DESIGN AND ANALYSIS OF MULTI-CHANNEL PROTOCOLS IN LOCAL AREA NETWORKS

by

HAMID KHOSHNEVIS

B.Sc., Purdue University, 1979

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

Department of Electrical Engineering

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

May 1982

Department of Electrical Engineering

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date June 7, 1982

ABSTRACT

Current implementations of local area networks utilize a single channel for packet transmission. The need for multi-channel networks arises whenever (a) the population of users is large, and the extra hardware cost incurred by increasing the data rates in response to higher user demands, is not justified, (b) the bandwidth is available only in segments (such as in CATV systems), or (c) the use of high data rates causes an unacceptably high bit-error probability.

We propose two different multi-channel access control protocols. The first, SRMA/m, utilizes a distributed scheme for request, and a central scheme for message packet scheduling. The second protocol, MCMA/m, is totally distributed, avoiding much of the hardware complexity of SRMA/m.

The performance of the protocols is studied using a combination of analytical and simulation methods. The performance criteria are efficient bandwidth utilization and small average delays. The same system parameters are used for both protocols, making a comparative study possible.

The delay vs. throughput performances of the two protocols are compared. The comparison indicates that for light to intermediate loads, MCMA/m shows a better performance (i.e. lower delays for the same throughputs). For higher loads, the relative performance depends on m, the number of channels. For small values of m ($1 \le m \le 4$), the overall performance of MCMA/m is superior. However as m increases,

SRMA/m shows a better performance in the form of lower delays and fixed capacity (versus decreasing capacity of MCMA/m).

In conclusion, the choice between SRMA/m and MCMA/m is based on a direct cost/performance tradeoff. For most ranges of parameters, SRMA/m achieves a superior performance at the expense of higher hardware complexity and cost.

TABLE OF CONTENTS

LIST OF FIGURES

NOTATION

ACK	Acknowledgement
AR	Answer-to-Request
CSMA	Carrier Sense Multiple Access Protocol
CSMA-CD	CSMA with Collision Detection
G/G/m	A queue with General (G) arrivals distribution and General (G) service time distribution and m servers
LAN	Local Area Networks
M/D/m	A queue with Poisson (M) arrivals and fixed (D) service time distribution and m servers
MCMA/m	Multi-Channel Multiple Access Protocol with m message sub-channels
R	Request
SRMA/m	Split-Channel Reservation Multiple Access Protocol with m message sub-channels

ACKNOWLEDGEMENTS

I am grateful to my research supervisor, Dr. R.W. Donaldson, for his enthusiasm, patience, and help throughout the life of the thesis.

Special thanks are extended to Dr. J.K. Cavers of MacDonald Dettweiler and Associates for helpful suggestions and providing many of the key references.

Grateful acknowledgement is given for the university research assistanship received.

I INTRODUCTION

1.1 Multi-Channel Protocols

The field of Local Area Networking (LAN) has witnessed rapid growth in the past few years. This growth is mainly due to great increase in the use of intelligent devices in office environments. It is now possible to establish communication among these often diverse devices for the purpose of sharing information, and facilitating access to expensive resources (such as mass-storage devices, printers, and office copiers).

A few characteristics distinguish LAN from other types of networks. The distances are short (typically <10 miles); as a result the ratio of end-to-end propagation delay to message transmission delay is small. The medium might be privately owned and inexpensive (such as a twisted pair or coaxial cable connecting various offices in a building), or leased and expensive (such as CATV[1] channel). The individual user demand is bursty in nature, i.e. the peak-to-average use ratio is large. The users do not utilize the network at all times, but when they do, a quick response is required.

A number of protocols have been devised to regulate access to the channel. The common feature of the physical implementations of these protocols is transmission rates typically less than 10 Mbits/sec. There are many applications where even though the traffic intensity dictates higher data

---

[1] Community Antenna TeleVision (Cable Television)

rates, a simple increase in the bit rate is not a practical solution.

As an example, we consider a system where the user population is large, and thus the hardware costs comprise a large part of the total system cost. As a general rule, the hardware costs increase as data rates increase, but beyond a point, these costs increase rapidly and disproportionately. The throughput/cost tradeoffs discourage the use of high data rates in these applications.

In some other applications, bandwidth (and not the hardware cost) is the restricting resource. A good example is CATV channel where the available bandwidth is in segments, i.e. interspersed among voice and video sub-channels. Although a bandwidth segment can probably support low-speed data rates (e.g. terminal-computer communication), in a very high-speed application (e.g. computer-computer communication), utilization of all the available bandwidth segments becomes necessary.

Finally, an increase in bit-rate may cause an unacceptable increase in bit-error probability, or in message retransmissions if error detection is used.

In summary, there can arise many situations where a simple increase in the transmission rate, in response to higher user demands from the system, is not an effective solution. For such applications, the alternative is to split the available bandwidth into a number of sub-channels. The additional hardware requirement (over single-channel systems) is for accessing a multiple number of sub-channels. This can be accomplished by using inexpensive frequency synthesizers, or crystal oscillators, to generate the appropriate sub-channel carrier frequencies.

The subject of the thesis is the design of protocols capable of utilizing the sub-channels in an efficient manner. The efficiency depends on many criteria. Besides high bandwidth utilization and small delays, the ability of the protocol to support different packet sizes, and different packet priorities, is worthy of further study. In this thesis however, the investigation of performance is limited to average delay as a function of throughput.

The delay performance of a single-channel network is always superior to that of a multi-channel network utilizing the same amount of bandwidth. In order to study the performance degradation as a function of m ($m>1$), we assume that the bandwidth available (for all values of m) is fixed.

We propose and analyze two multi-channel protocols. These protocols differ mainly in the control of access to the sub-channels. The control can be central, distributed, or a hybrid of the two. Distributed control refers to a situation where all users execute a common control strategy, as opposed to central control where a central computer schedules all the transactions.

One of the protocols considered is Split-Channel Reservation Multiple Access with m sub-channels (SRMA/m), a multi-channel extension of SRMA [TOBA 76]. The protocol employs distributed control for service requests, and central control for regulating access to the message sub-channels.

Although the SRMA/m protocol is very efficient for intermediate to heavy load conditions, its implementation might prove to be complicated because of the use of two different schemes (central and distribted). Therefore we propose a second protocol, Multiple Channel Multiple Access with m sub-channels (MCMA/m), which is a multi-channel superset of CSMA-CD[1] [TOBA 79]. The MCMA/m protocol is totally distributed, greatly simplifying the hardware design.

1.2 Review of Previous Work

One of the first protocols to exploit the characteristics of LAN was Carrier-Sense Multiple Access (CSMA). The protocol is described by Kleinrock and Tobagi [KLEI 75a], and its performance is derived using the infinite population model. A simulation study of the average packet delay for different loads is also carried out. The effect of acknowledgement traffic on the throughput was presented later [TOBA 78].

When the terminals are not all in line of sight or in range of each other, carrier sensing becomes impossible. The authors in their second paper [TOBA 75] propose a solution called Busy-Tone Multiple-Access (BTMA). A central controller detects carriers due to transmitting terminals and sends busy tones on a separate channel to inform the other terminals of the transmissions.

[1] Carrier-Sense Multiple Access with Collision Detection

Extending the work of Lam and Kleinrock [LAM 75], Tobagi and Kleinrock examine the stability of random access techniques [TOBA 77]. It is shown that for infinite population with bursty demand, the throughput of random-access protocols[1] goes to zero in finite time with probability of one! Heuristic dynamic control procedures are introduced that show good results under heavy load conditions. This paper includes the throughput-delay tradeoff analysis for finite population case [KLEI 75b].

The analysis of CSMA-CD is due to Tobagi and Hunt [TOBA 79]. Ethernet, a variation of CSMA-CD, is described by Metcalfe and Boggs [METC 76]. The retransmission delay is an exponential function of number of collisions, as a result the system introduces long delays under heavy loads. The Ethernet protocol has not yet yielded to analysis because of its collision-dependent retransmission delay. A number of clever variations have recently appeared in the literature [AGRA 77] and [TOKO 77].

The SRMA protocol is described by Tobagi and Kleinrock [TOBA 76]. The performance of the system, with the request channel operated using various random-access schemes, is also analyzed in the same paper.

The books by Kleinrock on queueing theory [KLEI 75b], and computer applications [KLEI 76], lay the foundation for much of

---

[1] Protocols in which the terminals transmit the packets on a common channel (or channels) with no pre-scheduling, i.e. the packets carry their own control information.

the theoretical work in the area. A book by Schwartz [SCHW 77] contains information on operational networks.

An exceptionally thorough survey on the subject of packet networks is presented by Tobagi [TOBA 80]. Clark et al. [CLAR 78], and Biba and Yeh [BIBA 79], offer useful insight into hardware implementation. Meisner et al. [MEIS 77] describe an interesting hybrid of broadcast and ring-oriented protocols.

1.3 Scope of the Thesis

The purpose of the thesis is to design, analyze, and compare efficient multi-channel protocols. Two protocols are considered: SRMA/m, a hybrid protocol, and MCMA/m, a totally distributed protocol. In both cases, the population of terminals M is assumed fixed (M=50).

The first protocol considered is SRMA/m. The protocol, when the request channel is operated in CSMA-CD, is analyzed in Chapter 2. The single-channel SRMA system is compared to the previously analyzed SRMA/1, where the request channel was operated in CSMA.

In Chapter 3, the error due to the assumption that the interdeparture times of the request packets are exponentially distributed, is studied. The analysis is carried out only for SRMA/1.

Chapter 4 deals with MCMA/m protocol. The protocol is described, and its performance is analyzed with the same parameter values used in the analysis of SRMA/m.

A summary of the work, and comparison of the two protocols, appears in Chapter 5. Suggestions for further work conclude the thesis.

II SRMA/m PROTOCOL

2.1 Introduction and Statement of Objectives

Split-Channel Reservation Multiple Access protocol (SRMA) was introduced by Tobagi and Kleinrock [TOBA 76]. In this protocol the total channel bandwidth is divided into three sub-channels (simply called channels if no ambiguity arises): a message sub-channel, a request (R) sub-channel, and an answer-to-request (AR) sub-channel. The name of the protocol is hereafter modified to SRMA/m, which stands for a SRMA system with the message sub-channel split into m separate channels.

First, we briefly describe the operation of SRMA/1. All terminals that have message packets ready for transmission, send request packets on the R channel. If a number of terminals submit time-overlapping R packets, these packets collide and destroy each other. As a result, the R channel requires an access control scheme (such as CSMA) for operation.

When a request packet is received at the controller, the controller determines the time when the message channel becomes free, and sends the timing information back to the transmitter (on AR channel). At the specified time, the terminal transmits the message packet.

In SRMA/m (m>1), the central controller schedules the message packet for the message channel that has the smallest number of queued packets awaiting transmission. Since there are m message channels in the m>1 case, the destination device must be informed of the channel for which the packet is scheduled. This is accomplished by requiring the receiver to listen to the AR channel, buffer any AR packets that contain the receiver's address in their header, and determine the channel assignment.

In the absence of central control, there can arise situations where two or more packets are addressed to the same terminal at overlapping times (and on different message channels). In this case, all but one of the packets are lost. The central controller avoids the problem by scheduling the packets for non-overlapping times.

The operation of the SRMA/m protocol for m=1 is illustrated in Fig. 2.1.

The objectives of this chapter can be stated as follows:

1) Revision of the performance of SRMA/1 with the request channel operated in CSMA: The CSMA delay curves employed in the previous study of SRMA/1 [TOBA 76] were taken from [KLEI 75a]. These CSMA curves were obtained by simulation and were rendered obsolete by a new analytical study [TOBA 79]. We revise the performance of SRMA/1 based on the new CSMA delay curves (reproduced in Appendix A).

2) Extension of the analysis to CSMA-CD: The performance of SRMA/1 is extended to the case where the request channel is operated in CSMA-CD (the curves for CSMA-CD also appear in Appendix A).

3) Extension of the analysis to SRMA/m (m>1): In this case, the R channel is operated in CSMA-CD because CSMA-CD is a more

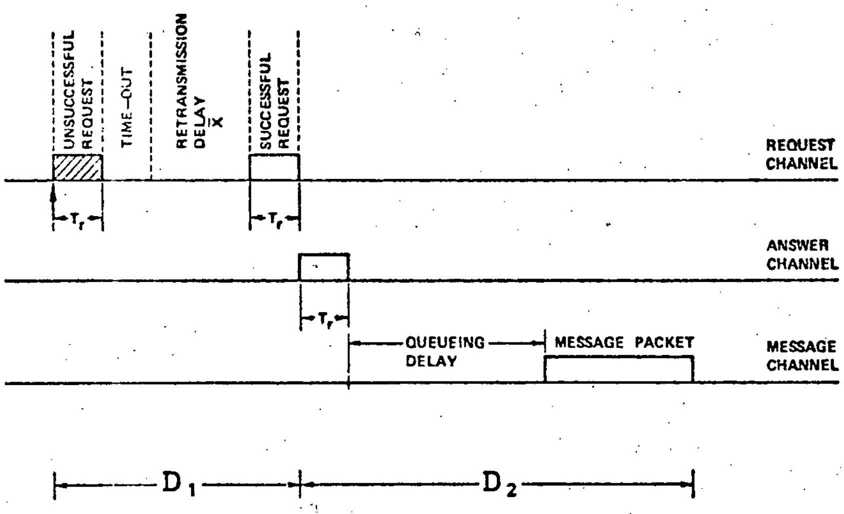


Fig. 2.1 SRMA/1 [TOBA 76]

efficient protocol (compared to CSMA).

2.2 Assumptions

Throughout the thesis, the population of terminals M is fixed (M=50). Furthermore, the terminals are assumed to have a constant rate of packet generation with an aggregate mean of λ packets/sec.

The processing time of the central controller is assumed negligible.

As discussed before, the central controller resolves conflicts due to overlapping message packets (destined to the same terminal on different channels). This is accomplished by scheduling the packets for non-overlapping times. In the following analysis of SRMA/m, we assume that the extra delay incurred due to this strategy is negligible.

The error due to neglecting the extra delay is greatest when:

1) the number of terminals is large; for a given load, the probability of addressing the same terminal increases.

2) the number of channels is large; the probability of time-overlapping transmissions (on different channels) increases (assuming that the number of terminals exceeds the number of channels).

In order to be consistent with the previous analysis of SRMA/1 [TOBA 76], the ratio of request packet's transmission delay to request channel's end-to-end propagation delay is assumed fixed at 100, independent of the percentage bandwidth allocated to the R channel. Therefore, the CSMA and CSMA-CD curves of Appendix A can be used for the analysis.

2.3 Terminology and Notation

The following terminology is used throughout the thesis:

Throughput — percentage of time a channel is occupied by successful transmissions

load — traffic intensity (which includes both successful and unsuccessful packets)

performance — average delay as a function of throughput (better performance corresponds to lower delay)

The following terminology and notation are used in the second and third chapters only:

Slot — end-to-end propagation delay of a sub-channel (typical value: 5 μs)

ν — retransmission coefficient of the request channel's random access protocol, defined as the probability of a terminal with a backlogged packet sensing the channel in the current slot

L — size of the request packet in request slots (L=100 throughout this chapter)

Furthermore we introduce the following notation:[1]

W — total bandwidth available

W_m — bandwidth assigned to the entire message channel

---

[1] The notation and the theoretical development of SRMA/1 protocol strongly follows the work of Tobagi and Kleinrock [TOBA 76].

W_r bandwidth assigned to the request channel

W_a bandwidth assigned to the answer-to-request (AR) channel

b_m number of bits in a message packet

b_r number of bits in a request packet

b_a number of bits in an AR packet

θ fraction of bandwidth assigned to the entire message channel; $\theta = W_m / W$

In addition:

T_w transmission time of the message packet on the entire bandwidth; $T_w = b_m/W$

T_m transmission time of the message packet on the entire message channel; $T_m = b_m/W_m$

T_r transmission time of the request packet on the request channel; $T_r = b_r/W_r$

T_a transmission time of the AR packet on the AR channel; $T_a = b_a/W_a$

n_r b_r/b_m

n_a b_a/b_m

S_r throughput of the request channel

S_m throughput of the total message channel

S throughput of the SRMA/m system over the entire bandwidth; $S = S_m \theta$

D_r delay due to the request packet, from the time packet is generated until the time it is received by the controller

SRMA/m/RAND/y SRMA/m system for which the request channel is operated in RANDom-access scheme (e.g. CSMA), and η=y

2.4 Components of the Delay

The total delay of the SRMA/m system consists of the following delays:

D_1 delay from the time a request is made until the time the answer-to-request packet is initiated (by the controller)

D_2 delay from the time the AR packet is initiated until the time the message packet is completely transmitted

2.5 Analysis of D_1

Since the request and AR packets basically carry the same kind of information, they are assumed to have the same size:

$$b_r = b_a \tag{2.1}$$

We can therefore define η as $\eta=\eta_r=\eta_a$.

Under steady-state conditions, the AR channel should have enough capacity to sustain the fastest rate at which AR packets are generated. This rate is the same as the rate at which (successful) R packets arrive at the controller. Since in CSMA-type protocols, the rate of arrival of R packets can approach 1 packet/T_r, the capacity of AR channel should at least be 1 packet/T_r ; therefore we allocate equal amounts of bandwidth to R and AR channels. If θ is the percentage bandwidth allocated to the message channel, the amounts of bandwidth allocated to the R and AR channels in terms of W (the total bandwidth) are:

$$W_r = W_a = (1-\theta)W/2 \qquad (2.2)$$

Then using (2.1), transmission times of the R and AR packets are equal:

$$T_r = T_a \qquad (2.3)$$

The delay D_1 is equivalent to $D_r(S_r)$. For each value of S_r, D_r (in units of T_r) can be obtained from Appendix A. We can write:

$$D_1 = D_r(S_r).T_r \quad \text{(seconds)} \qquad (2.4)$$

2.6 Analysis of D_2

2.6.a Components of D_2 and Interdeparture Times Distribution

The delay D_2 consists of (a) the delay from the time the AR packet is initiated until the time it is received at the terminal, and (b) the delay from the time the AR packet is received until the time transmission of message packet is completed.

Part (a) is a simple transmission delay, so we first concentrate on Part (b). Part (b) of D_2 is the "queueing delay" of message channel. We introduce "waiting delay" as the queueing delay less the transmission delay.

In order to analyze the queueing delay, it is necessary to know the distribution of the arrival of AR packets at the terminals. We observe that this arrival has the same distribution as the departure of request packets from the request channel, since the processing delay of the controller is assumed negligible and the AR channel introduces only a constant transmission delay.

In order to find the interdeparture times distribution of the request packets, a simulation study of CSMA-CD protocol with M=50 terminals was carried out (Appendix B). The probability density function of the interdeparture times are presented in Fig. 2.2 for different values of S_r. Exponential pdf's with mean $1/S_r$ are also shown for comparison.

Since no more than one packet can be successfully transmitted on the request channel at a time, the interdeparture times in the interval of 0 to 1 unit of T_r can be seen to have 0 density. Compared to the exponential pdf, the interdepartures in the interval between 1 and 1.5 units show unusually high pdf's; as the throughput increases, the pdf's in this interval tend to increase.

In general, the exponential distribution seems to fit the histograms of Fig. 2.2 well. In order to develop a uniform interdeparture model for all load levels, we assume that the interdeparture times are exponentially distributed. In other words, the departing R packets are assumed to form a Poisson source.

