ADVANCED TELEPROCESSING SYSTEMS
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

SEMI-ANNUAL TECHNICAL REPORT

MARCH 31, 1983

Principal Investigator: Leonard Kleinrock

Computer Science Department
School of Engineering and Applied Science
University of California
Los Angeles

UNIVERSITY OF CALIFORNIA, LOS ANGELES

SEMI-ANNUAL TECHNICAL REPORTS

Sponsored by the

Defense Advanced Research Projects Agency

Contract Nos.: DAHC-15-C-0368, MDA 903-77-C-0272; DARPA Order No. 2496

COMPUTER NETWORK RESEARCH

DATES	DDC ACCESSION NUMBER
August 1969 to February 1970:	AD 705 149
August 1970:	AD 711 342
June 1971:	AD 727 989
December 1971:	AD 739 705
June 1972:	AD 746 509
December 1972:	AD 756 706
June 1973:	AD 769 706
December 1973:	AD A004167
June 1974:	AD A008422
December 1974:	AD A016823
June 1975:	AD A020671
December 1975:	AD A025914
June 1976:	AD A034171 (Final)

ADVANCED TELEPROCESSING SYSTEMS

June 1976 to December 1976:	AD A039018
June 1977:	AD A047496
June 1978:	AD A077404
September 1979:	AD A081938
March 1980:	AD A088839
September 1981:	to be assigned

ADVANCED TELEPROCESSING SYSTEMS

Semi-Annual Technical Report

March 31, 1983

Contract Number: MDA 903-82-C-0064
DARPA Order Number: 2496
Contract Period: October 1, 1981 to September 30, 1983
Report Period: October 1, 1982 to March 31, 1983

Principal Investigator: Leonard Kleinrock

Co-Principal Investigator: Mario Gerla

(213) 825-2543

Computer Science Department
School of Engineering and Applied Science
University of California, Los Angeles

Sponsored by

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: UCLA-ENG-83-26	2. GOVT ACCESSION NO.: AD-A133555
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): ADVANCED TELEPROCESSING SYSTEMS: SEMI-ANNUAL TECHNICAL REPORT	5. TYPE OF REPORT & PERIOD COVERED: Semi-Annual Technical 1 OCT 82 through 31 MAR 83
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Leonard Kleinrock	8. CONTRACT OR GRANT NUMBER(s): MDA 903-82-C-0064
9. PERFORMING ORGANIZATION NAME AND ADDRESS: School of Engineering and Applied Science, University of California, Los Angeles, CA 90024	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: DARPA Order No. 2496
11. CONTROLLING OFFICE NAME AND ADDRESS: Defense Advanced Research Projects Agency, 1400 Wilson Boulevard, Arlington, VA 22209	12. REPORT DATE: March 31, 1983
	13. NUMBER OF PAGES: 230
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Computer communication networks, Packet switching, Multi-hop networks, Packet radio networks, Channel access protocols, ALOHA network, TDMA, CSMA (carrier-sense multiple-access), Markov chains, Resource sharing, Multi-access communications, Computer communications, Channel sharing.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This Semi-Annual Technical Report covers research carried out by the Advanced Teleprocessing Systems Group at UCLA under DARPA Contract No. MDA 903-82-C-0064 covering the period from October 1, 1982 to March 31, 1983.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(20)

This contract has three primary designated research areas: packet radio systems, resource sharing and allocation, and distributed processing and control.

This report contains the abstracts of the publications which summarize our research results in these areas during this reporting period, followed by the main body of the report which is devoted to a treatment of single- and multi-hop packet radio systems. In particular, the main body consists of the Ph.D. dissertation by Hideaki Takagi, "Analysis of Throughput and Delay for Single- and Multi-Hop Packet Radio Networks," conducted under the supervision of Professor Leonard Kleinrock (Principal Investigator for this research). The work presents a new approach to evaluating the performance of multi-hop packet radio networks, namely, a study of the times between successful transmissions. Also studied is the behavior of packets in a multi-hop system when a fixed transmission radius is specified and this radius is then optimized for throughput. A Markov chain model is also introduced and solved numerically to evaluate transmission and flow control strategies in these systems. In this work we construct models and methodologies to analyze the throughput and delay for a given network by exploring its underlying stochastic processes. We study the packet interdeparture times for single-hop systems in ALOHA and nonpersistent CSMA.

ADVANCED TELEPROCESSING SYSTEMS

Defense Advanced Research Projects Agency
Semi-Annual Technical Report

March 31, 1983

INTRODUCTION

This Semi-Annual Technical Report covers research carried out by the Advanced Teleprocessing Systems Group at UCLA under DARPA Contract No. MDA 903-82-C-0064 covering the period from October 1, 1982 through March 31, 1983. Under this contract we have three designated tasks as follows:

TASK I. PACKET RADIO SYSTEMS

The extension of our analytic and design techniques to modern multi-hop packet radio networks will be studied. The applications and extensions include access methods, large network control and management, queueing network models, approximation methods, capture phenomena, conflict-free algorithms, reliability, routing procedures, topological studies, TDMA in a multi-hop environment, and multiplexing methods.