Since the exponential assumption does not fit the

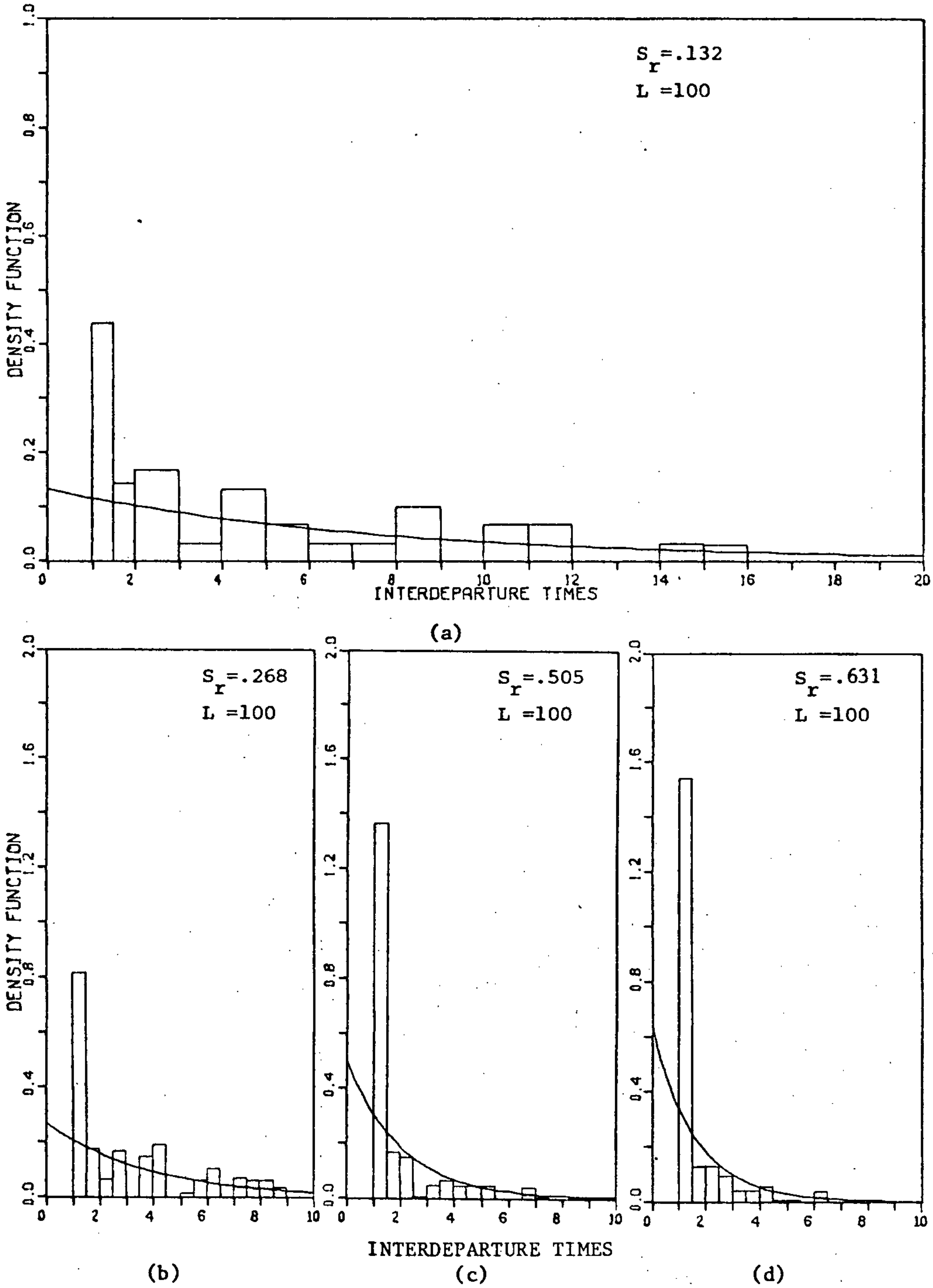


Fig. 2.2. Probability density function of interdeparture times in units of packet transmission times for CSMA-CD. (a) S_r=.132 (b) S_r=.268 (c) S_r=.505 (d) S_r=.631

histograms in the 0 to 1.5 units interval, this assumption introduces errors that are analyzed in the next chapter.

2.6.b Performance of M/D/m Queue

Because of the exponential interarrivals assumption and constant message packet size assumption, the delay due to Part (b) of D_2 corresponds to that of the M/D/m queue. In this section we analyze the delay of the M/D/m queue. The notation used is local to this section and does not correspond to that of SRMA.

An exact solution for M/D/m queueing delay does not exist. The delay, however, can be bounded. We carry out the analysis with the hope of finding tight bounds which can be directly used in the delay analysis.

For the G/G/m queue, the following bounds on the waiting delay exist [KLEI 75b]:

$$\hat{W}_t - \frac{\{(m-1)/m\} \; \overline{X^2}}{2\,\overline{X}} \leq W_t \leq \frac{\sigma_a^{\,2} + (1/m)\sigma_b^{\,2} + \{(m-1)/m^2\}\,\overline{X}^2}{2\,\overline{t}\,(1-S)} \qquad (2.5)$$

Where:

$\hat{W}_t$	waiting time for a G/G/1 system
m	number of servers (channels)
x	service time random variable
t	interarrival time random variable
S	throughput of the queueing system
σ_a	equivalent to σ_t, standard deviation of t
σ_b	equivalent to σ_x, standard deviation of x

For M/D/1 queue:

$$\hat{W}_t = \frac{\overline{X}\,S}{2(1-S)} \tag{2.6}$$

The transmission delay $\overline{X}$ of (2.6) corresponds to a M/D/1 queue with the same total bandwidth as a M/D/m queue. In order to apply the delay of (2.6) to (2.5), $\overline{X}$ has to be normalized to that of a M/D/1 queue with the same bandwidth as a M/D/m queue; therefore $\overline{X}/m$ replaces $\overline{X}$ [BRUM 74]:

$$\hat{W}_{t,norm} = \frac{1}{m}\,\frac{\overline{X}\,S}{2(1-S)} \tag{2.7}$$

For M/D/m queue:

$$\overline{t} = \frac{\overline{X}}{m\,S} \tag{2.8}$$

Since the service time is fixed:

$$\sigma_b = \sigma_x = 0 \tag{2.9}$$

Interarrival times are exponentially distributed; thus:

$$\sigma_a^{\,2} = \overline{t}^2 \tag{2.10}$$

By substituting from (2.7)-(2.10) into (2.5), the delay of M/D/m queue can be bounded by:

$$\frac{S}{2m(1-S)} - \frac{m-1}{2m} \leq \frac{W_t}{\overline{X}} \leq \frac{1+S^2(m-1)}{2mS(1-S)} \tag{2.11}$$

where $\overline{X}$ is the packet transmission time on a sub-channel.

The total queueing delay is:

$$D_{M/D/m} = (W_t + 1)\overline{X} \qquad \text{(seconds)} \tag{2.12}$$

It should be noted that for m=1, the bounds of (2.12) converge to (2.6). The performance bounds for m=1,4,8 are shown in Fig. 2.3 where the delay is in units of packet transmission times over the entire bandwidth.

Unfortunately the bounds are not tight and cannot be directly used in the analysis. In order to find the M/D/m delay, a simulation study was carried out (Appendix C). The results from the simulation study also appear in Fig. 2.3. Except for throughputs close to 1, the simulation results appear to fall within the theoretical performance bounds.

For light loads, the waiting delay is very small and the packet delay is equal to the transmission delay; therefore, the delay of M/D/m system is m times the delay of M/D/1 (that has the same total bandwidth). For higher loads, the delays tend to converge. We conclude that the degrading effect of splitting the bandwidth decreases with increasing loads.

2.6.c Total Value of D_2

The total value of D_2 can be written as:

$$D_2 = \underbrace{T_a + T_a/L}_{\text{part (a) of } D_2} + D_{M/D/m}(S_m) . T_m \quad \text{(seconds)} \tag{2.13}$$

where T_a/L is the delay before the first bit of AR packet arrives at the terminal, and $D_{M/D/m}$(in units of T_m) can be read directly from Fig. 2.3.

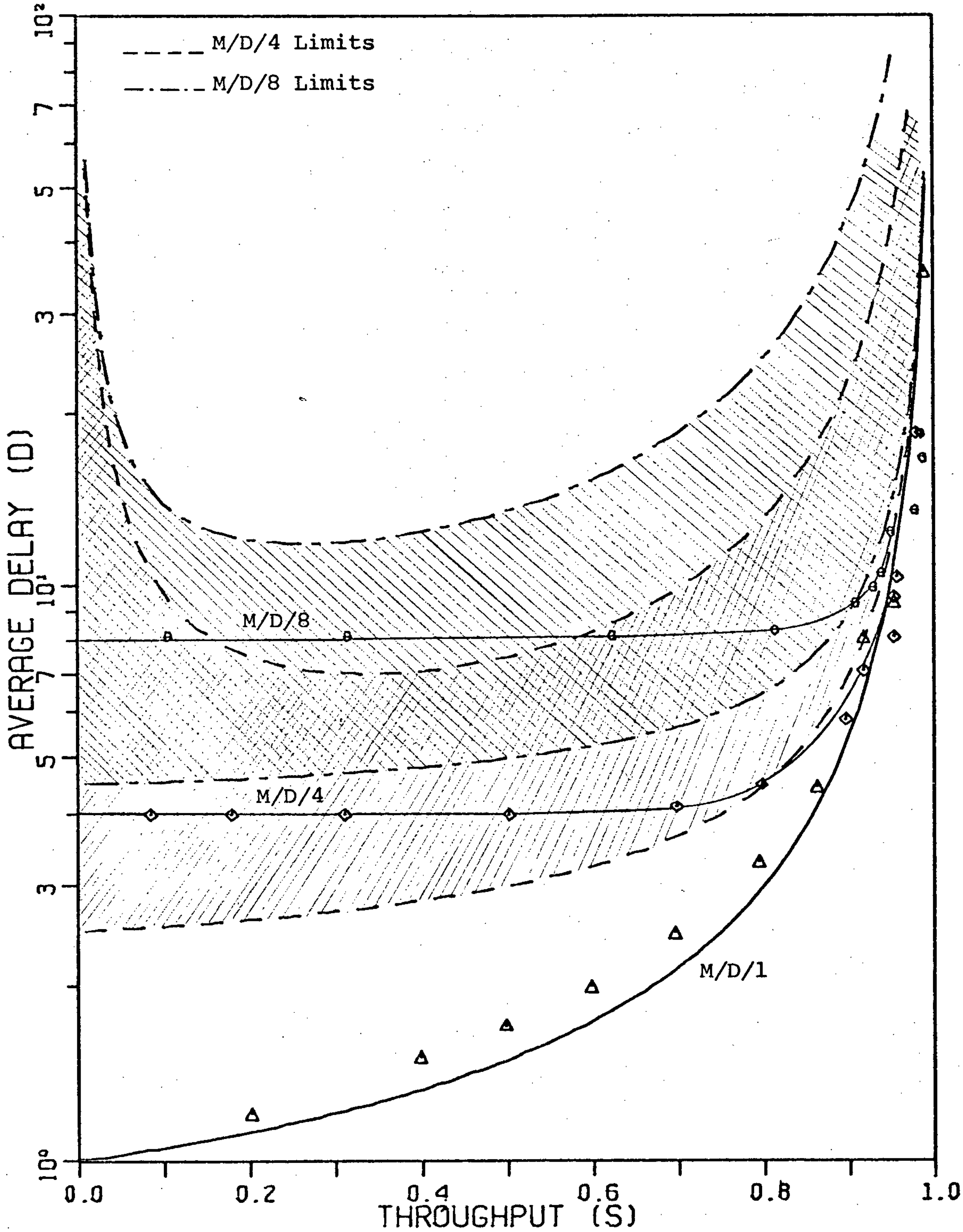


Fig. 2.3. M/D/m: Delay in units of packet transmission times vs. throughput , from simulation study and delay bounds.

2.7 Total Delay D

The total delay D consists of D_1 (2.4) and D_2 (2.13):

$$D = D_r(S_r).T_r+T_a+T_a/L+D_{M/D/m}(S_m).T_m \quad \text{(seconds)} \tag{2.14}$$

By using definitions of T_r and T_m:

$$\frac{T_r}{T_m} = \frac{b_r}{b_m} \cdot \frac{W_m}{W_r} = \eta \frac{\Theta}{(1-\Theta)/2} \tag{2.15a}$$

$$\frac{T_m}{T_w} = \frac{b_m}{b_m} \cdot \frac{W}{W_m} = \frac{1}{(1-\Theta)} \tag{2.15b}$$

$$\frac{S_r}{S_m} = \frac{\lambda T_r}{\lambda T_m} = \frac{\Theta}{(1-\Theta)/2} \cdot \eta \tag{2.15c}$$

$$S_r = \frac{S_r}{S_m} \cdot S_m = \frac{\eta}{(1-\Theta)/2} \cdot S \tag{2.15d}$$

The total delay (in units of T_w) using (2.15a) and (2.15b) is written as:

$$D = D_1+D_2 = D_r(S_r)\frac{\eta}{(1-\Theta)/2}+\frac{\eta(1/L+1)}{(1-\Theta)/2}+D_{M/D/m}(S_m)/\Theta \tag{2.16}$$

where $S_m=S/\theta$ and S_r is given by (2.15d).

It is clear from (2.16) that θ, as well as η affect the delay in a complicated way. By definition, η is proportional to b_r. Addressing information comprises a substantial part of the R packet, because the R packet mainly carries the source and destination addresses. Since the address size is logarithmically proportional to M, η is roughly proportional to

Log(M). In this study η is limited to .01 and .10.

Our goal now is to find a value of θ that minimizes the delay. The delay vs. θ curves for different values of S are presented in Fig. 2.4. In this figure, for constant values of S, θ is bounded by $S \le \theta \le 1-2\eta S/C_r$ [TOBA 76].

We define θ(opt) as the value of θ that minimizes the delay. It is clear from Fig. 2.4 that a single value of θ cannot minimize the delay for all ranges of S. The sensitivity of delay to the value of θ increases dramatically with the increase in throughput, so it is reasonable to choose a θ that minimizes the delay for higher throughputs.

By examination of Fig. 2.4:

For	SRMA/1/CSMA/.01	θ(opt)=.97
For	SRMA/1/CSMA/.1	θ(opt)=.8
For	SRMA/1/CSMA-CD/.01	θ(opt)=.97
For	SRMA/1/CSMA-CD/.1	θ(opt)=.83

Figs. 2.5a and 2.5b show the performance curves for SRMA/4 and SRMA/8 respectively. The computer-generated contour lines are drawn using interpolated data from M/D/m delay curves. The lines, for most ranges of S seem to be broken. This is due to the rapid change in the M/D/m delay for high throughputs. The value of θ(opt) for the case η=.01 appears to be roughly equal to θ(opt) of SRMA/1. For η=.1 θ(opt)=.78 which is slightly smaller than before.

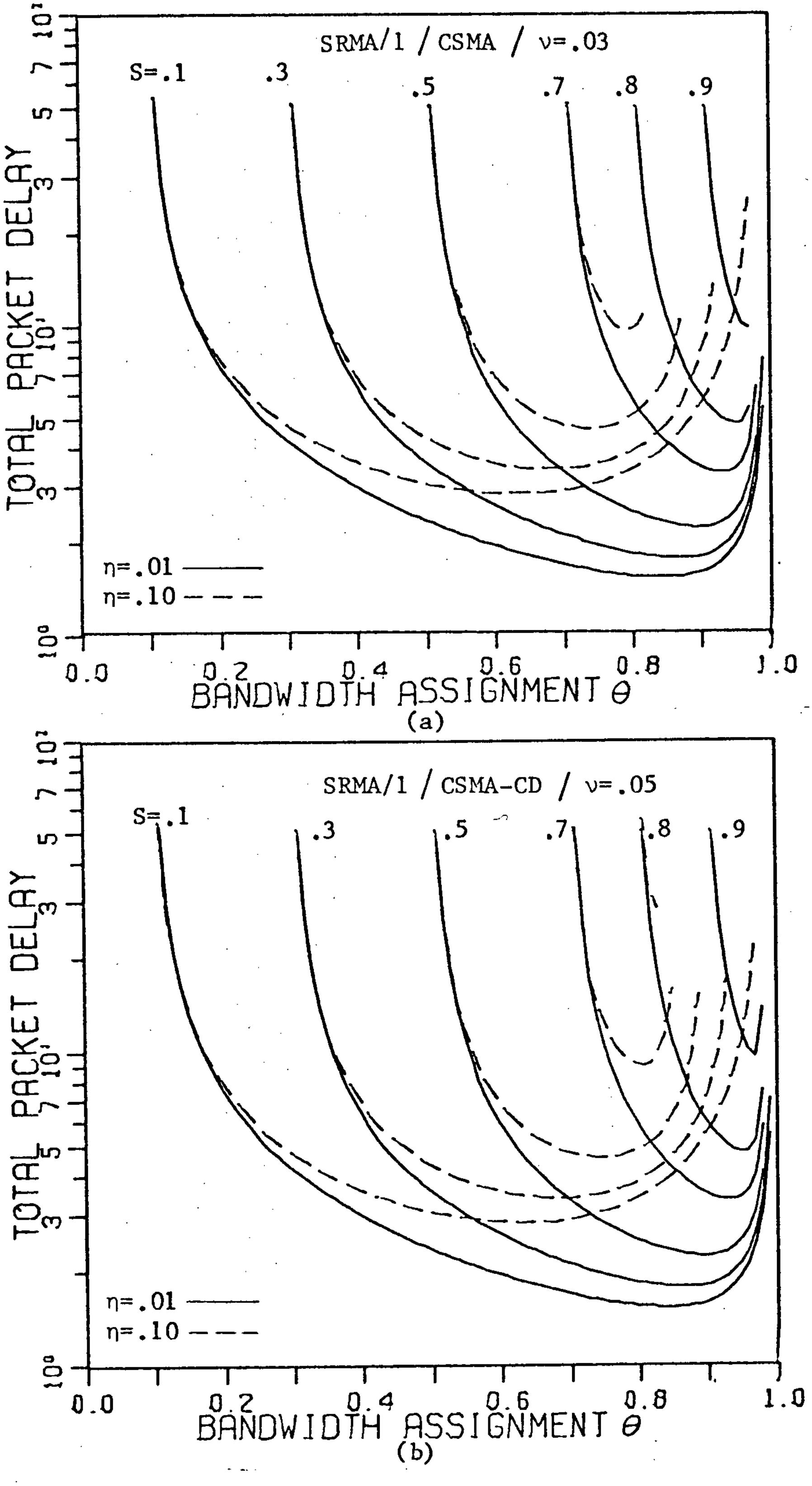
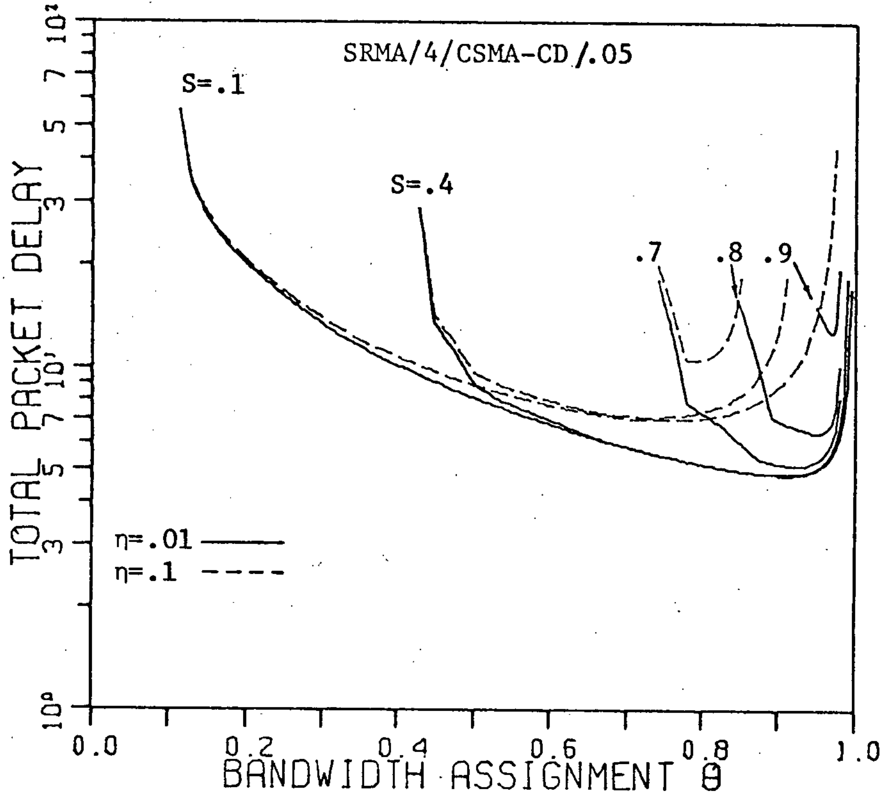
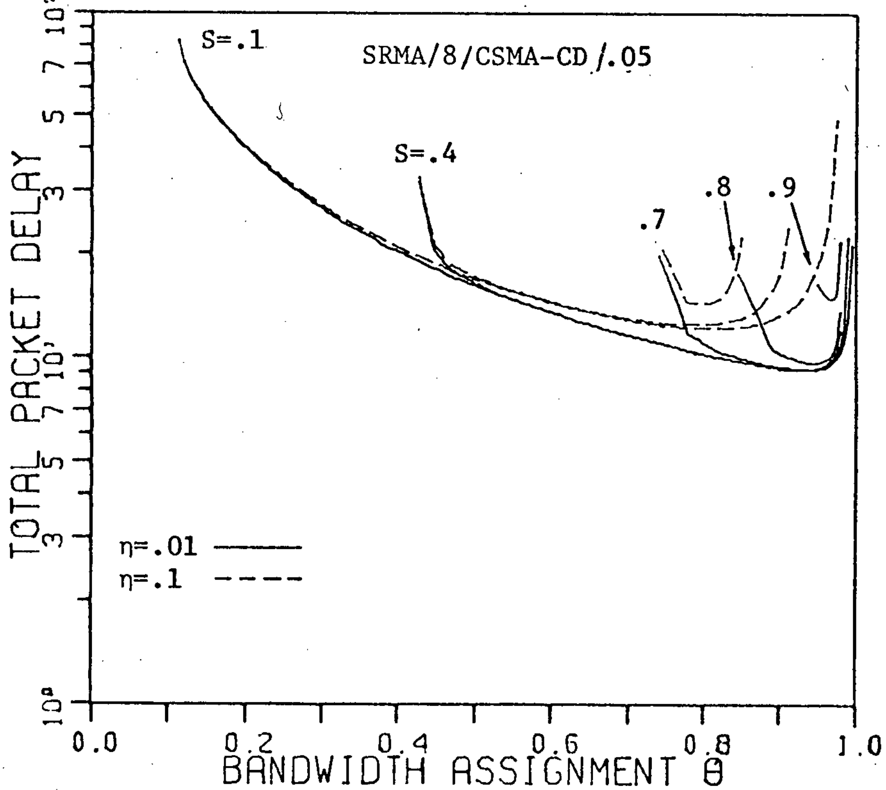