TASK II. RESOURCE SHARING AND ALLOCATION

Extended concepts of "power" in networks will be studied. The extensions include more complex topologies and configurations, extended queueing disciplines, general distributions, other definitions of power, effects of varying the traffic matrix, fairness. The problems of large scale internetting with respect to resource allocation and sharing will also be studied further.

TASK III. DISTRIBUTED PROCESSING AND CONTROL

Overall principles of distributed processing and distributed control will be studied. The issues of sequencing in data base updates, distributed control, and distributed processing (involving the calculation of concurrency of processing) are the subjects of concern here.

A major contribution of our research during this reporting period is contained in Reference 4 listed below, namely, "Analysis of Throughput and Delay for Single- and Multi-Hop Packet Radio Networks," by Hideaki Takagi. This dissertation was supervised by Professor Leonard Kleinrock (Principal Investigator for this research). It introduces a new technique for evaluating the throughput for single-hop systems, namely, it evaluates the time between successful transmission times and in so doing we are able to find not only the mean throughput but the higher moments of the time between successful departures. The first two moments of the interdeparture distribution are then used to determine the

coefficients needed in a diffusion model which approximates the joint queue length behavior for terminals in such a single-hop environment. This technique is then used to evaluate the performance of various ALOHA and various CSMA (Carrier Sense Multiple Access) systems, including also hidden terminal cases. These diffusion approximation results are shown to be in good agreement with the simulation results. Furthermore, fixed transmission radius systems are studied in an attempt to find the optimal radius to be used in maximizing the expected progress that a packet makes in the way toward its destination. This system is evaluated for a variety of access protocols. Lastly, an exact Markov chain model is formulated and solved numerically which examines the effect of transmission and flow control strategies, as well as finite buffer strategies in a multi-hop environment. The entire dissertation is reproduced as the main body of this report. The following list of research publications summarizes the results of this semi-annual period and the abstract of each paper is given along with the reference itself.

RESEARCH PUBLICATIONS

1. **Belghith, A., "Distributed Routing Scheme in Stationless Multi-hop Packet Radio Networks 'Mobility Handling'." Master's Thesis, Computer Science Department, University of California, Los Angeles, December 1982, also published as a Packet Radio Temporary Note, #313R.**

A mobile stationless multi-hop packet radio network consists of a set of mobile and geographically distributed nodes (e.g., computers, terminals, etc., equipped with radio units), called packet radio units (PRUs), which communicate using a shared broadcast radio channel without a central station control. In this thesis, we present distributed routing schemes for mobility handling and real-time data traffic transport.

In chapter one, we discuss two main issues. First, we provide a validation of the use of tiered-ring architecture for stationless multi-hop packet radio networks. Second, we study a scheme for handling packet radio units mobility by the use of links traversal approach when a PRU is no longer in possession of an outgoing link to communicate with the rest of the network.

In the second chapter, we discuss and study the transport of real-time data traffic. We first discuss the issue of real-time data transport, then we overview a transport scheme called "the duct routing scheme", and provide a detailed comments and critiques of it. Finally, we present a new distributed and highly dynamic scheme called "Tiered Based Dynamic Scheme: TBDS" for the transport of real-time data traffic in highly mobile stationless multi-hop packet radio networks.

Throughout the thesis we give examples and discuss in some details the routing specifications of some existing packet radio networks, in particular the *PRNET* of the Defense Advanced Research Projects Agency (DARPA) and the Advanced Mobile Phone Service (AMPS) of Bell Laboratories.

2. **Naylor, W. E. and L. Kleinrock, "Stream Communication in Packet-Switched Networks,"** *IEEE Transactions on Communications*, **December 1982, pp. 2527-2534.**

In this paper we consider the problem of sending a stream of data (speech, for example) through a packet-switched network which introduces variable source-to-destination delays for different packets of the stream. Ideally, this delay difference should be smoothed so as to preserve the continuity of the stream. We investigate an adaptive destination buffering scheme which may be used to achieve the smoothing of the output stream. The scheme uses delay information, measured for previous streams, in order to compute destination buffering information. Specifically, of the last m packet delays, one discards the largest k and then the range of this partial sample is used for the destination wait time D. We obtain a rule of thumb for choosing m and k, and demonstrate its applicability on some empirical delay distributions from ARPANET measurements. It is, in general, necessary to deal with discontinuities which occur even after smoothing. To this end, we consider two possible playback schemes: method E (time expanded in order to preserve information) and method I (late data ignored in order to preserve timing). The two methods are at opposite ends of a continuum of possible playback schemes. We study the implication of methods E and I on the choice of smoothing parameters and establish a foundation for evaluating all schemes in this continuum.