Fig. 2.4. SRMA/1: Packet delay in units of T_w vs. bandwidth assignment Θ for different throughputs. Request channel is operated in (a) CSMA (b) CSMA-CD



(a)



(b)

Fig. 2.5. SRMA/m: Packet delay in units of T_W vs. bandwidth assignment Θ for different throughputs. (a) m=4 (b) m=8
Request channel is operated in CSMA-CD

2.8 Discussion

The delay vs. throughput curves for SRMA/m are presented in Fig. 2.6. In order to investigate the effect of the random access scheme used, we substitute C_r for $S_r(max)$ in (2.15d):

$$S_{max} = \frac{(1-\Theta)/2}{\eta} C_r \tag{2.17}$$

If $\eta=.1$ then:

For CSMA: $C=.8$ $\theta(opt)=.8$ so: $S(max)=.80$

For CSMA-CD: $C=.95$ $\theta(opt)=.83$ so: $S(max)=.81$

As the length of the message packet increases (smaller η), the effect of the request channel's capacity on the total capacity diminishes. As an example, for $\eta=.01$, the effect of the random access scheme on the capacity is negligible and the system can approach throughputs close to 1. For $\eta=.1$, however, the effect of random access scheme used is more important and results in capacity roughly equal to the capacity of the random access scheme.

Comparing the effect of the random access scheme on the throughput vs. delay curves of Fig. 2.6, we can see that for $\eta=.01$, SRMA/1/CSMA and SRMA/1/CSMA-CD curves overlap. For $\eta=.1$, SRMA/1/CSMA has a better performance for lighter loads, although for heavier loads its performance is worse than that of SRMA/1/CSMA-CD. This effect is due to the chosen $\theta(opt)$ which optimizes the SRMA/1/CSMA-CD performance for relatively high throughputs.

The previous conclusions about M/D/m vs. M/D/1 are applicable to SRMA/m vs. SRMA/1. For low throughputs, an

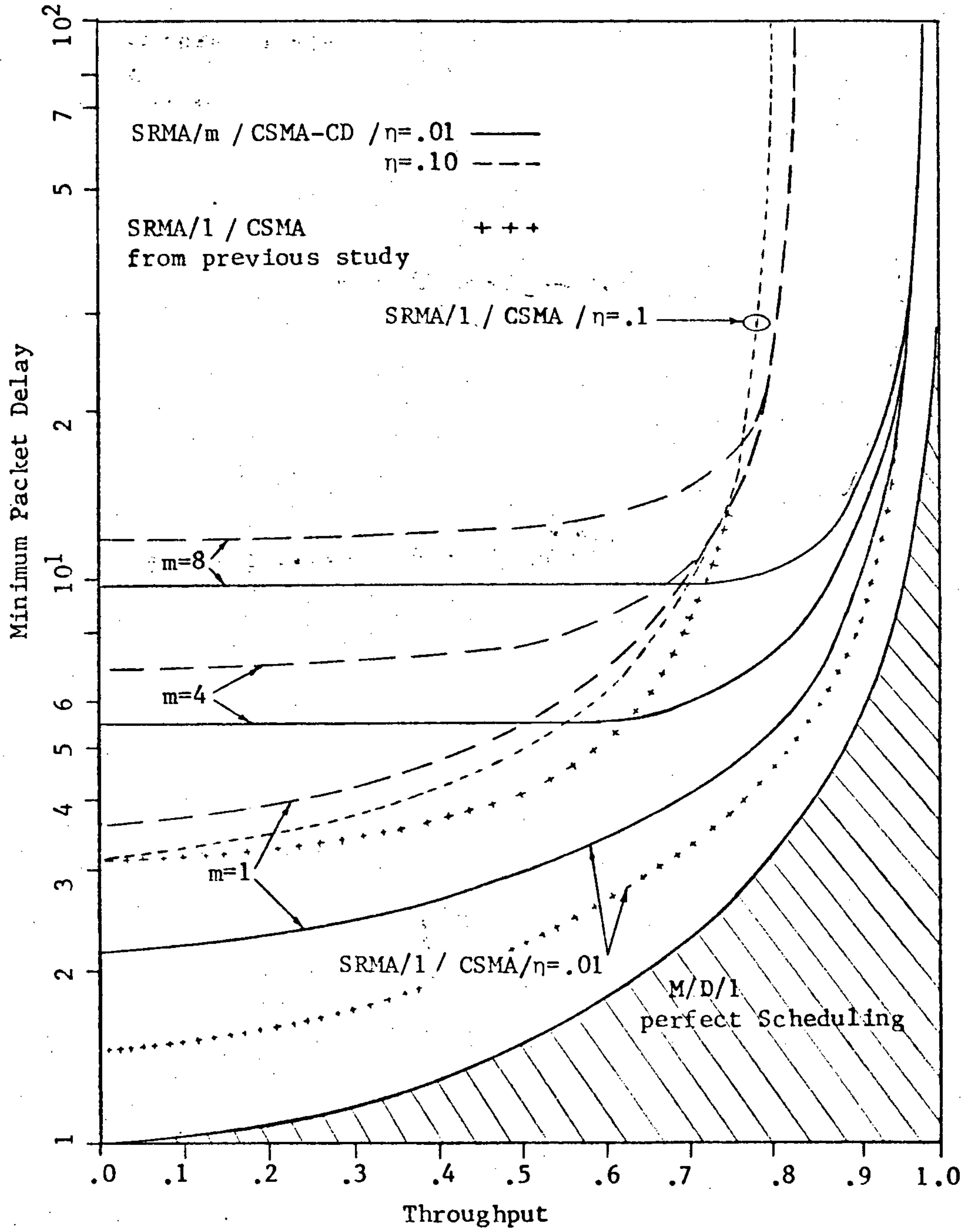


Fig. 2.6. SRMA/m: Delay in units of T_w vs. throughput

increase in the number of sub-channels increases the delay dramatically. For high throughputs (e.g. heavy loads) the delays converge. Compared to the M/D/m, the delay convergence starts at even lower throughputs (e.g. .65 for η=.1).

We are generally more interested in the heavy-load performance of multi-channel protocols, since if the system was characterized by moderate loads, we could use a single-channel system. From Fig. 2.6, for loads close to the capacity of the system, the increase in the delay resulting from the channel splitting is negligible.

In order to compare the new and old studies of SRMA/1/CSMA, the old SRMA/1/CSMA curves are reproduced in Fig. 2.6 (from [TOBA 76]). For light loads and for both values of η, the previously obtained delays are smaller. The difference can be explained by the fact that the CSMA delay curves used, as well as the values of θ(opt) obtained, are different from before.

In summary, SRMA/m appears to be an efficient multi-channel communication protocol that offers good performance for intermediate to heavy load conditions.

III ERROR ANALYSIS IN SRMA

3.1 Introduction and Statement of Objectives

In the previous chapter, it was assumed that the arrival of AR packets to the terminals follows a Poisson distribution. By using this assumption, we were able to carry out the performance analysis based on the vast amount of information available in the literature dealing with Poisson arrivals.

In this chapter, we take a second look at the assumption. Although from Fig. 2.2, the assumption seems to fit for most ranges of interarrivals, it does not fit for interarrivals in the 0 to 1.5 units interval. The error due to this discrepancy would affect the performance curves.

Our objectives in this chapter are:

(a) analysis of the error due to exponential arrivals assumption, using values of θ(opt) found in the previous chapter

(b) making appropriate corrections to the performance curves of SRMA (plotted in the previous chapter).

For simplicity, we limit the scope of the study to SRMA/1/CSMA-CD and both values of η.

3.2 Examination of the Validity of Exponential Assumption

In order to examine the validity of the exponential assumption for interarrivals greater than 1.5 units, we employ the graphical representation of Kolmogorov-Smirnov goodness of fit test [BENJ 70]. In this test the cumulative distribution function (CDF or PDF) of the random variable is drawn on the probability paper[1], along with the negative and positive difference curves for different values of α, where α is the probability of rejecting a correct hypothesis based on the test. If the CDF crosses the difference lines then the hypothesis is rejected, otherwise it is accepted.

The test can be shown by the following equation:

$$D_n = \sup_t |F_n(t) - F_o(t)| \tag{3.1}$$

where $F_n(t)$ is the experimental CDF and $F_o(t)$ is the hypothesized CDF. If D_n is greater than the confidence interval for a value of α, then the hypothesis is rejected.

The test, for two values of S_r, is shown in Figs. 3.1a and 3.1b. From Fig. 3.1a the hypothesis can be accepted. Fig. 3.1b, however, shows that the match for small values of interarrival times are not acceptable, as is expected.

Our general conclusion is that the exponential assumption, for ranges of interarrivals greater than 1.5 units and for most ranges of S_r, is acceptable.

---

[1] The probability axis is scaled so that the CDF plots as a straight line.

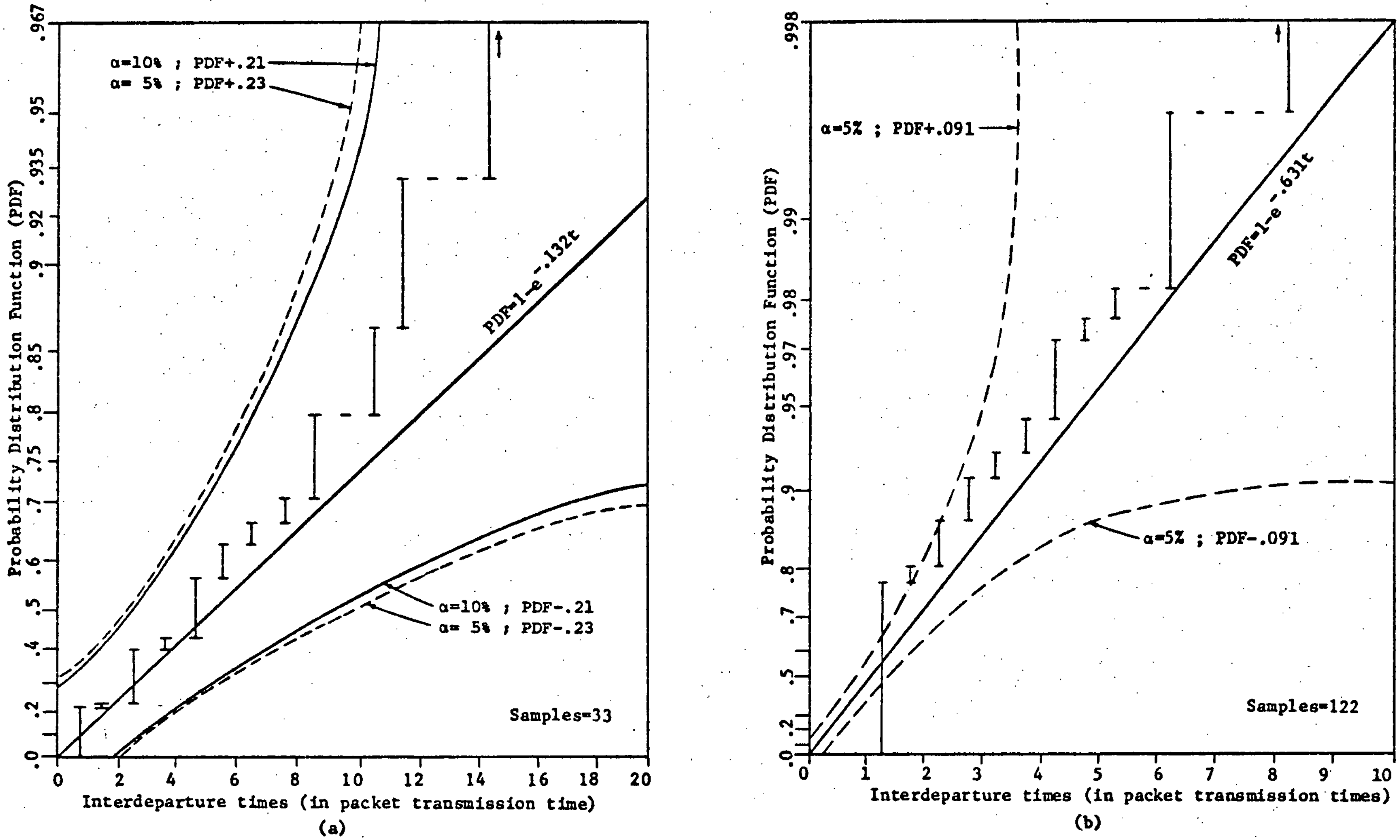


Fig. 3.1 The Kolmogorov-Smirnov goodness of fit test for the CDF of the interarrival times (a) $S_r=.132$ (b) $S_r=.631$

3.3 Interdeparture Model

In order to analyze the delay, it is necessary to find an equation that describes the pdf of interarrivals for all ranges of S_r. By inspection of Fig. 2.2, it can be concluded that the excess pdf in the 1 to 1.5 units interval is roughly equal to the missing pdf of the 0 to 1 units interval. We further assume that this pdf is uniformly distributed in the 1 to 1.5 units interval (Fig. 3.2). In the following analysis, unless otherwise specified, time is measured in unit of T_r.

The pdf of interarrival times can be written as:

$$\text{pdf} = f(t) = \begin{cases} 0 & 0 \leq t \leq 1 \\ \lambda e^{-\lambda t} + 2(1-e^{-\lambda}) & 1 \leq t \leq 1.5 \\ \lambda e^{-\lambda t} & 1.5 \leq t \end{cases} \tag{3.2}$$

where λ is the mean arrival rate in units of packets/T_r, and t is the interarrival time.

In order to determine the mean and variance of $f(t)$ we use the following equation [CRC 78]:

$$\int x^m e^{ax}\, dx = e^{ax} \sum_{r=0}^{m} (-1)^r \frac{m!\, x^{m-r}}{(m-r)!\, a^{r+1}} \tag{3.3}$$

The mean can be written as:

$$\bar{t} = \int_1^{1.5} \{\lambda e^{-\lambda t} + 2(1-e^{-\lambda})\}\, t\, dt + \int_{1.5}^{\infty} \lambda t e^{-\lambda t}\, dt \tag{3.4}$$

Using (3.3), (3.4) is finally written as:

$$\bar{t} = (1/\lambda - .25)e^{-\lambda} + 1.25 \tag{3.5}$$

The second moment of t is:

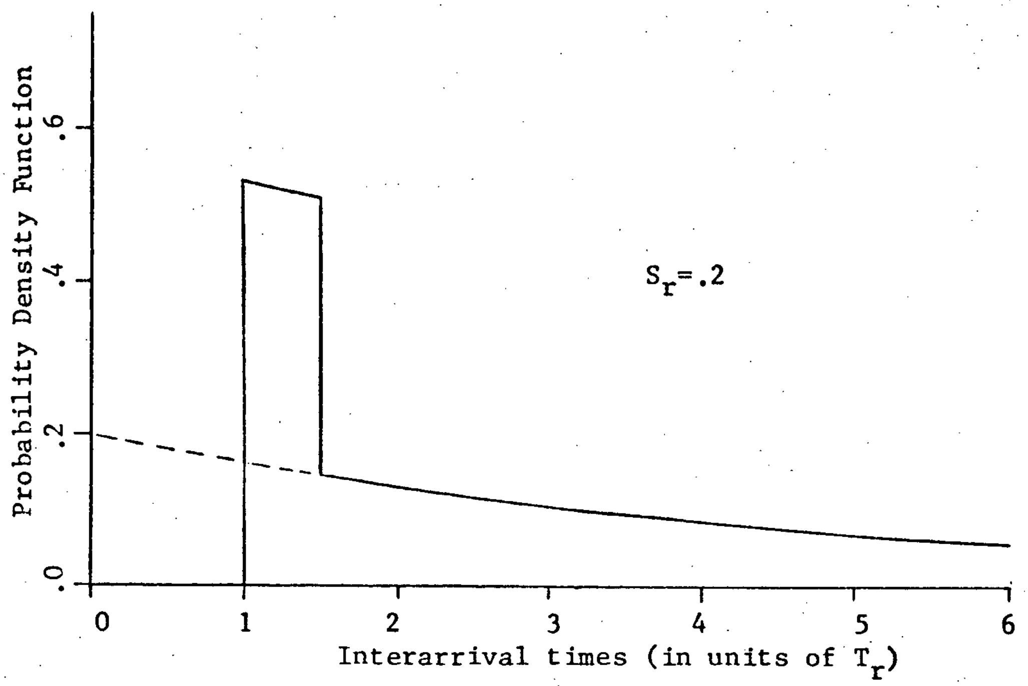


Fig. 3.2 The assumed interarrival times density function. The interarrivals are in units of T_r.

$$\bar{t^2} = \int_1^{1.5} \{\lambda e^{-\lambda t} + 2(1-e^{-\lambda})\}t^2\, dt + \int_{1.5}^{\infty} \lambda t^2\, e^{-\lambda t}\, dt \tag{3.6}$$

which simplifies to:

$$\bar{t^2} = e^{-\lambda}\,(-.583 + 2/\lambda + 2/\lambda^2\,) + 1.583 \tag{3.7}$$

The variance is then written as:

$$\sigma_t{}^2 = \bar{t^2} - \bar{t}^2 \tag{3.8}$$

3.4 Analysis of Delay vs. Throughput

The queueing delay of G/G/1 is given as [KLEI 75b]:

$$D = \frac{\sigma_t^2 + \sigma_b^2 + (\overline{t^2})(1-S)^2}{2\,\bar{t}\,(1-S)} - \frac{\overline{I^2}}{2\bar{I}} + 1 \qquad (3.9)$$

where I is the idle time, and for fixed packet sizes $\sigma_b = 0$.

The idle times distribution depends on how the previous busy period ended. For arbitrary interarrival times distributions, the idle times are very hard, if not impossible, to calculate. So the idle times pdf's presented in the following sections were obtained by simulation (Appendix B).

From (2.14c), transmission time of the message packet is:

$$T_m = \frac{(1-\Theta)/2}{\Theta} \cdot \frac{1}{\eta} \qquad (3.10)$$

3.5 Analysis of Performance for G/D/1 Queue

3.5.a $\eta = .01$

For SRMA/1/CSMA-CD/.01, $\theta(opt) = .97$, and T_m using (3.10) is:

$$T_m = 1.55$$

Using the interarrivals pdf of (3.2), a simulation study was carried out to determine the idle times distribution. The idle times pdf appears in Fig. 3.3. Instead of showing the whole histogram, we only show the pdf at the midpoints of histogram. Exponential density functions with mean $1/S_r$ are superimposed. For small S_r, the exponential assumption seems to be a good

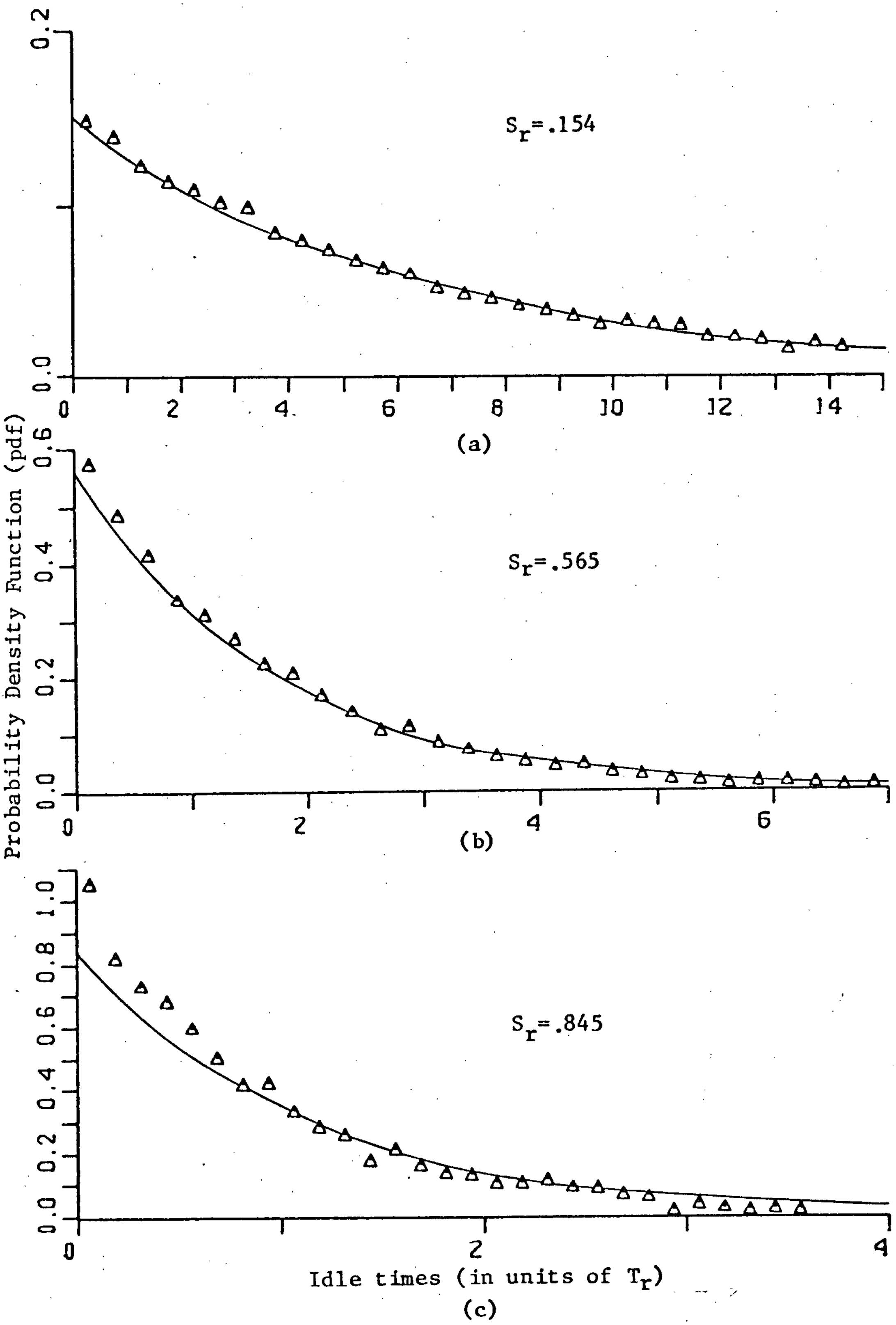


Fig. 3.3 The probability density function of the idle times for T_m/T_r=1.55 for (a) S_r=.154 (b) S_r=.565 (c) S_r=.845 Exponential pdf's with means of $1/S_r$ are superimposed.

choice. For larger S_r (Fig. 3.3c), there is a slight difference between the two curves for small idle times. In general, the exponential pdf appears to fit the experimental curves well.

Since I is exponentially distributed (using the same T_r units):

$$\bar{I} = T_m / S^2 \qquad (3.11a)$$

$$\bar{I^2} = 2T_m^{\ 2}/S^2 \qquad (3.11b)$$

All that remains in the calculation of D is S which is the throughput of the G/D/1 queue. Using the definition of S:

$$S = \bar{X} / \bar{t} = T_m / \bar{t} \qquad (3.12)$$

where $\bar{X}$ is the packet size. Substituting for S from (3.12), and using the idle times moments ((3.11a) and (3.11b)), and the interarrival times moments ((3.5) and (3.7)), the total queueing delay of (3.9) can be calculated. The G/D/1 delay is shown in Fig. 3.4. The delay curve of M/D/1 system, which corresponds to the original exponential assumption, is superimposed.

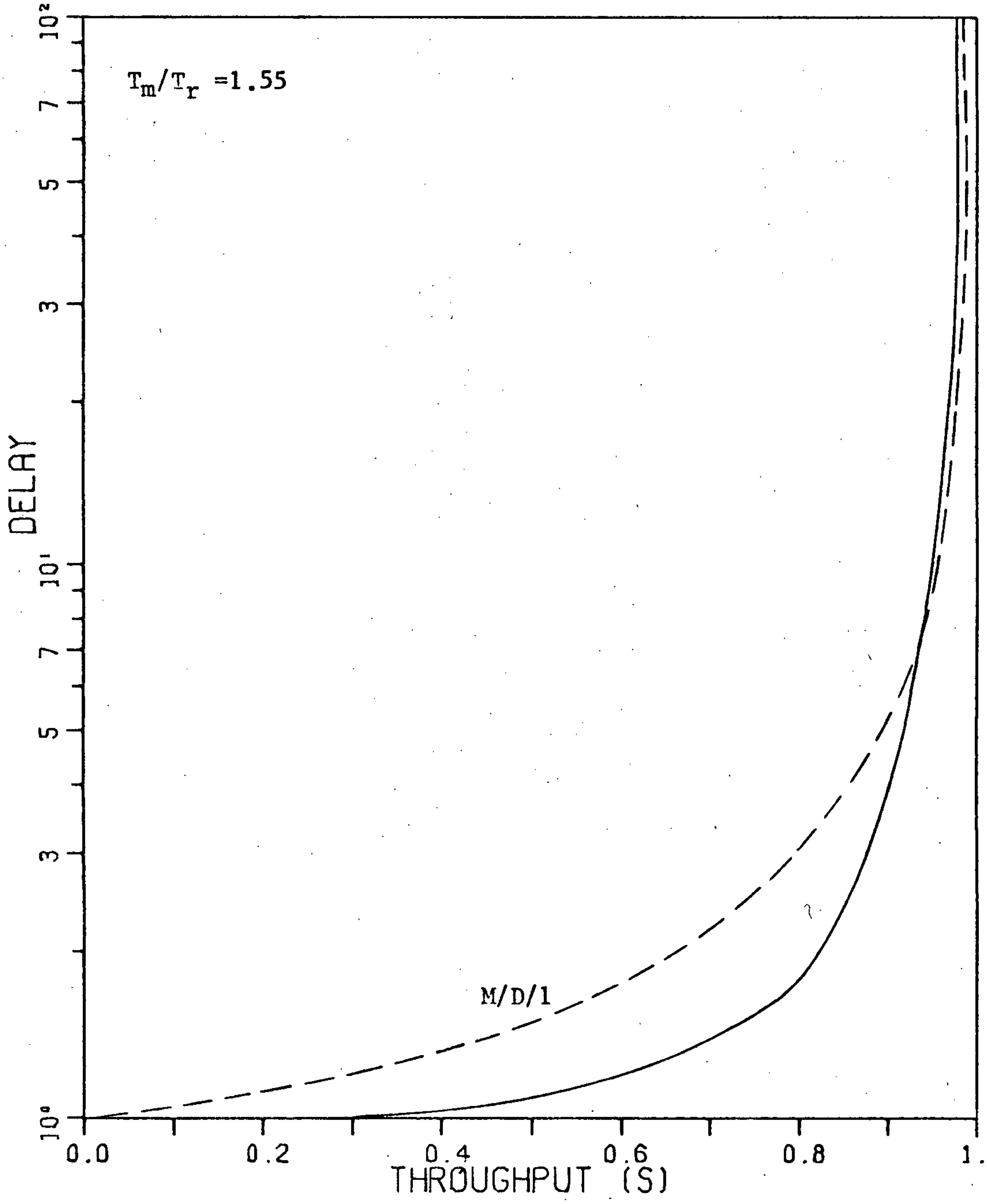


Fig. 3.4 The queueing delay vs. throughput for T_m/T_r=1.55. The delay is in units of packet transmission times. The dashed line is the M/D/1 queueing delay produced for comparison.

3.5.b $\eta=.1$

The idle times pdf for $\eta=.1$ is shown in Fig. 3.5. Although after some tedious work, the idle times distribution can be modeled, we use a simpler procedure to approximate the delay.

For $\eta=.1$, $\theta(opt)=.83$; therefore using (3.10):

$$T_m = 1.02$$

Waiting delay happens if two packets arrive less than 1.02 units apart and:

$$p(\tilde{t} \leq 1.02) = \int_{1}^{1.02} \{\lambda e^{-\lambda t} + 2(1-e^{-\lambda})\}\, dt \tag{3.14}$$

$$= .04 + .96e^{-\lambda} - e^{-1.02\lambda}$$

The maximum probability of two packets arriving less than 1.02 units apart is .03 ((3.14) is maximized for $\lambda=1$, $0\leq\lambda\leq1$ packets/T_r). So for all practical purposes, the waiting delay can be assumed to be zero, and the delay in this case, is only due to the transmission time of the message packet. Simulation studies support this conclusion.

The maximum throughput of the system is found by evaluating (3.5) at $\lambda=1$, and then by using (3.12):

$$S(max)=.67$$

3.6 Performance Comparison

The delay vs. throughput curves for the new assumption along with the exponential assumption are presented in Fig. 3.6. It is clear from the figure that for $\eta=.1$, the

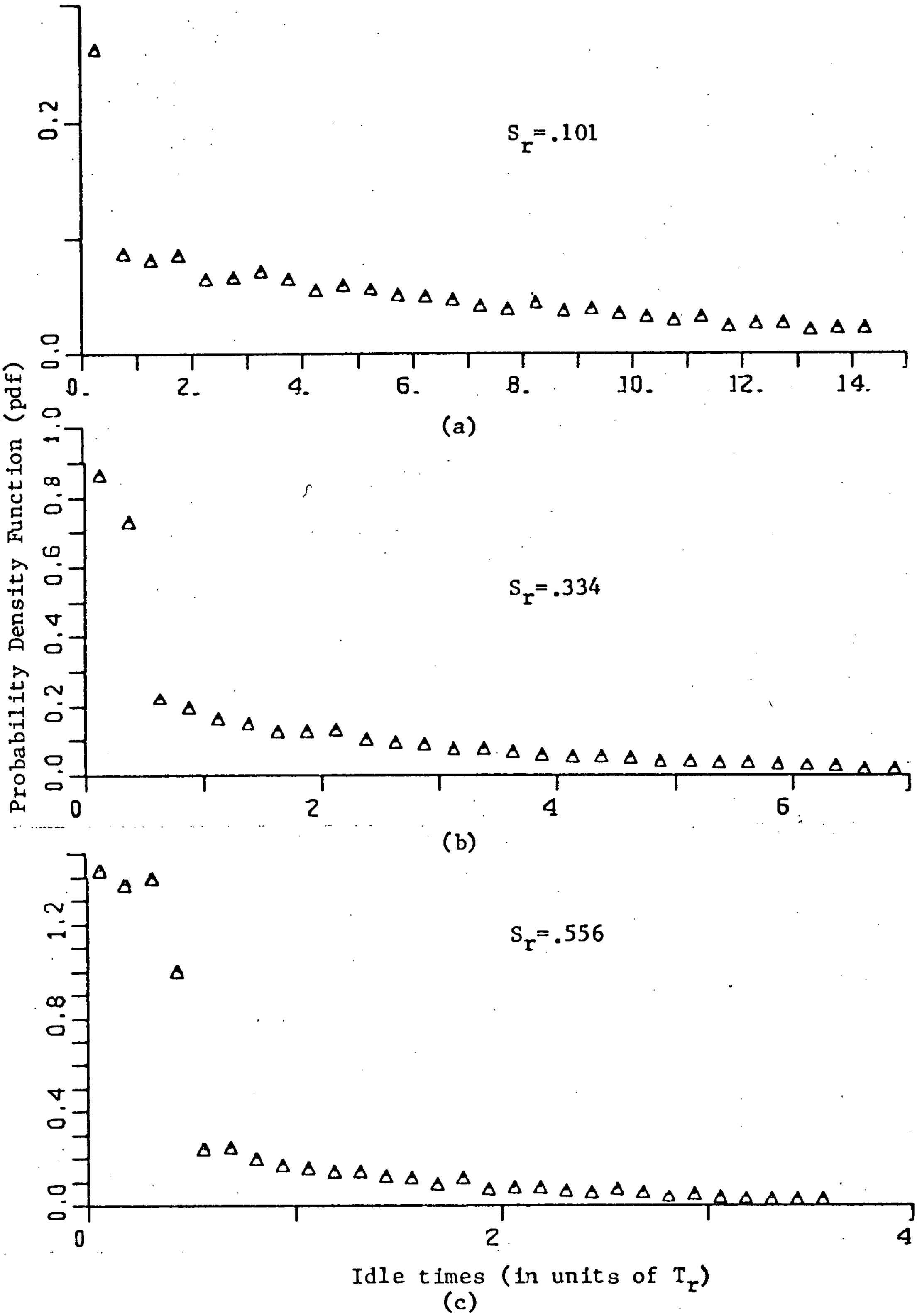


Fig. 3.5 The probability density function of the idle times for T_m/T_r=1.02 for (a) S_r=.101 (b) S_r=.334 (c) S_r=.556

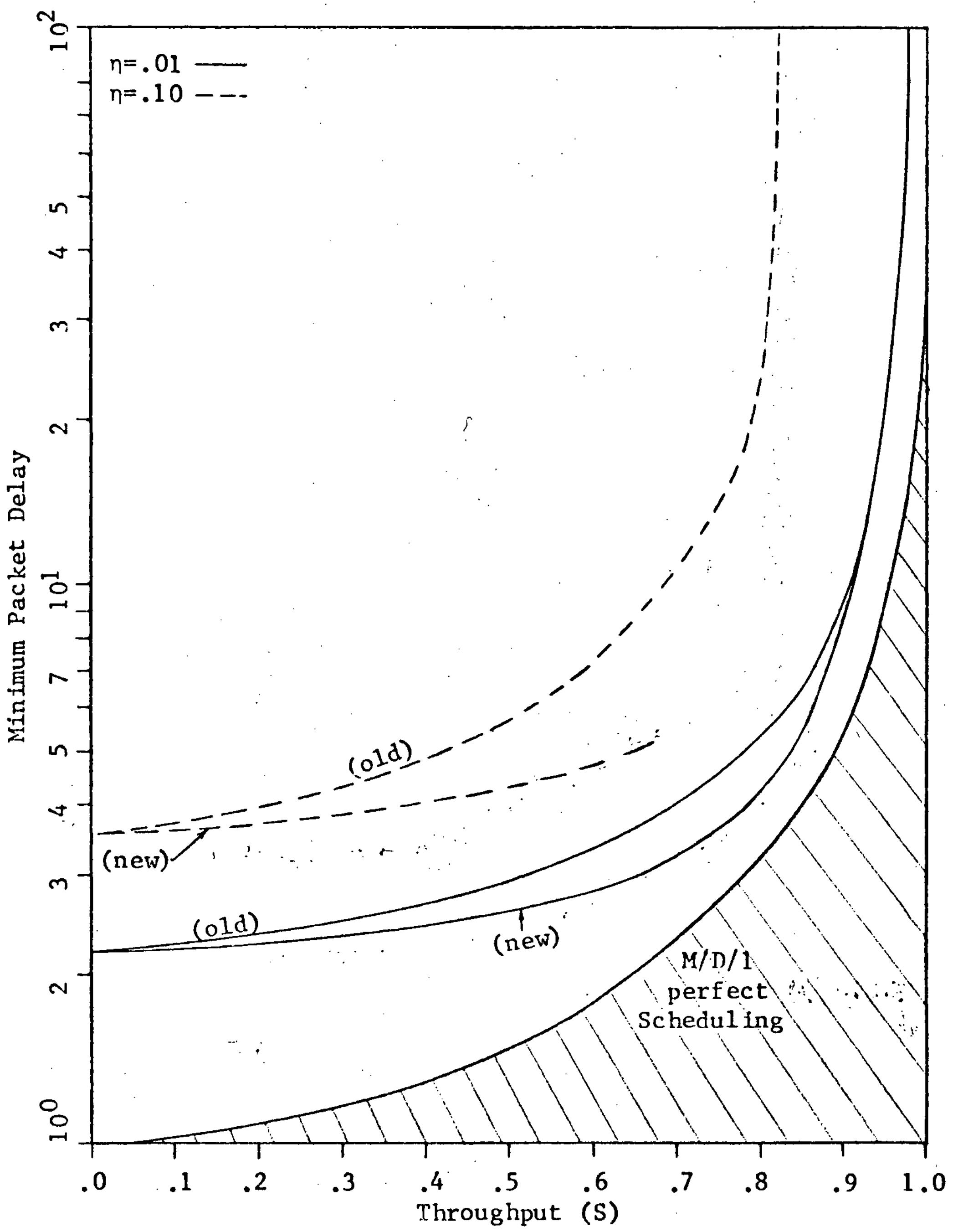


Fig. 3.6 SRMA/1 (new): Packet delay in units of T_w vs. throughput. The delay curves of Poisson assumption are superimposed for comparison (old).

effect of the new interarrivals assumption on the overall delay is negligible.