3. **Belghith, A., "A Distributed Routing Scheme in Stationless Multi-hop Packet Radio Networks 'Mobility Handling',"** *ACM SIGCOMM '83 Proceedings*, **March 9, 1983, Austin, Texas.**

A mobile stationless multi-hop packet radio network consists of a set of mobile and geographically distributed nodes (e.g., computers, terminals, etc., equipped with radio units), called packet radio units (PRUs), which communicate using a shared broadcast radio channel without a central station control. In this paper we consider the routing in highly mobile stationless multi-hop packet radio networks. We provide a validation of the use of the tier-ring architecture, and we present a scheme for handling mobile packet radio units in stationless environment by the use of a link-traversal approach when a node (packet radio unit) is no longer in possession of an outgoing link to communicate with the rest of the network.

4. **Takagi, H., "Analysis of Throughput and Delay for Single- and Multi-Hop Packet Radio Networks," Ph.D. Dissertation, Computer Science Department, University of California, Los Angeles, March 1980, Report No. CSD-830523.**

UCLA-ENG-8326
MAY 1983

ANALYSIS OF THROUGHPUT AND DELAY FOR SINGLE- AND MULTI-HOP PACKET RADIO NETWORKS

by

Hideaki Takagi

This research, conducted under the chairmanship of Professor Leonard Kleinrock, was sponsored by the Defense Advanced Research Projects Agency, Department of Defense.

Computer Science Department
School of Engineering and Applied Science
University of California
Los Angeles

Acknowledgments

This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense under Contracts MDA 903-77-C-0272 and MDA 903-82-C-0064, and by IBM Japan, Ltd. through the Overseas Scholarship Program.

I would like to express my appreciation to my doctoral committee consisting of Professors Mario Gerla, James R. Jackson, Leonard Kleinrock, Steven A. Lippman and Richard R. Muntz. I am particularly grateful to the committee chairman and my advisor, Dr. Leonard Kleinrock for his enlightening and stimulating discussions throughout the course of this dissertation work. Special thanks also go to Doctors Hisashi Kobayashi and Stephan S. Lavenberg of IBM Corporation for their constant encouragement.

I am also thankful to the following staff (DARPA employees) at the UCLA Computer Science Department for administrative and technical assistance: Amanda Daniels, George Ann Hornor, Linda Infeld, Lillian Larijani, Terry Peters and Ruth Pordy (for drawing figures).

Chatting with insightful friends has always been fun throughout the four years of my Ph.D. program. I shall not forget endless discussions with Randolph Nelson and Richard Gail with equations written on napkins at lunch time. Other students of the ATS (Advanced Teleprocessing Systems) group have also been nice to talk with. They include Abdelfettah Belghith, Rina Dechter, Joseph Green, Kenneth C. Kung and Hanoch Levy. During the last three quarters, I presided an informal seminar to study stochastic processes as applied to computer science. To its regular members (Hak-Wai Chan, Luis De Moraes, Paulo Rodrigues, Behrokh Samadi and Edmundo Souza e Silva), I offer my sincere thanks for exciting discussions. Other computer science students whose friendship I appreciate include Rodolfo Pazos and Yoshitaka Shibata.

Finally, I dedicate this dissertation to Yoko, my wife, who has been a great partner of life and so understanding in this foreign country. Playing with my children, Mayu and Takuma, has been a relief from often tough research. My last and greatest appreciation goes to my parents, Yoshio and Kiyo Takagi, in Japan for their everlasting encouragement through numberless letters.

Abstract

A packet radio network is a packet-switching, computer-communication network for geographically distributed fixed and mobile users over a radio channel. When the final destinations of packets are generally beyond a transmission radius, they are relayed by users in a store-and-forward fashion (multi-hop case); otherwise, we have a single-hop system. We focus on the performance evaluation for random channel-access protocols such as ALOHA and carrier-sense multiple-access (CSMA).

Two fundamental performance measures of concern in a packet radio network are the maximum-achievable throughput and the average packet delay given the throughput requirement. We construct models and methodologies to analyze the throughput and delay for a given network by exploring its underlying stochastic processes. In cases where exact analysis fails, we provide approximate solutions backed up with simulations of the exact models.

Specifically, we study for single-hop systems the packet interdeparture times (i.e., the intervals between two successive successful transmissions) in ALOHA and nonpersistent CSMA (including the hidden-terminal environment). The reciprocal of the mean interdeparture time is the channel throughput traditionally studied. The first two moments of the distribution for the interdeparture time are used to determine the coefficients in the diffusion process approximation to the joint queue length distribution for buffered users. The proposed average packet delay formulas (for statistically identical users) based on the diffusion approximation are shown to be in good agreement with simulation results. For persistent CSMA protocols, we analyze the throughput only. Our results for single-hop systems are readily applicable to local-area computer networks.