For the case of $\eta=.1$, the delay, instead of going to infinity for heavier loads, rises to a maximum value of about 5 units (of T_{ω}). The tradeoff is in the maximum capacity that the system can achieve; as shown before, S(max)=.67 which is smaller than S(max)=.79 achieveable under the exponential arrivals assumption.

In summary, the error in delay, due to the exponential interarrivals assumption, increases with increasing η. In comparison to previous delay curves, the corrected curves show smaller delays and reduced capacities.

IV MCMA/m PROTOCOL

4.1 Introduction

In the previous two chapters, the performance of the SRMA/m protocol was analyzed. The protocol employs aspects of both central and distributed control to optimize the performance. A good performance is achieved at the expense of system hardware complexity since there are two completely different control schemes to deal with.

In this chapter we propose a new protocol called Multiple Channel Multiple Access with m sub-channels (MCMA/m), which is based entirely on a distributed control strategy, therefore avoiding the hardware complexity of SRMA/m.

4.2 Assumptions and Protocol Description

The channel is divided into a number of sub-channels of which one, the ACK sub-channel, is used exclusively for carrying acknowledgement traffic. The remaining sub-channels are used for carrying message packets. From now on, if no ambiguity arises, the term channel is used instead of sub-channel.

Under software control, the terminals have the capability of switching among channels. The synchronization delay depends on the system parameter values and the modulation technique used. For the purpose of this study, the synchronization delay is assumed negligible.

When the terminals are free to receive, they continuously scan the channels to detect packets addressed to them. Since there is no call establishment, a problem arises: when a scanning terminal is monitoring a channel and a packet addressed to the terminal arrives on another channel, the packet is lost. To solve the problem, the address is repeated m times in the packet header; thus a scanning terminal has a 100% chance of properly decoding one of the m addresses in the header. The overhead due to this strategy, if message packets are long, and if m and the addresses are small (i.e. number of terminals small), can be assumed negligible.

Each terminal is equipped with a receive and a transmit timer. These timers keep track of the time spent on the ACK channel. When timeout on the ACK channel expires, they inform the terminal, and the terminal disengages from the ACK channel.

Since in a multi-channel system, message transmissions can finish in such a way that ACK packets are initiated at overlapping times, the ACK channel is operated in CSMA-CD.

It is assumed that the population of terminals is large enough to form a Poisson message source, independent of the delay due to the ACK packets. This assumption enables us to analyze the delay vs. throughput performance of the message packet, independent of the ACK packet.

The flowchart diagram of the protocol for both transmit and receive sequences appears in Fig. 4.1. In order to describe the protocol, the possible operation modes are presented in the following sub-sections.

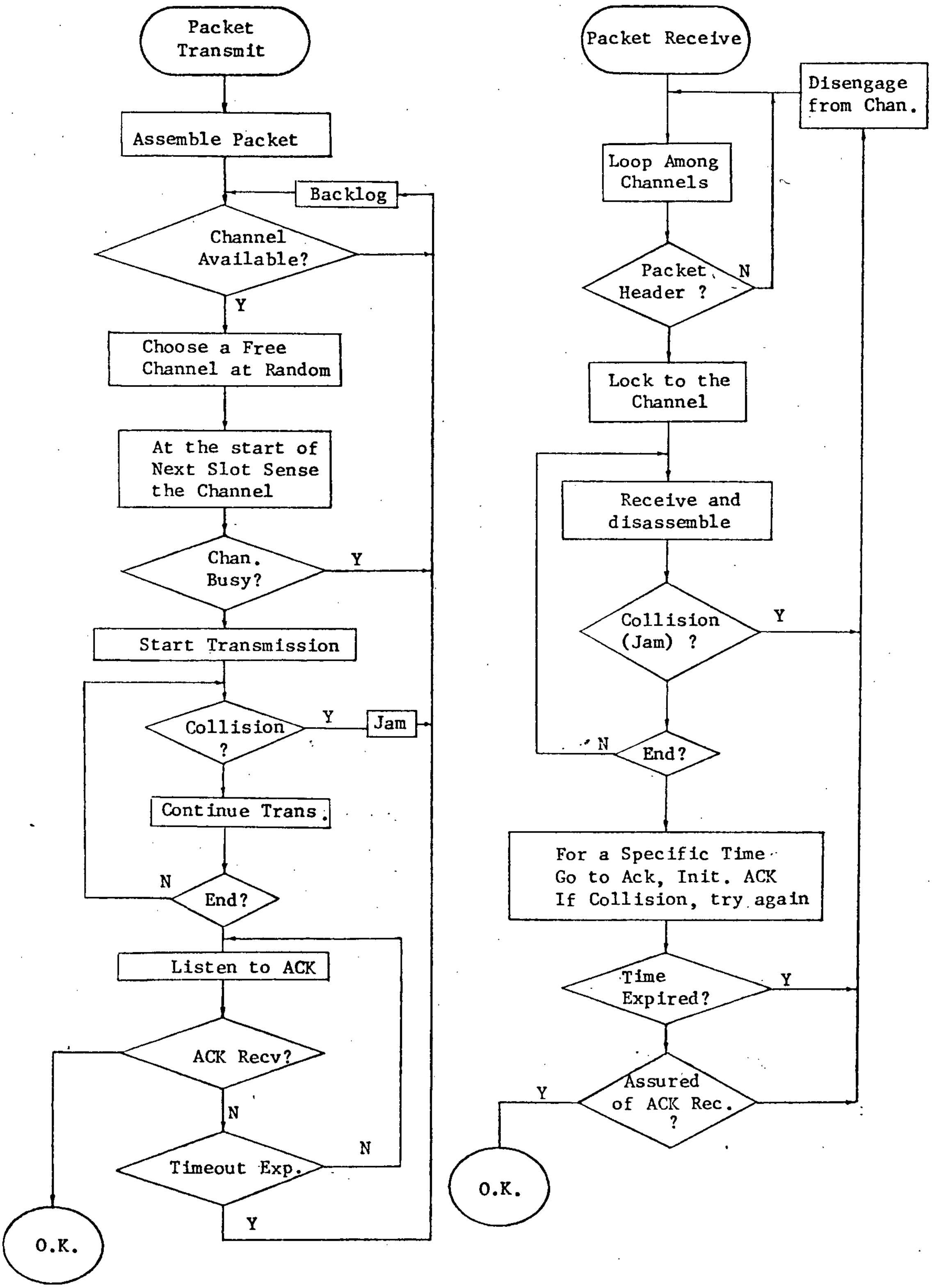


Fig. 4.1 MCMA/m: TRANSMIT and RECEIVE sequences

4.2.a Transmission without Collision

The terminals keep a running inventory of free channels. When a message packet is generated and no message channel is available, the packet is backlogged (i.e. scheduled for later retransmission). If a free channel is available, at the start of the next slot, the terminal again senses the channel; if the channel is free, transmission is initiated, otherwise the message packet is backlogged.

The destination terminal (assumed not busy in this sub-section) is scanning all of the message channels and if a proper address header is detected on a channel, the terminal starts buffering the message packet. When packet reception is completed, the transmitter and receiver reset their timers and switch to the ACK channel. The destination terminal sends an ACK (in CSMA-CD fashion) which is received by the source device and the terminals disengage from the ACK channel. The timer is pre-set to a finite time. In case the ACK packet does not get through because of excessive collisions on the ACK channel, the source and destination terminals disengage from the ACK channel after the timeout, and the message packet is backlogged.

4.2.b Transmission with Collision

If two or more message packets start transmission on the same channel within an end-to-end propagation delay period, the packets collide and destroy one another. After this critical period, all terminals detect carriers due to packet transmissions, and refrain from packet initiation. The source terminals send jam signals on the same channel to inform the destination terminal(s) (and others) of the collision. The destination terminal(s) (if already locked to that message channel) disengage immediately. The source terminals, in turn, reschedule the packets for some later time.

4.2.c Transmission with Non-scanning Destination Terminal

Message transmission is initiated but the destination terminal is not scanning the channel because it has either failed, is servicing another packet (on another channel), or is locked to the ACK channel. After completing the transmission, the source switches to the ACK channel, resets its timer, and starts listening for ACK. Since the source does not get an answer, after the timeout period, the packet is backlogged.

Since missed packets run to completion without truncation, (as opposed to collided packets that are truncated), they have the potential of heavily degrading the throughput. The probability of missed packets in a multi-channel system was discussed in Chapter 2.

4.3 Terminology and Notation

We introduce the following notation:

slot — end-to-end propagation delay of a sub-channel ($slot_m$ and $slot_a$ correspond to the message and ACK sub-channels respectively)

ν — retransmission coefficient as defined in Chapter 2, but pertaining to the message sub-channels

b_m number of bits in message packet

b_a number of bits in ACK packet

W_m bandwidth of the entire message channel

W_a bandwidth of the ACK channel

η ratio of the size of ACK packet to the size of message packet; $\eta = b_a/b_m$

L_m size of message packet in message slots[1]

L_a size of ACK packet in ACK slots

D_m delay (wait+transmission) of the message packet

D_a delay (wait+transmission) of the ACK packet

T_m transmission time of the message packet on the entire message channel (so mT_m is the transmission time on a sub-channel); $T_m = b_m/W_m$

T_a transmission time of the ACK packet on the ACK channel; $T_a = b_a/W_a$

In addition θ, S, and S_m have the same definitions as in Chapter 2.

4.4 Performance of MCMA/1 vs. Number of Terminals

In the MCMA/1 protocol, no ACK channel is necessary since the message packets cannot be lost; therefore, the performance of MCMA/1 directly corresponds to that of CSMA-CD, and the

---

[1] Note that L_m is independent of m since splitting the available bandwidth increases the end-to-end propagation delay and the packet transmission delay by the same factor.

simulation program used in the study of CSMA-CD can also be used in the study of MCMA/1.

The optimum value of ν for CSMA-CD with M=50 terminals was found to be ν=.05 [TOBA 80]. As M changes, the optimum value of ν also changes. For larger M, ν is selected smaller to prevent excessive collisions in heavy-load situations. We postpone the study of ν and choose ν=.05 as a compromise.

Fig. 4.2 shows the delay vs. throughput curves of MCMA/1 for different numbers of terminals.[1] The CSMA-CD curve is superimposed. Comparing the M=50 and the CSMA-CD curves, we observe that for light loads, the delays are equal; for higher loads, there is a small difference between the two curves. We take the closeness of the two curves as an indication of the correctness of the simulation program.

Comparing the performance curves for different values of m, we see that as m increases, the delay of the system as a function of throughput also increases. This effect is apparent over the entire range of loads.

The delay vs. throughput curves, however, do not show the rate of change of throughput with respect to changes in load. For that purpose, we plot the throughput vs. G curves in Fig. 4.3.[2] The curves corresponding to M=50 and M=100 appear to

---

[1] In this section and the next one, the delay is defined as the total delay (wait+transmission) of the message packet over the message bandwidth. In the same manner, throughput is the throughput over the message bandwidth.

[2] We define G as the total number of packet transmissions attempted in L_m slots (whether or not actual transmission occurs).

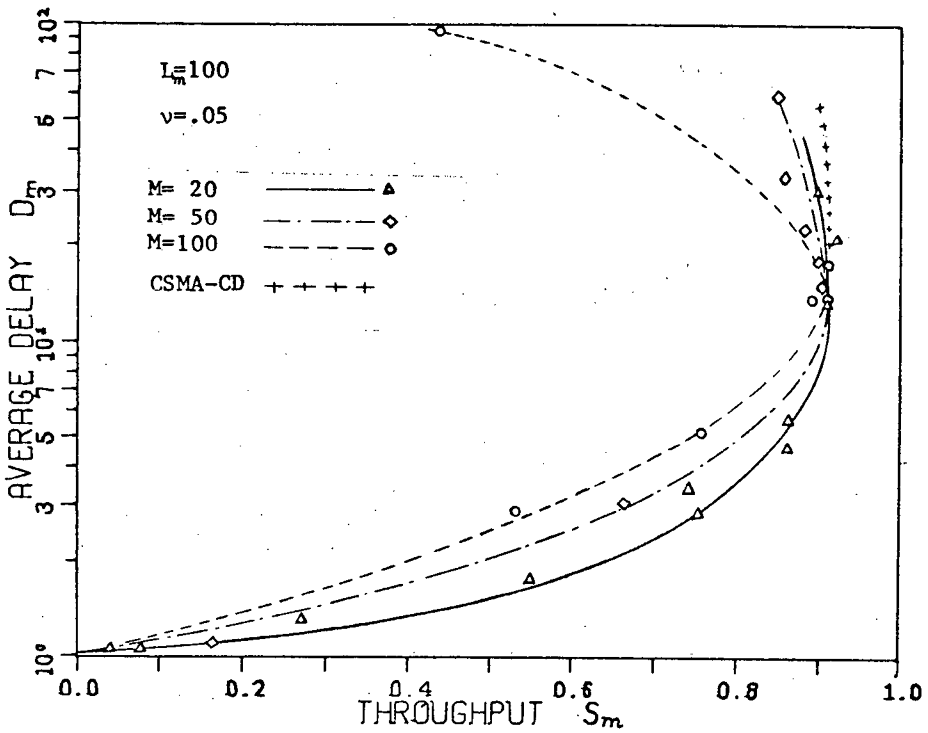


Fig. 4.2 MCMA/1: Delay vs. throughput for different values of M

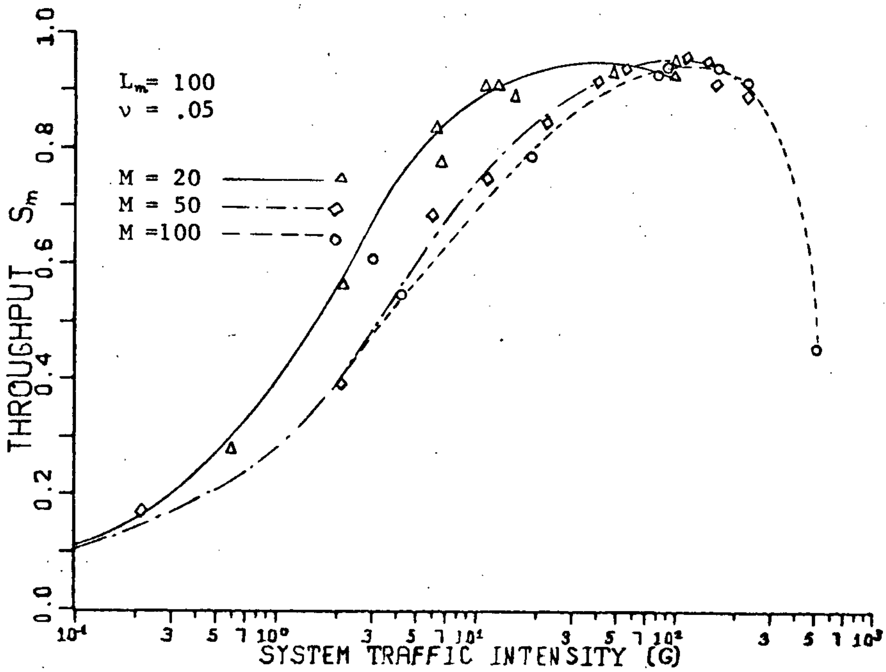


Fig. 4.3 MCMA/1: Throughput vs. G for different values of M

be quite close, with the exception of high loads where the throughput of the M=100 curve rapidly deteriorates. The deterioration is due to the chosen value of ν. For a fixed ν (ν=.05), increasing the population size causes a progressively larger throughput degradation.

The relative "flatness" of the curves at the top is a measure of the sensitivity of the optimum performance to changes in load condition. The M=20 curve is flatter at its top than the other two, so a system with 20 terminals stays at peak performance for a wider range of loads. For purposes of comparison with the previous results, we choose M=50.

4.5 Comparative Study of MCMA/1 and MCMA/m (m>1)

The simulation of MCMA/m (m>1) is only performed for the message packets (and not the ACK packets); so the simulated system loosely corresponds to a m-channel CSMA-CD. If $D_a(S_a)$ is independent of m and ν, then in comparing MCMA delays for different values of m and ν, the ACK delay can be ignored. Later on, it is shown that the arrival of ACK packets to the ACK channel can always be assumed to form a Poisson source (independent of m), therefore D_a is not a function of m. It is further assumed that the delay of ACK packets is independent of ν.

We begin the study of MCMA/m (m>1) by examining the performance as a function of ν. The delay vs. throughput curves for m=4 and different values of ν are shown in Fig. 4.4. Small values of ν result in good throughputs, and the throughputs do not degrade substantially for the heavy-load tails of the curves. However, a tradeoff is made in the delay at capacity. Increasing ν decreases the delay at capacity, until the capacity itself starts to degrade. A choice has to be made among different values of ν. For a reasonable overall performance, ν=.05 seems to be a good choice. Note that this is the same value of ν chosen in CSMA-CD, so we conclude that the optimum value of ν is independent of m.

In order to investigate the changes in the delay as a function of ν, the delay vs. ν curves for specific throughput values are plotted in Fig. 4.5. The corresponding CSMA-CD curves are superimposed. The rate of change of the two delay curves appears to be quite similar. We conclude that the effect of ν on the performance of the multi-channel system is not noticeably different from that of the single-channel system.

A shortcoming of the protocol is the problem of lost packets. In MCMA/m (m>1), packets are lost when two or more packets are addressed to the same terminal at overlapping times. In this case, no more than one packet is properly received (in the absence of collisions). For the system parameters under consideration (m=4, M=50), this effect is significant. In order to show the resulting performance degradation, the delay vs. throughput curves of MCMA/m are plotted in Fig. 4.6 (for m=1,4,8). In Fig. 4.7, we plot the maximum capacity (over the entire range of ν) as a function of

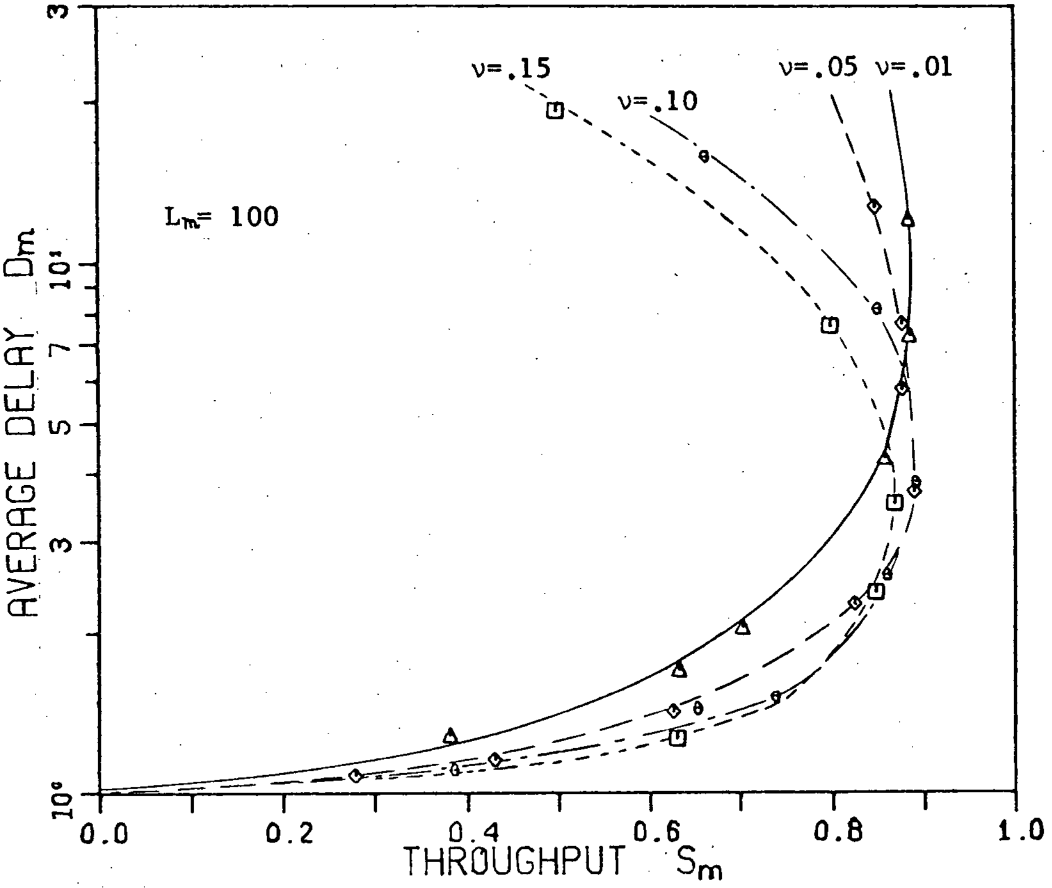


Fig. 4.4 MCMA/4: Delay (in units of mT_m) vs. throughput (S_m) for different values of ν

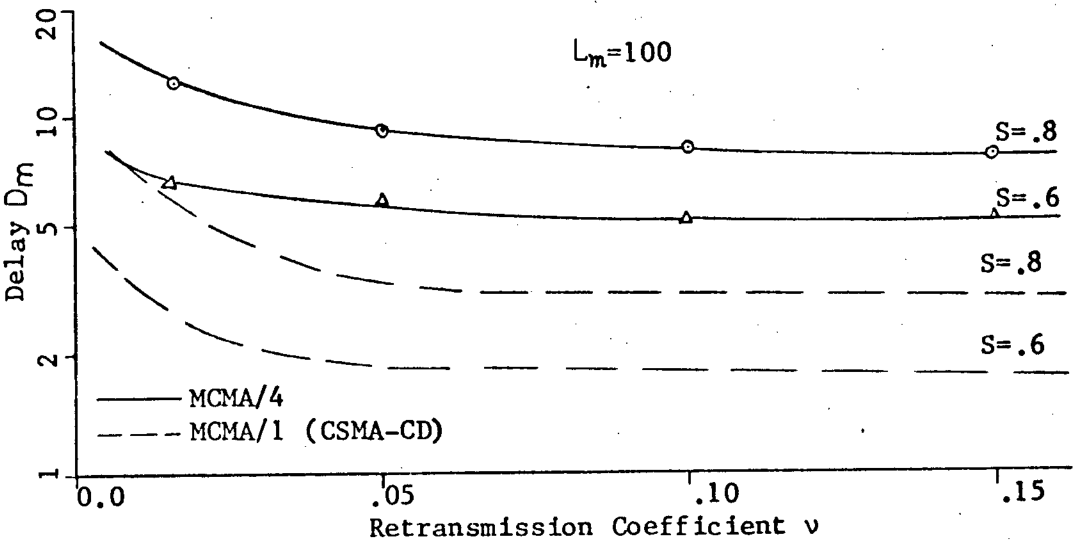


Fig. 4.5 Delay (in units of T_m) vs. ν for CSMA-CD and MCMA/4

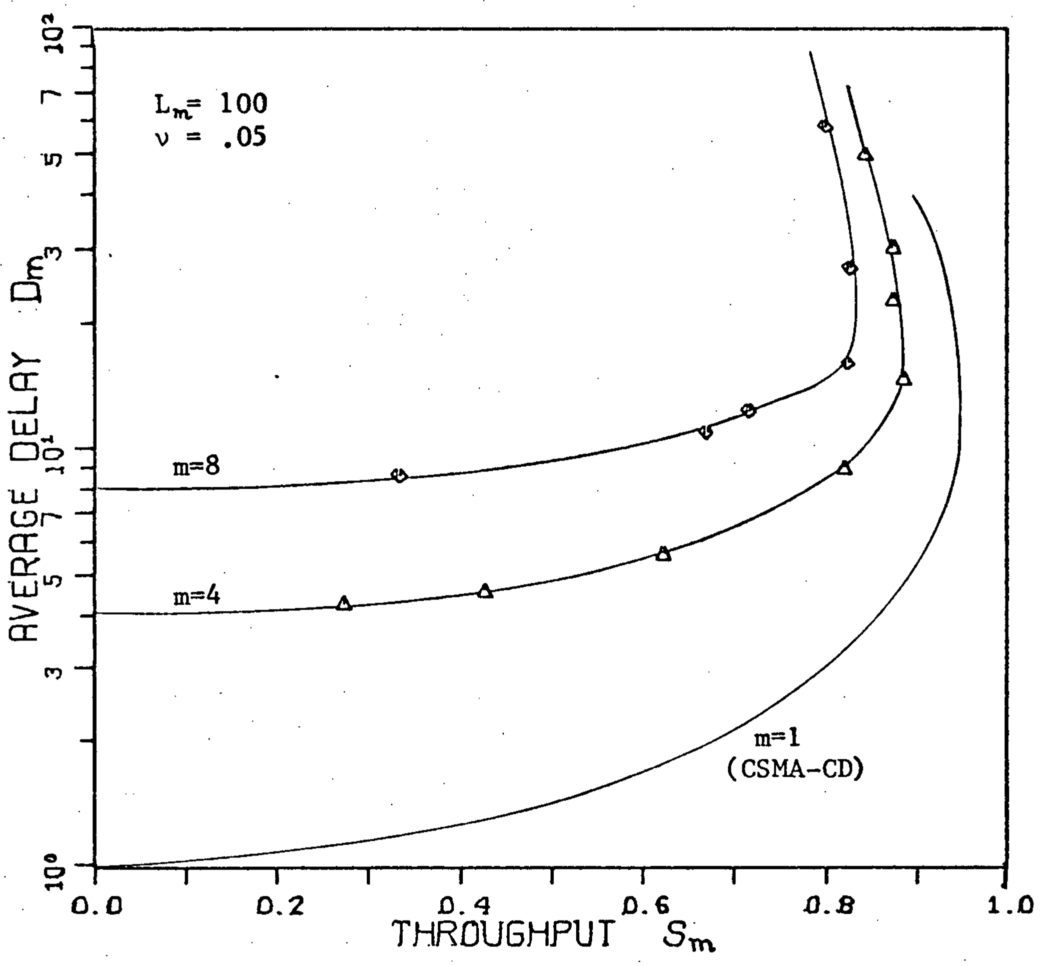


Fig. 4.6 MCMA/m: Delay (in units of T_m) vs. throughput for m=1,4,8

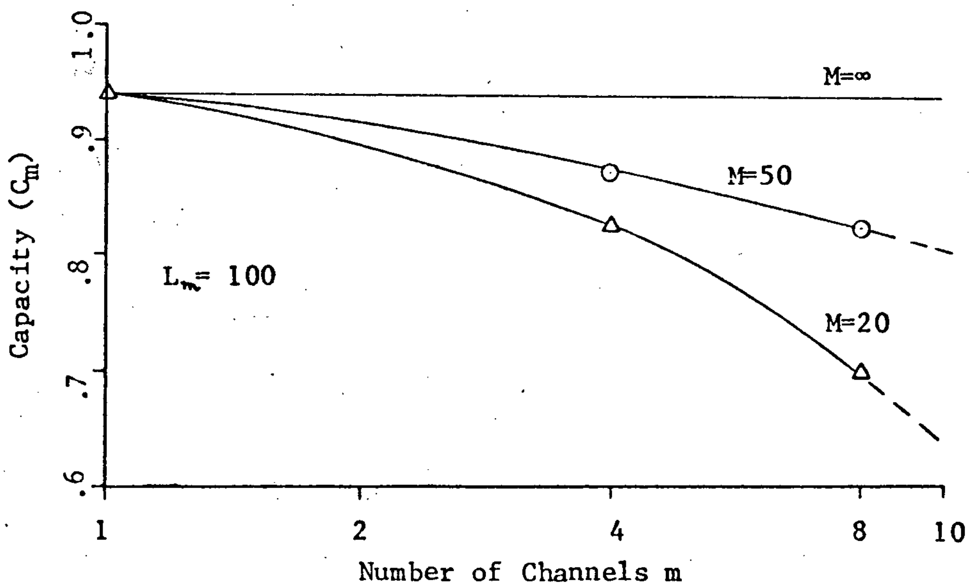


Fig. 4.7 Capacity vs. number of channels for different terminal populations (M)

m for different values of M.[1] From Fig. 4.7, the capacity degradation due to lost packets increases with increasing m. The rate of capacity degradation as a function of m depends on M. For small values of M, the effect is rather significant.

The performance of the system is optimum at capacity (i.e. offering the optimum tradeoff between delay and throughput). In order to compare the optimum performance of MCMA/1 and MCMA/m (m>1), the capacity as a function of ν for m=1,4 is plotted in Fig. 4.8. The capacity difference between MCMA/1 and MCMA/4 is greatest for ν>.1. In this region, the number of collisions per channel is larger for MCMA/1 (since there is only one channel); as a result, the capacity of MCMA/1 degrades at a faster rate.

4.6 Analysis of the Total Delay

4.6.a Components of the Delay

Until this point, our attention has been focused on the delay of the message packet on the message channel. In this section we analyze the total delay on the entire system bandwidth.

The ACK packet carries the same kind of information that the request and AR packets carry in SRMA, namely addresses, timing, and error detection and correction information. In order to make the comparison with SRMA possible, we assume L_a=100 (which corresponds to L=100 for request and AR packets

[1] The delay vs. throughput curves of M=20 appear in Appendix D.

in SRMA).

The delay of the system consists of D_m and D_a. The total delay can be written as:

$$D = D_m(S_m) + D_a(S_a) \tag{4.1}$$

where S_a is the throughput of the ACK channel, and D_m and D_a are in units of T_m and T_a respectively.

4.6.b Delay of the Message Packet

The delay of the message packet depends on the value of L_m, the size of the message packet in units of message slots. First we find a relationship between L_m and L_a, and then we use L_a=100 to determine L_m.

Slot is defined as the end-to-end propagation delay of a sub-channel:

$$\text{slot} = X/W \quad \text{(seconds)} \tag{4.2}$$

where W is the channel bandwidth (in bits/sec.), and X is the "cable length" (in bits), where cable length denotes the maximum number of bits that can be in transmit on the cable at any instant.

By definition:

$$L = T/\text{slot} \tag{4.3}$$

where T is the packet transmission time.

Using (4.2)-(4.3) and definitions of T_m and T_r:

$$\frac{L_m}{L_a} = \frac{T_m}{T_r} \cdot \frac{\text{slot}_r}{\text{slot}_m} = \frac{b_m}{b_r} \cdot \frac{W_r}{W_m} \cdot \frac{X}{X} \cdot \frac{W_m}{W_r} \tag{4.4}$$

We can now write:

$$L_m = L_a/\eta \tag{4.5}$$

From (4.5) for $\eta=.01$, $L_m=10000$ and for $\eta=.1$, $L_m=1000$.

Simulation of MCMA/m, for such large values of L_m, is very costly. Therefore we use an approximation to the delay vs. throughput curves.

Assumption: For a constant m, the message packet delay $D_m(S_m)$ in units of T_m, is independent of L_m (i.e. follows a fixed trajectory).

In order to show this property, in Fig. 4.10 we plot $D_m(S_m)$ curves for m=4 and several values of L_m. For throughputs less than capacity (i.e. the range of interest), the delay curves are very close. Therefore a fixed trajectory extended to $S_m=1$ can represent the curve for all values of L_m.

In summary, the delay D_m (in units of T_m) for different values of S_m, can be read from the trajectory.

4.6.c Delay of the ACK Packet

The analysis of D_a is relatively easy. The arrival of ACK packets to the ACK channel has the same distribution as the departure of the message packets from the message channel. It was shown in Fig. 2.2 that the output of CSMA-CD system can be modeled as exponential (with the exception of 0 to 1 unit interval).

Since in a multi-channel system, packets can depart less than one packet transmission time apart, we expect the assumption to improve (i.e. also fit the pdf of the 0 to 1 unit interval). Fig. 4.9 shows the interdeparture times pdf for m=4. The exponential assumption seems to fit the experimental data very well. Although the interdeparture times pdf varies with L_m, we assume that for the range of L_m under study, the exponential assumption is valid.

4.7 Total Delay and Discussion

The total delay as a function of θ for m=4 is shown in Fig. 4.11a. The delay is based on the trajectory of Fig. 4.10. From the figure, for η=.1 θ(opt)=.97, and for η=.1 θ(opt)=.90. The contour lines for m=8 are plotted in Fig. 4.11b, which shows that the values of θ(opt) are about the same as in m=4.

The delay vs. throughput curves of MCMA/m are shown in Fig. 4.12. Note that MCMA/1 directly corresponds to CSMA-CD because no ACK channel is necessary for m=1.

In Fig. 4.12, there are two sources of analytical error for which adjustments should be made. The first is due to the trajectory being extended all the way to S_m=1. As discussed before, lost packets degrade the capacity of multi-channel systems. This degradation increases with increasing m. For the values of L_m under study, we assume that the percentage degradation as a function of m, is equal to that of L_m=100 (since the probability of lost packets is, almost, independent of L_m). The degradation (corresponding to L_m=100 as shown in Fig. 4.7) is applied to Fig. 4.12. The capacity correction in Fig. 4.12 is designated by the area called unattainable region.

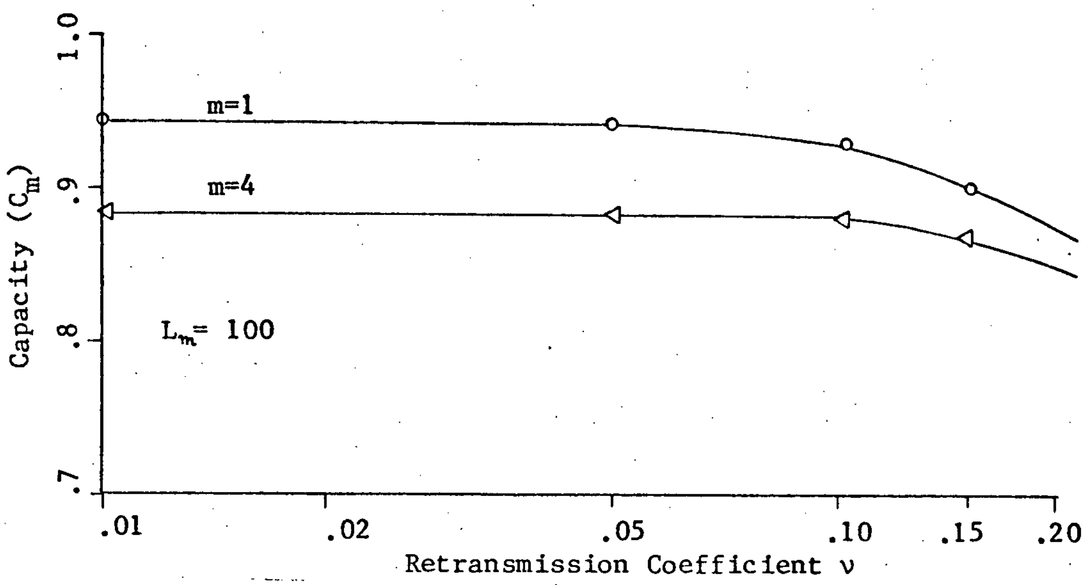


Fig. 4.8 Capacity vs. retransmission coefficient ν for MCMA/1 and MCMA/4

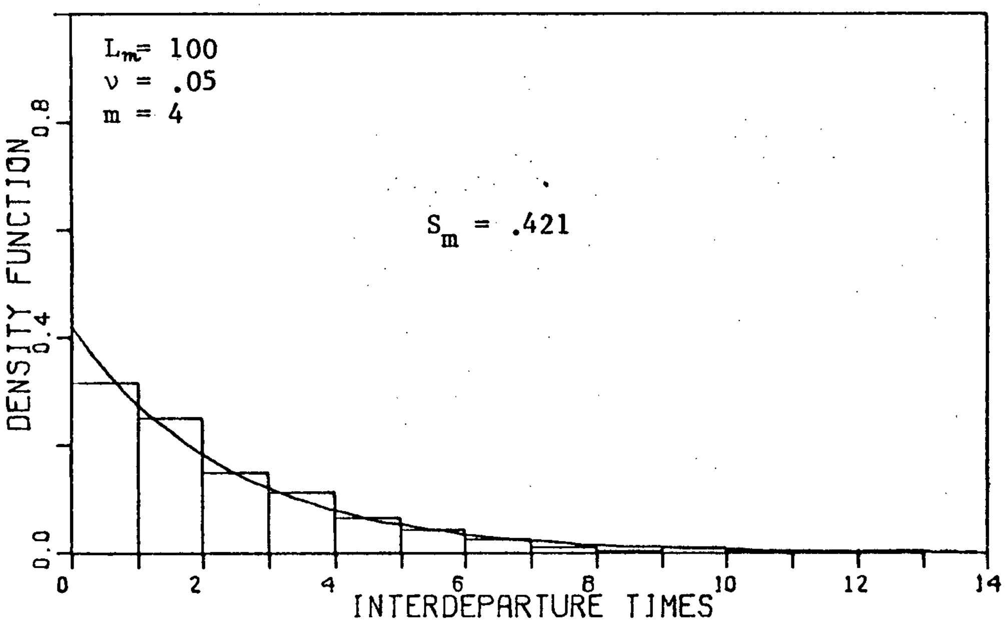


Fig. 4.9 Probability density function of the interdeparture times of the message packets (in units of mT_m)