For multi-hop systems, we evaluate the expected progress of packets towards their final destinations when a finite transmission radius is given. We find the optimal radius to maximize the expected progress under a variety of conditions. Also, for a store-and-forward network of slotted ALOHA users, an exact Markov chain model is formulated and solved numerically to examine the effects of some transmission and flow-control strategies and multiple buffers.

Table of Contents

page

Acknowledgments iii

Abstract v

Table of Contents vii

List of Figures x

List of Tables xiii

Glossary of Notation xiv

1 Introduction 1
1.1 Overview of the Packet Radio Technology 1
1.1.1 Evolution of Packet Radio Networks 1
1.1.2 System Functional Characteristics 3
1.1.3 Some Existing Ground Radio Networks 8
1.2 Outline of the Dissertation 12
1.2.1 Models, Measures and Methodologies 12
1.2.2 Summary of the Results 13
1.3 Survey of Related Work 16
1.3.1 Single-Hop Systems 16
1.3.2 Multi-Hop Systems 19

2 Throughput Analysis for Finite-Population Packet Broadcasting Systems 21
2.1 Introduction 21
2.2 Throughput Analysis 22
2.2.1 Pure ALOHA 23
2.2.2 Pure ALOHA with Delay Capture 27
2.2.3 Slotted Persistent CSMA 28
2.2.4 Slotted Persistent CSMA with Collision Detection 35
2.2.5 Unslotted Persistent CSMA 40
2.2.6 Unslotted 1-Persistent CSMA 44
2.2.7 Unslotted Persistent CSMA with Collision Detection 46
2.3 Conclusion 51

3 Output Processes in Contention Packet Broadcasting Systems 52
3.1 Introduction 52
3.2 The Number of Successful Transmissions 53
3.3 Output Processes for Identical Users 55
3.3.1 Slotted ALOHA 56
3.3.2 Pure ALOHA (Infinite Population) 57
3.3.3 Slotted CSMA and CSMA with Collision Detection 60
3.3.4 Unslotted CSMA 64
3.3.5 Unslotted CSMA with Collision Detection 65
3.3.6 Unslotted CSMA with Delay Capture 66
3.3.7 Discussions of the Numerical Results 69

3.4 Output Processes for Nonidentical Users 76
3.4.1 Slotted ALOHA 76
3.4.2 Slotted CSMA and CSMA with Collision Detection 77
3.4.3 Unslotted CSMA 79
3.4.4 Unslotted CSMA with Collision Detection 81
3.5 Conclusion 83

4 Approximate Output Processes in Hidden-User Packet Radio Systems 84
4.1 Introduction and Assumptions 84
4.2 Superposition of Independent Renewal Processes 86
4.2.1 Continuous-Time Systems 87
4.2.2 Discrete-Time Systems 90
4.3 Analysis of Output Processes 92
4.3.1 Pure ALOHA 93
4.3.2 Unslotted CSMA with Zero Propagation Delay 96
4.3.3 Unslotted CSMA with Delay Capture 101
4.3.4 Unslotted CSMA with Nonzero Propagation Delay 106
4.3.5 Slotted CSMA 114
4.4 Discussion 118
4.5 Conclusion 119

5 Queueing Delays in Contention Packet Broadcasting Systems 121
5.1 Introduction 121
5.2 Exact Analysis for Two Identical Users 122
5.2.1 Slotted ALOHA 123
5.2.2 Slotted CSMA with Collision Detection 124
5.3 Bounds on the Mean Delay in Slotted ALOHA 130
5.4 Diffusion Process Approximation for a Contention System 132
5.4.1 Diffusion Equation for a Contention System 135
5.4.2 Stationary Solution to the Diffusion Equation 137
5.4.3 Proposed Mean Delay Formulas 138
5.4.4 Discussion of the Numerical Examples 140
5.5 Conclusion 142

6 Optimal Transmission Ranges for Randomly Distributed Packet Radio Terminals 145
6.1 Introduction 145
6.2 Optimal Transmission Radii for Slotted ALOHA 146
6.3 Optimal Transmission Radii for ALOHA with Capture 151
6.4 Optimal Transmission Radii for CSMA 155
6.5 Optimal Transmission Radii in an Inhomogeneous Density of Terminals 160
6.6 Optimal Number of Neighbors for ALOHA with Capture 164
6.7 Conclusion 169

7 Throughput-Delay Characteristics of Some Slotted-ALOHA Multi-Hop Packet Radio Networks 172
7.1 Introduction 172
7.2 Basic Model 173
7.3 Formulation 176
7.4 Improvement by Transmission Suppression 183
7.5 Improvement by Transmission Acceleration 184
7.6 Improvement by Multiple Buffers for Repeaters 185
7.7 Conclusion 186

8 Concluding Remarks 189
8.1 Application to Local Area Networks 189
8.2 Considerations on the Model and the Methodology 189
8.3 Suggestions for Future Research in Multi-Hop Systems 190