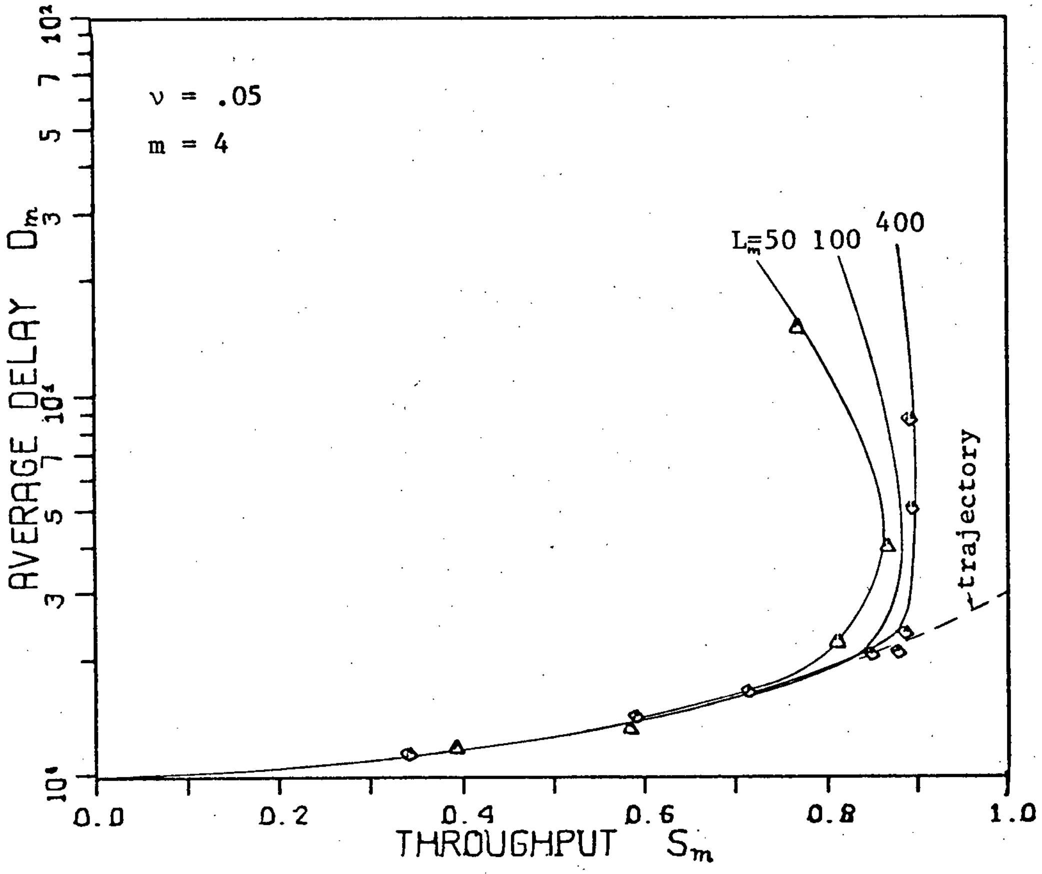


Fig. 4.10 Delay in units of mT_m vs. throughput S_m for different packet lengths

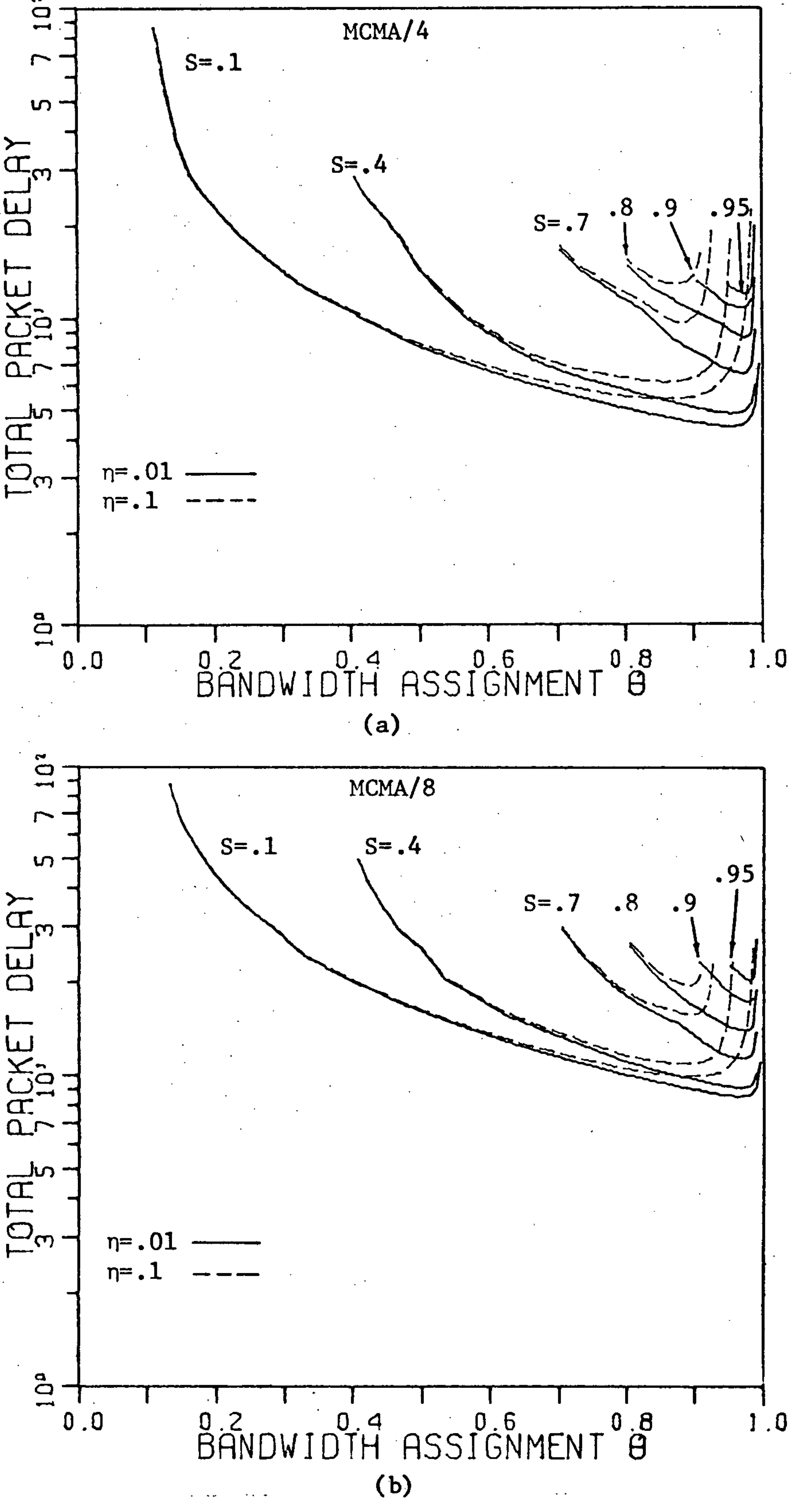


Fig. 4.11 Delay (in units of T_w) vs. bandwidth assignment Θ for (a) m=4 (b) m=8

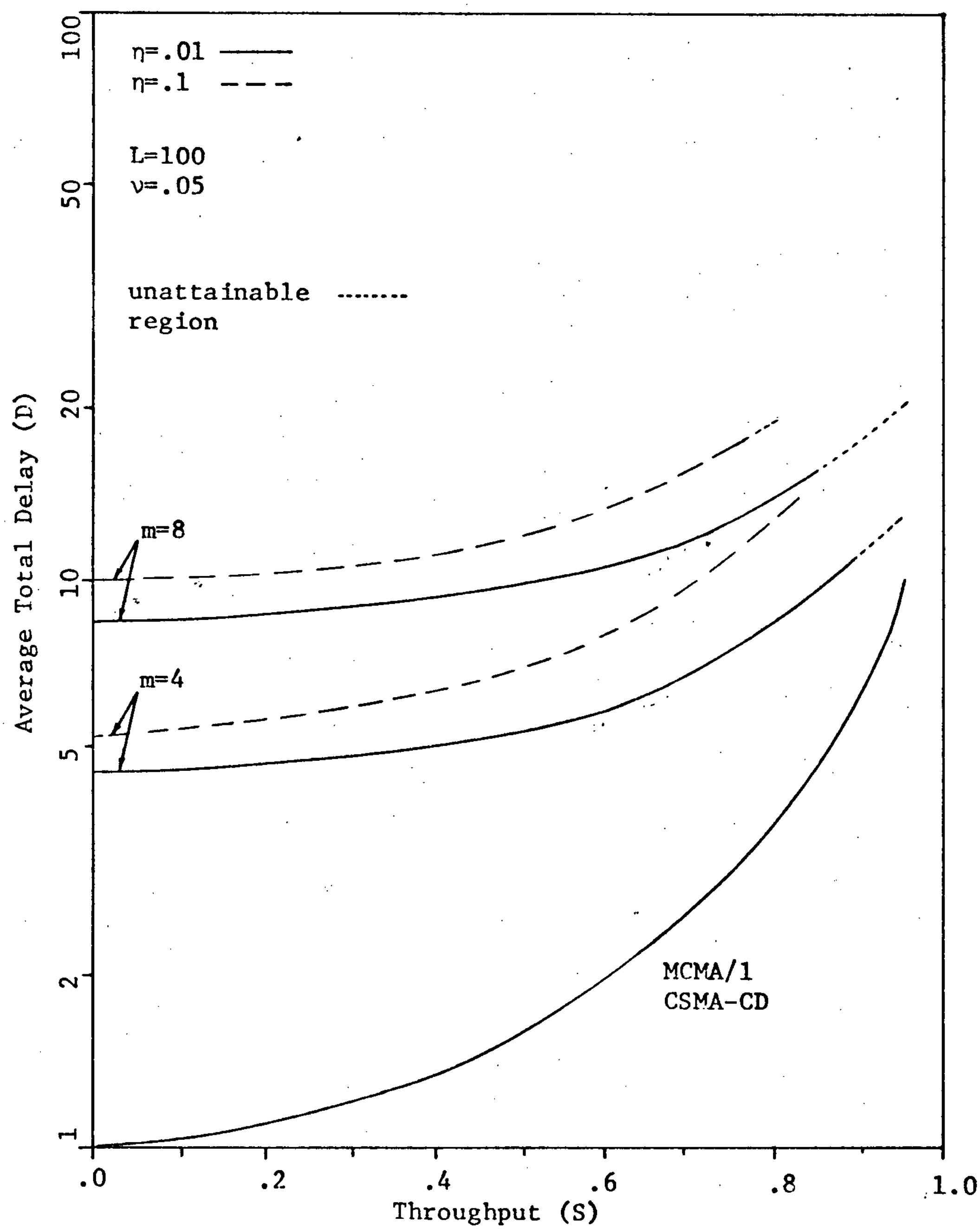


Fig. 4.12 MCMA/m: Average total delay (in units of T_w) vs. throughput

The second source of error is due to the (implied) assumption that the delay on the ACK channel is always equal to the delay of successful ACK packets. This assumption is correct, except when lost packets are involved. In that case, the delay is equal to the maximum value of the timer timeout delay. The choice of a timeout value depends on the load conditions, and the pdf of the CSMA-CD delay. Because of the extremely complicated nature of the analysis, the amount of error is left unspecified.

We defer the discussion of performance to the next chapter where the SRMA/m and MCMA/m systems are compared.

V CONCLUSION

5.1 Summary

In this thesis, the performance of two multi-channel multi-access protocols has been investigated. The first one, SRMA/m, utilizes distributed control for request scheduling and central control for message scheduling. The second protocol, MCMA/m, uses a totally distributed scheduling strategy.

Single-channel SRMA (SRMA/1) was proposed and analyzed by Tobagi and Kleinrock [TOBA 76]. In the analysis, the request channel was operated in CSMA. In Chapter 2, the analysis was extended to CSMA-CD. The distribution of interdeparture times of message packets was determined by simulation. It was found that except for the 0 to 1 unit interval, the interdeparture times pdf can be modeled as exponential. Using this assumption θ(opt), the percentage bandwidth allocated to the message channel, was found and the total delay was plotted. The changes in overall delay resulting from switching to CSMA-CD, was concluded to be minimal.

The SRMA/1 analysis was then extended to SRMA/m. Since SRMA/m involves M/D/m queueing, and exact solutions for M/D/m queue do not exist, the delay for various values of m was obtained by simulation. The simulation results were found to fall within the calculated theoretical bounds. In order to find θ(opt), delay contour lines for m=4 and m=8 were plotted. The value of θ(opt) was not significantly different from the m=1 case.

In Chapter 3, the error due to the assumption that the interdeparture times of request packets are exponentially distributed, was analyzed. A unified interdeparture model was developed. Since the analysis involved idle times, the idle times pdf was determined by simulation. The total delay, according to the values of θ found in Chapter 2, was calculated. For small values of η (request packet to message packet size ratio), the corrections made to the delay curves were quite small. For larger η, the correction was substantial and resulted in a maximum delay of 5 units (of message packet transmission time), instead of infinity.

The MCMA/m protocol was proposed and analyzed in Chapter 4. The Delay vs. throughput curves of MCMA/1 and MCMA/4 were compared for various values of ν (retransmission coefficient of the message packet). It was found that the effect of changes in ν on the delay curves of m=4, is very similar to that of m=1. The total delay of MCMA/m was then analyzed based on the same parameters used in the SRMA analysis. The optimum θ was calculated and the delay curves for m=1,4,8 were plotted. These curves were corrected for degradation resulting from lost packets.

5.2 Comparison of SRMA and MCMA Protocols

The delay vs. throughput curves for both protocols and for η=.01 and η=.1 appear in Figs. 5.1 and 5.2 respectively. We can see that no protocol consistently outperforms the other. The tradeoffs are similar to what Tobagi and Kleinrock faced in

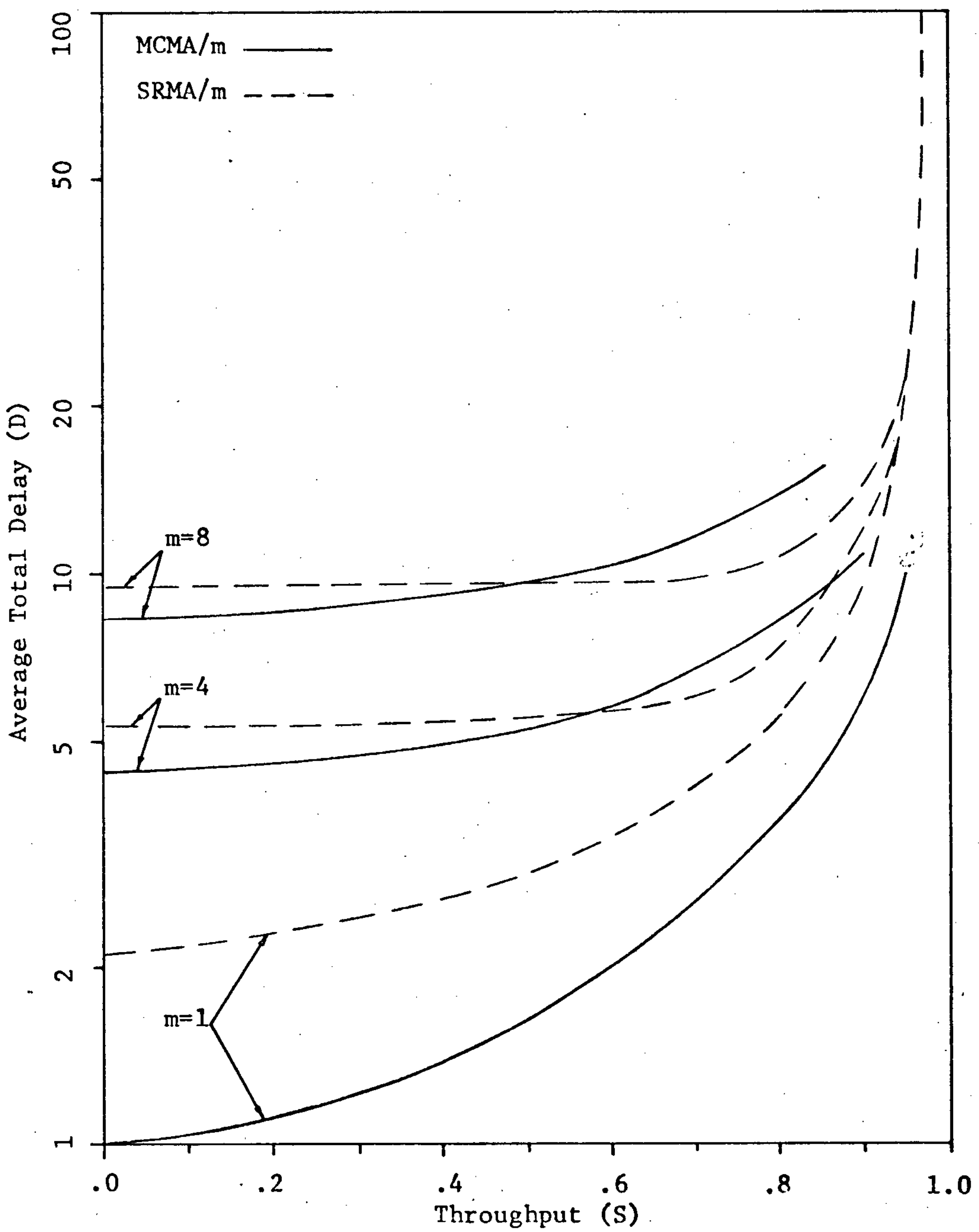


Fig. 5.1 Average total delay (in units of Tω) vs. throughput for SRMA/m and MCMA/m, η=.01

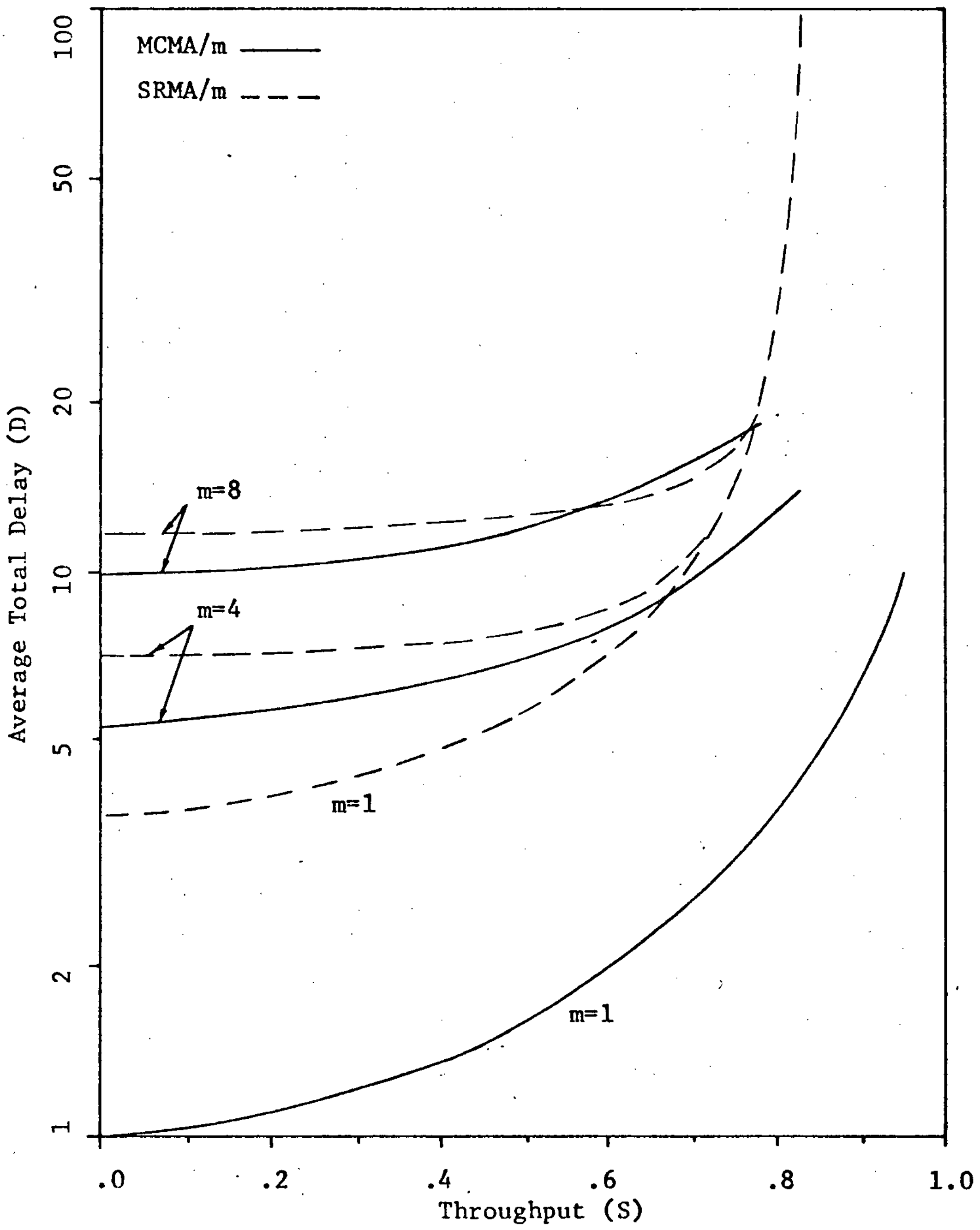


Fig. 5.2 Average total delay (in units of T_w) vs. throughput for SRMA/m and MCMA/m, $\eta=.1$

comparing random access protocols [TOBA 76]. They stated:

> ... It is to be noted that there exists no scheme which is consistently superior to others. The performance of each is dependent on several system parameters...; so also is the selection of the best scheme.