Appendices 191
A. Derivation of Eq.(3.4) 191
B. Derivation of Eq.(3.22) 192
C. Derivation of Eqs.(3.45) and (3.47) 193
D. Determination of the Transmission Rate for CSMA 195
E. A Numerical Technique to Solve a Large Sparse Markov Chain 195

References 198

List of Figures

page

1.1 Characterization of the multi-hop packet radio network 2

2.1 Allowable source rates for pure ALOHA 26

2.2 Capacity of optimaliy-controlled pure ALOHA with delay capture 29

2.3 Channel state in slotted persistent CSMA (a) Overview of channel state (b) Close look at the jth sub(busy)period 30

2.4 Throughput of slotted h-persistent CSMA 36

2.5 Throughput of slotted h-persistent CSMA with collision detection 38

2.6 Throughput of slotted 0.03-persistent CSMA with collision detection 39

2.7 Channel state in unslotted persistent CSMA (a) Unslotted persistent CSMA (b) Unslotted persistent CSMA with collision detection 41

2.8 Throughput of unslotted h-persistent CSMA 45

2.9 Throughput of unslotted persistent CSMA with collision detection 48

2.10 Throughput of unslotted 1-persistent CSMA with collision detection 50

3.1 Packet interdeparture times X (a) Slotted ALOHA (b) Slotted CSMA with collision detection 58

3.1 Packet interdeparture times X (c) Unslotted CSAM (d) Unslotted CSMA with collision detection 59

3.2(a) Packet interdeparture time X in pure ALOHA 61

3.2(b) Unsuccessful transmission period in pure ALOHA 61

3.2(c) Probability density function for packet interdeparture time in pure ALOHA 62

3.3 Collision detection timing in unslotted CSMA/CD 67

3.4 Transmission periods in unslotted CSMA with delay capture (a) Successful transmission period (b) Unsuccessful transmission period 70

3.5(a) The (maximized) channel throughput S 71

3.5(b) The coefficient of variation C^2 for packet interdeparture time 72

3.6(a) The channel throughput (S) of unslotted CSMA with delay capture 73

3.6(b) The coefficient of variation (C^2) for packet interdeparture time in unslotted CSMA with delay capture 74

3.7 The channel throughput (S) and coefficient of variation (C^2) for packet interdeparture time in slotted CSMA/CD 75

3.8 $\{S_i\}$ and $\{C_i^2\}$ for 5 nonidentical users of slotted CSMA 78

4.1 Interevent time in the superposition process 88

4.2 Relative error of the mean duration of an unsuccessful transmission period ($\bar{F}$) in pure ALOHA 97

4.3 The hearing graphs of example configurations (a) Symmetric hidden-user configuration ($M=8$, $m=7$) (b) Wall configuration ($M=10$) 100

4.4(a) Throughput (S) in unslotted CSMA (zero propagation delay) for symmetric hidden-user configurations 102

4.4(b) Coefficient of variation of the packet interdeparture time (C^2) in unslotted CSMA (zero propagation delay) for symmetric hidden-user configurations 103

4.5(a) Individual user throughput (S_i) and channel throughput (S) in unslotted CSMA (zero propagation delay) for a wall configuration in Figure 4.3(a) 104

4.5(b) Coefficient of variation of the packet interdeparture time (C^2 for the system and C_i^2 for user i) in unslotted CSMA (zero propagation delay) for a wall configuration in Figure 4.3(b) 105

4.6(a) Throughput (S) in unslotted CSMA (zero propagation delay) with perfect capture for symmetric hidden-user configurations 107

4.6(b) Coefficient of variation of the packet interdeparture time (C^2) in unslotted CSMA (zero propagation delay) with perfect capture for symmetric hidden-user configurations 108

4.7(a) Throughput (S) in unslotted CSMA (nonzero propagation delay) for a symmetric hidden-user configuration ($M=20$, $m=19$) 112

4.7(b) Coefficient of variation of the packet interdeparture time (C^2) in unslotted CSMA (nonzero propagation delay) for a symmetric hidden-user configuration ($M=20$, $m=19$) 113

4.8(a) Throughput (S) in slotted CSMA for a symmetric hidden-user configuration ($M=20$, $m=19$) 116

4.8(b) Coefficient of variation of the packet interdeparture time (C^2) in slotted CSMA for a symmetric hidden-user configuration ($M=20$, $m=19$) 117

5.1 An illustration of the channel state in slotted CSMA with collision detection (The embedded Markov epochs are shown by •.) 125