The important system parameters affecting the performance are m and η. First we consider m. For m=1, MCMA has a definite superiority over SRMA, expressed as smaller delays. As m increases, the difference in performance becomes dependent on load levels. For moderate to heavy loads, MCMA always achieves a smaller delay. For heavier loads (e.g. S>.6) and for small m's (e.g. m≤4), it is the MCMA system that shows a smaller delay; but as m increases, SRMA appears to have a progressively better performance.

Another measure of performance is the effect of overhead resulting from increases in η. The delay degradation (for m>1) does not seem to be very sensetive to η. On the other hand, the capacity degradation as a result of increasing η, appears to affect SRMA far more than it affects MCMA.

In some applications, the capacity (and not the delay) might be the crucial factor. In SRMA, the capacity is independent of m. In MCMA however, the capacity is a decreasing function of m, and for large m's, the capacity of MCMA shows severe degradation.

In summary, if we assume that a typical system is characterized by heavy loads and large number of channels, then SRMA outperforms MCMA in both delay and capacity. If because of moderate throughput/delay requirements, the choice between the two protocols is not influenced by performance, MCMA is the more suitable protocol because of its easier hardware implementation.

5.3 Suggestions for Further Work

We consider the case where short (interactive) and long (file transfer) packet types are involved. If each packet type has its own delay constraint (e.g. interactive packets requiring short delays), then it is conceivable to schedule each packet type for a different set of channels to meet the delay constraints. It remains to determine the optimum percentage bandwidth allocated to each channel set, so as to optimize the performance of each packet type. The analysis of optimum bandwidth assignment is complicated by the fact that specific mixes of packet types require different bandwidth allocations.

One application of multi-channel protocols is fail-safe communication. If a channel malfunctions, after a number of tries, the controllers erase the channel from their channel look-up table and resume transmission on the remaining channels. The task at hand is the design of higher-level protocols stressing reliability rather than performance.

REFERENCES

AGRA 77 A.K. Agrawala, R.M. Bryant, and J.A. Agre , "Analysis of an Ethernet-like Protocol", Proceedings of Computer Communication Symposium, pp. 104-111, 1977.

BENJ 70 J.R. Benjamin and C.A. Cornell, "Probability, statistics, and Decision for Civil Engineers", McGraw Hill, New York, 1970.

BIBA 79 K.J. Biba and J.W. Yeh, "Fordnet: A Front-end Approach to Local Communication Networks", Proceedings of LACN Symposium, pp. 199-213, May 1979.

BRUM 71 S.L. Brummel, "Some Inequalities for Parallel Server Queues", Operations Research, 19, pp. 402-413, 1971.

CLAR 78 D.D. Clark, K.T. Pogran, and D.P. Reed, "An Introduction to Local Area Networks", Proceedings of IEEE, Vol. 66, No. 11, pp. 1497-1517, November 1978.

CRC 78 -"CRC Standard Mathematical Tables", 25th Edition, CRC Press, Florida, 1978.

ETHE 80 -"The Ethernet, a Local Network, Data Link Layer & Physical Layer Specifications", Version 1.0, Intel Corporation, California, September 1980.

KLEI 75a L. Kleinrock and F.A. Tobagi, "Packet Switching in Radio Channels: Part I- Carrier Sense Multiple-Access Models and their Throughput-Delay Characteristics", IEEE Transactions on Communication, Vol. COM-23, No. 12, pp. 1400-1416, December 1975.

KLEI 75b L. Kleinrock, "Queueing Systems, Vol I: Theory", Wiley Interscience, New York, 1975.

KLEI 76 L. Kleinrock, "Queueing Systems, Vol. II: Computer Applications", Wiley Interscience, New York, 1976.

LAM 75 S.S. Lam and L. Kleinrock, "Packet Switching in Multiaccess Broadcast Channels: Dynamic Control Procedures", IEEE Transactions on Communication, Vol. COM-23, No. 9, pp. 891-904, September 1975.

MEIS 77 N.B. Meisner, D.G. Willard, and G.T. Hopkins, "Time Division Digital Bus Techniques Implemented on Coax Cable", Proceedings of Computer Networking Symposium, June 1977.

METC 76 R.M. Metcalfe and D.R. Boggs, "Ethernet, Distributed Packet Switching for Local Networks", Communications of ACM, Vol. 19, No. 7, pp. 395-404, July 1976.

TOBA 75 F.A. Tobagi and L. Kleinrock, "Packet Switching in Radio Channels: Part II- The Hidden Terminal Problem in Carrier Sense Multiple Access and Busy-Tone Solution", IEEE Transactions on Communication, Vol. COM-23, pp. 1417-1433, December 1975.

TOBA 76 F.A. Tobagi and L. Kleinrock, "Packet Switching in Radio Channels: Part III- Polling and (Dynamic) Split-Channel Reservation Multiple Access", IEEE Transactions on Communication, Vol. COM-24, No. 8, pp. 832-844, August 1976.

TOBA 77 F.A. Tobagi and L. Kleinrock, "Packet Switching in Radio Channels: Part IV- Stability Considerations and Dynamic Control in Carrier Sense Multiple-Access", IEEE Transactions on Communication, Vol. COM-25, No. 10, pp. 1103-1115, October 1977.

TOBA 78a F.A. Tobagi and L. Kleinrock, "The Effect of Acknowledgement Traffic on the Capacity of Packet-Switched Radio Channels", IEEE Transactions on on Communication, Vol. COM-25, No. 10, pp. 815-825, June 1978.

TOBA 78b F.A. Tobagi, M. Gerla, R.W. Peebles, and E.G. Manning, "Modeling and Measurement Techniques in Packet Communicaton Networks", Proceedings of IEEE, Vol. 66, No. 11, pp. 1423-1447, Nov. 1978.

TOBA 79 F.A. Tobagi and V.B. Hunt, "Performance Analysis of Carrier Sense Multiple Access with Collision Detection", Proceedings of LACN Symposium, pp. 217-245, May 1979.

TOBA 80 F.A. Tobagi, "Multiaccess Protocols in Packet Communication Systems", IEEE Transactions on Communication, Vol. COM-28, No. 4, pp. 468-488, April 1980.

TOKO 77 M. Tokoro and K. Tamaro, "Acknowledging Ethernet", Proceedings of LACN Symposium, pp. 120-128, May 1977.

SCHW 77 M. Schwartz, "Computer Communication Network Design and Analysis", Prentice-Hall, New Jersey, 1977.

APPENDIX A

Performance curves of CSMA and CSMA-CD

Reproduced from [TOBA 79]

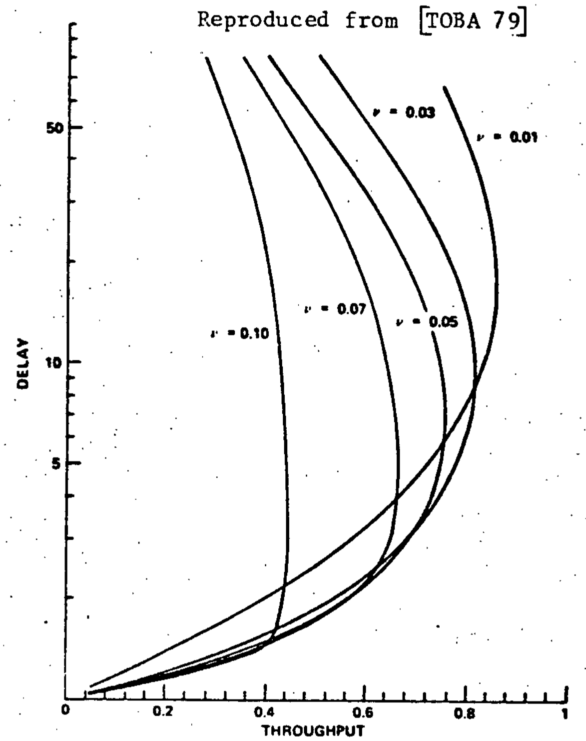


Fig. A-1 *The Throughput-Delay Tradeoff in CSMA at Fixed* **ν**

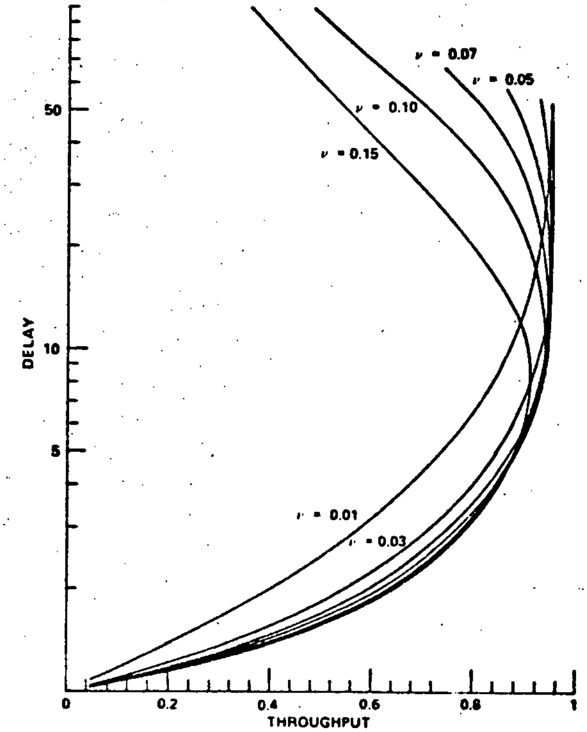


Fig. A-2 *The Throughput-Delay Tradeoff in CSMA-CD at Fixed* **ν**

APPENDIX B

Simulation of CSMA-CD Protocol

The simulation program to study CSMA-CD is written in FORTRAN. The packets, in this program, are represented by an array. These packets have a set of "attributes", represented by other arrays. These attributes are: source and destination addresses, packet size (in slots), packet born time, and collision flag. Time is in units of slots.

Each complete cycle of the running program increments a global timer (by one slot). To generate a packet, one element of the packet array is selected and the global time is assigned to the packet's born time. Source and destination addresses and size are also assigned. In each cycle, the program compares the finish time (i.e. born+size) to the global time; if there is a match, the packet is designated as properly received, and is counted in the throughput and delay statistics. In the program, collisions are checked by comparing the born times and sizes of all the packets to determine conflicts.

The program is typically run for 10000 slots. When the results are overly unexpected (e.g. if the delay for a heavier load is smaller than the delay for a lighter load, or a sample point is radically out of line with other sample points), then the program is run for 20000 slots and the new results are substituted. However, running the program for 10000 slots has been found to be quite satisfactory in terms of the results obtained.

APPENDIX C

Simulation of M/D/m Queue

In order to find the queueing delay of a M/D/m system, a simulation program written in FORTRAN was developed. The Poisson source is simulated by generating packets t slots apart where t is an exponentially distributed random variable. To generate a random number with the same CDF as t, we observe that for any random variable:

$$Pr(CDF(t_1)<CDF(t)\leq CDF(t_2))=Pr(t_1<t\leq t_2)$$

Thus:

$$Pr(CDF(t_1)<CDF(t)\leq CDF(t_2))=CDF(t_2)-CDF(t_1)$$

The above equation is the necessary and sufficient condition for the random variable CDF to be uniformly distributed so:

$$CDF(t)=U$$

Interarrival times are exponentially distributed, as a result [KLEI 75b]:

$$U=CDF(t)=1-exp(-\lambda t)$$

where λ is the mean arrival rate in packets/slot.
Substituting 1-U for U and solving for t:

$$t=-\ln(U)/\lambda$$

The packets are added to the queue at the arrival instances. At the arrival time of a packet, all of the packets that have finished transmission are removed. This strategy reduces the computation overhead significantly, because there is no need to look at both the arrival and the departure instances. The queue is represented by a circular buffer with a capacity of 1000 packets. Experiment shows that this queue size is sufficient to handle ranges of λ not very close to 1 (e.g. $\lambda \leq .98$).

The simulations are performed with different values of λ for 5000 arrivals. An increase in the number of arrivals does not appear to change the results significantly. The program is validated by drawing the simulation results of M/D/1 along with the theoretical delay curve in Fig. 2.3.

APPENDIX D

MCMA DELAY CURVES FOR M=20

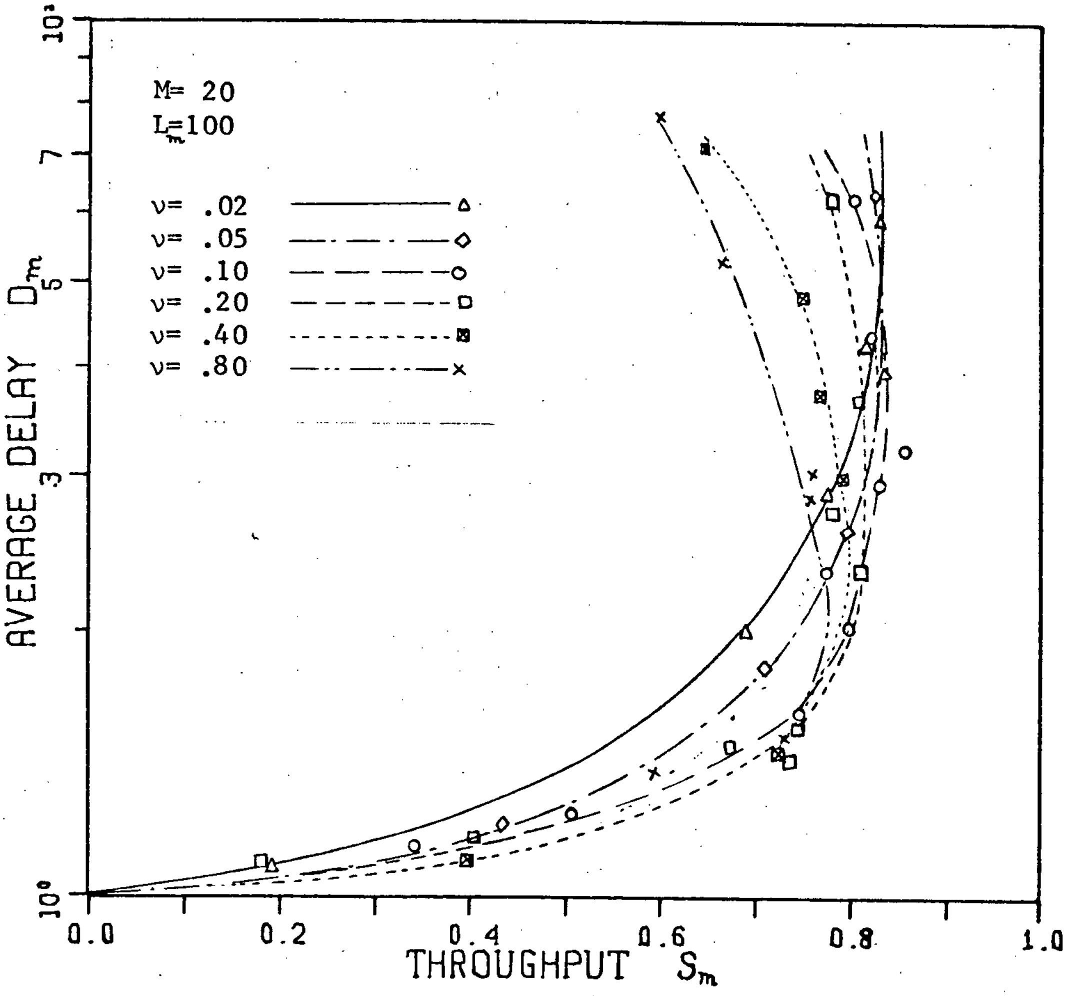


Fig. D-1 MCMA/4: Delay (in units of transmission time of a packet on a sub-channel) vs. throughput

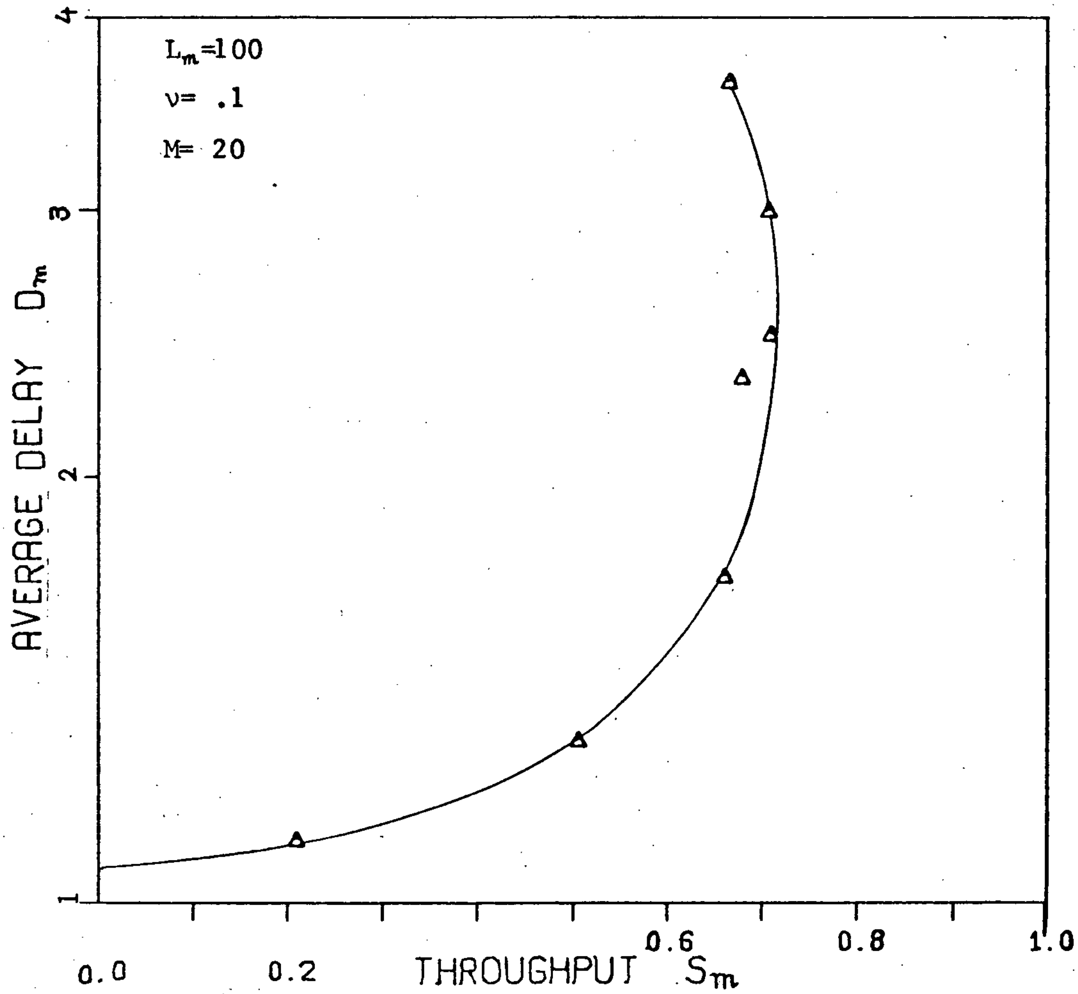


Fig. D-2 MCMA/8: Delay vs. throughput
Delay is in units of packet transmission time
on a sub-channel