5.2 Mean delay for two identical users of CSMA with collision detection 129

5.3 Mean delay for a system of 3 identical users of slotted ALOHA 133

5.4 Mean delay for a system of 10 identical users of slotted ALOHA (with simulation results) 134

5.5 Comparison of the diffusion approximation with simulation results in pure ALOHA 141

5.6 Comparison of the diffusion approximation with simulation results in hidden-user, unslotted CSMA 143

5.7 Comparison of the diffusion approximation with simulation results in wall configuration, unslotted CSMA 144

6.1 The area of interference with the transmission $P \rightarrow Q$ 150

6.2 The position of the receiver Q 150

6.3 The optimal transmission for slotted ALOHA networks without capture 152

6.4 The area of terminals possibly interfering the transmission $P \rightarrow Q$ (a) $\alpha r < R$ (b) $\alpha r > R$ 154

6.5 The position of the receiver Q with respect to the transmitter P 154

6.6 The optimal transmissions for slotted ALOHA with capture 156

6.7 Slotted nonpersistent CSMA (transmission and idle periods) 158

6.8 The period for the transmission P to Q vulnerable to the transmissions from areas A and B (a) Configuration (b) Time line (•:vulnerable points to A and B, ○:vulnerable points to B) 158

6.9 Comparison of the optimized expected progress among ALOHA with and without capture, and CSMA networks 161

6.10 Three cases of the position of receiver Q (P:the transmitter, ////:the area of possibly interfering terminals (a) $x \leqslant b - R$ (b) $b - R \leqslant x \leqslant 0$ (c) $b \leqslant x \leqslant R$ 162

6.11 The optimal transmission radii and expected progress for packets crossing a gap 165

6.12 The angular position of the jth nearest neighbor 166

6.13 The optimal transmissions for ALOHA with capture and without transmission radius 170

7.1 Examples of networks (a) Network 1 — 5 terminals, 3 repeaters and 3 paths (b) Network 2 — 7 terminals, 5 repeaters and 4 paths 175

7.2 Throughput-delay characteristics for network 1 with single-buffered repeaters 180

7.3 Throughput-delay characteristics for network 2 with single-buffered repeaters 181

7.4 Throughput-delay characteristics for individual paths in network 2 (basic protocol and single buffer) 182

7.5 Throughput-delay characteristics for network 1 (basic protocol) with m buffers for repeaters ($m = 1,2,3$) 187

7.6 Throughput-delay characteristics for network 1 with double-buffered repeaters 188

List of Tables

Page

1.1 Coverage of each chapter 17

1.2 Chapter reference to protocol model 17

6.1 The routing probabilities $a_j(N)$ 167

E.1 The number of states and the sparseness of $\underline{P}$ for some network models depicted in Figure 7.1. Suppression or acceleration of transmission is not employed. m = number of buffers in each repeater 197

Glossary of Notation

Notation	Definition	Referenced chapter(s)
$A_i(t)$	Number of packet arrivals at user i during $[0,t]$.	5
a	Propagation delay (also slot size in slotted CSMA).	2-6
B	Duration of the channel busy (transmission) period.	2
$B^{(j)}$	Duration of the jth sub(busy)period in B.	2
B_n	Mean duration of the busy period initiated by n packets.	2
R	Mean backlog of packets summed over the interval between two successive Markov epochs in the analysis of slotted CSMA.	5
b	Collision detection parameter.	2,3,5
b	Width of a vacant strip in an otherwise randomly distributed population of terminals.	6
BTMA	Busy tone multiple access.	
CD	Collision detection.	
CDMA	Code division multiple access.	
CSMA	Carrier sense multiple access.	
C^2	Coefficient of variation for the packet interdeparture time. $C^2 = Var[X]/\bar{X}^2$.	3-5
C_i^2	Coefficient of variation for the packet interdeparture time from user i. $C_i^2 = Var[X_i]/\bar{X}_i^2$.	3-5
$C_{a,i}^2$	Coefficient of variation for the packet interarrival time at user i.	5
$Cov[f,g]$	Covariance for the random variables f and g.	
c	Delay capture parameter. The time needed to capture a leading packet.	2-4
c^2	$= C_i^2$ for identical users.	5
c_a^2	$= C_{a,i}^2$ for identical users.	5
D	Mean packet delay. $D = \bar{Q}/S$.	5,7
$D_i(t)$	Number of packet departures from user i during $[0,t]$.	3,5
$D(k)$	Mean packet delay for path k.	7
D_{HT}	Upper bound on D based on the heavy-traffic assumption.	5
D_U	A tighter upper bound on D.	5
D_i^*, D_i^{**}	Proposed mean delay approximation for user i.	5
E	Probability of an idle slot. Generally, $E = \prod_{i=1}^{M}(1-p_i)$. For identical users, $E = (1-p)^M$.	3,4
$E[f]$	Expected value of the random variable f.	

$E[f\|A]$	Expected value of the random variable f conditioned on the event A.	
e_i	Behavior of user i; i.e., $e_i = k\ (\neq 0)$ if user i transmits a packet of path k, and $e_i = 0$ if he does not transmit.	7
e	Behavior of system. $\mathbf{e} = (e_1, e_2, \cdots, e_M)$.	7
FM	Frequency modulation.	
$F^{(n)}, F$	Duration of the (nth) unsuccessful transmission period.	3,4
$F^*(s)$	Laplace transform of pdf for F.	3
$f^{(n)}, f$	Duration of the (nth) interval between two successive transmission start times such that it is shorter than $1+a$.	3,4
G	Rate of starting transmission by breaking the channel idle period. For identical users: $G = gM$ (pure ALOHA), $G = pM$ (slotted ALOHA), $G = gM/a$ (unslotted CSMA), $G = pM/a$ (slotted CSMA).	2-5
$G(\mathbf{z})$	z-transform of the joint queue length distribution, where $\mathbf{z} = [z_1, z_2, \cdots, z_M]$.	5
g	Rate of starting transmission by each of identical users in a unslotted system.	2-5
g_i	Rate of starting transmission by user i in a unslotted system.	2-5
g'_i	Reduced rate (due to carrier-sensing) of of starting transmission by user i.	4
$H(i)$	Set of user indices who can hear transmission by user i and vice versa (including i).	4
$H_i(i)$	$\equiv H(i) - \{i\}$.	4
h	Persistence parameter in persistent CSMA. $0 < h \leqslant 1$ (slotted); $0 < h$ (unslotted).	2
h_{ij}	$= 1$ if users i and j can hear each other, $= 0$ otherwise.	7
$I^{(n)}, I$	Duration of the (nth) channel idle period.	2-4
$I^*(s)$	Laplace transform of pdf for I.	3
$\bar{I}_i$	Mean duration of the idle (non-transmitting) state at user i.	2,4
J	Number of sub(busy)periods in a busy period B.	2
K	Number of transmission periods until success.	3,4
$K^*(z)$	z-transform of the distribution for K.	3
L	Number of transmissions (excluding the initial one) involved in an unsuccessful transmission period.	3,4
L	Mean length of the interval between two successive Markov epochs in the analysis of slotted CSMA.	5
M	Number of users.	2-5
m	Number of hearable users in symmetric hidden-user configurations.	4
$\mathbf{m} = [m_i]^t$	Column vector of the stationary infinitesimal means.	5
MFN	Most forward within N (routing).	6
MFR	Most forward within R (routing).	6

N	Mean number of terminals within a circle of radius R. $N = \lambda\pi R^2$.	6
P_0	Probability that no users are transmitting at an arbitrary time.	2
pdf	Probability density function.	
p	Probability of transmission in a slot by each of identical users.	2-7
p_i	Probability of transmission in a slot by user i.	3,4
p'_i	Reduced rate (due to carrier-sensing) of starting transmission in a slot by user i.	4
$p(k)$	Probability that a user holding a packet of path k transmits it in a slot.	7
p_{nl}	Probability of k arrivals in a sub(busy)period initiated by n packets.	2
$Prob[A]$	Probability of the event A.	
$Prob[A\|B]$	Probability of the event A conditioned on the event B.	
PRU	Packet radio unit.	1
Q	Number of packets in system.	5,7
Q_i	Number of packets buffered in user i in stationary state.	5
$Q_i(t)$	Number of packets buffered in user i at time t.	5
$Q(k)$	Number of backlogged packets on path k.	7
q_i	Probability that a successful transmission is achieved by user i. For identical users, $q_i = 1/M$.	3-5
R	Radius of packet transmission.	6
$R^{(l)}$	Duration of the transmission delay in $B^{(l)}$.	2
R_n	Duration of the transmission delay in a sub(busy)period initiated by n packets.	2
ROP	Radio-on packet.	1,8
S	System throughput. Generally, $S = \sum_{i=1}^{M} S_i$. For identical users $S = sM$.	2-7
S_i	Throughput for user i. $S_i = q_i S$.	3-5
$S(k)$	Throughput for path k.	7
s	$= S_i$ for identical users.	5
s	Laplace transform variable.	3
s_i	State of user i; i.e., $s_i = k\ (\neq 0)$ if user i has a packet of path k, and $s_i = 0$ if his buffer is empty.	7
$\mathbf{s}$	State of system. $\mathbf{s} = (s_1, s_2, \cdots, s_M)$	7
T	Duration of a successful transmission period.	3,4
$T^*(s)$	Laplace transform of pdf for T.	3
$T^{(l)}$	Duration of the transmission time in $B^{(l)}$.	2
t	Time. The unit of time is the transmission time of a constant-length packet.	2-7
U	Time spent for successful transmission(s) in a busy period B. In a slotted system, it is identical to U below.	2
$U^{(l)}$	Time spent for a successful transmission in $B^{(l)}$.	2

U_n	Expected number of successful transmissions in the busy period initiated by n packets.	2
U_i	$= \nu_i\gamma_i = u_i/(1-E)$. Note that $\gamma = \sum_{i=1}^{M} U_i$ and $U = (1-E)\sum_{i=1}^{M} U_i$.	3
U	Probability of starting a successful transmission in a slotted system. Generally, $U = \sum_{i=1}^{M} u_i$. For identical users, $U = pM(1-p)^{M-1}$.	3
u_i	Probability of starting a successful transmission by user i in a slotted system. $u_i = p_i \prod_{j=1(j\neq i)}^{M} (1-p_j)$.	3
$Var[f]$	Variance of the random variable f.	
ν_i	Probability that user i initiates a transmission period.	3,4
X	Packet interdeparture time.	3,4
$X^*(s)$	Laplace transform of pdf for X.	3
X_i	Packet interdeparture time for user i.	3
Y	Transmission start time of the last overlapping packet.	2,3
$Y(i)$	Transmission start time of the last overlapping packet in an unsuccessful transmission period initiated by user i.	3,4
Y_1	Transmission start time of the first overlapping packet.	2,3
$Y_1(i)$	Transmission start time of the first overlapping packet in an unsuccessful transmission period initiated by user i.	3
$Y_{(n)}$	Transmission start time of the last overlapping packet in a sub(busy)period initiated by n packets.	2
Z	Expected progress of a packet.	6
ZAP	Zero-and-pole approximation for throughput-delay curves.	
z	z-transform variable.	
α	FM capture parameter.	6
β	Dimensionless gap intensity. $\beta = \lambda b^2$.	6
γ	Probability that a transmission period is successful. $\gamma = \sum_{i=1}^{M} \nu_i\gamma_i = U/(1-E)$.	3,4
γ_i	Probability that a transmission period initiated by user i is successful.	3,4
$\gamma_{(n)}$	Probability of a successful transmission in a sub(busy)period initiated by n packets.	2
δ	Probability that an arbitrary interval between two successive transmission start times is no shorter than the packet transmission time $1+a$.	4
δ_{ij}	Kronecker's delta. $\delta_{ii} = 1$, and $\delta_{ij} = 0$ for $i \neq j$.	
λ	Area density for the planar distribution of terminals.	6
λ	$= \lambda_i$ for identical users.	5
λ_i	Mean packet arrival rate at user i.	5
$\lambda(k)$	Probability that a thinking terminal for path k generates a packet in a slot.	7
$\pi(s)$	$Prob$[system state $= s$].	7

$\sigma = [\sigma_{ij}^2]$	Matrix of the stationary infinitesimal covariances.	5
τ	Number of slots included in the transmission time of a packet (in slotted CSMA). $\tau = \lceil (1-b)/(b+a) \rceil$ in Section 5.2.2. Elsewhere, $\tau = 1/a$.	4-6
$\bar{f}$	$= E[f]$, expected value of the random variable f.	
$\bar{x}$	$= 1 - x$.	5
G_i	Derivative of $G(\mathbf{z})$ with respect to z_i.	5
$\|A\|$	Number of elements in the set A.	4
$\lceil x \rceil$	Ceiling of x.	5
$\det(A)$	Determinant of the matrix A.	5

CHAPTER 1
Introduction

This introductory chapter first delineates the background technology for the subject of this dissertation. Then we summarize our contribution. This is followed by a brief survey of other related work. Throughout the chapter, key words are italicized.

1.1 Overview of the Packet Radio Technology

1.1.1 Evolution of Packet Radio Networks

Recently we have been witnessing exciting progress in packet radio technology as it applies to the fields of digital radio and computer communication networks. The *packet radio network* provides a capability for geographically distributed fixed and mobile digital equipment to reliably communicate with each other.

A typical packet radio network contains three primary functional elements: terminals, stations and repeaters. We call them *packet radio units* (PRUs) collectively. All of these are assumed to share a common radio frequency (RF) bandwidth for communication. The users access the network via *terminals* which support devices for a human interface, unattended sensors and instruments, or host computers. Some terminals act as *gateways* to other networks. The *stations* have the responsibility for overall management of the network including initialization, routing, traffic control and accounting functions. Depending on the terrain, *repeaters* may be used for extending the geographic range of the network to provide connectivity for terminals not in line-of-sight of others. In such a case, packets are transported through repeaters in a store-and-forward fashion, and so the network is called *multi-hop*. If a single receiver (e.g., a central station) is in the transmission range of all users, we have a *single-hop* system.

Thus the multi-hop packet radio technology may be thought of as a union of the two technical developments in the early 1970's: *packet-switching store-and-forward wire-based networks* (originated in the ARPANET by the Department of Defense [Klei76]) and *packet-broadcasting single-hop radio-based systems* (pioneered by the ALOHA project at the University of Hawaii [Abra73a]). *Packet switching*, as opposed to message or circuit switching, was designed to provide efficient network communications for bursty traffic, and to facilitate resource sharing. In Figure 1.1, we itemize some functional characteristics and performance modeling methodologies for the two networking technologies. Some terminology in this figure is explained in the next section.

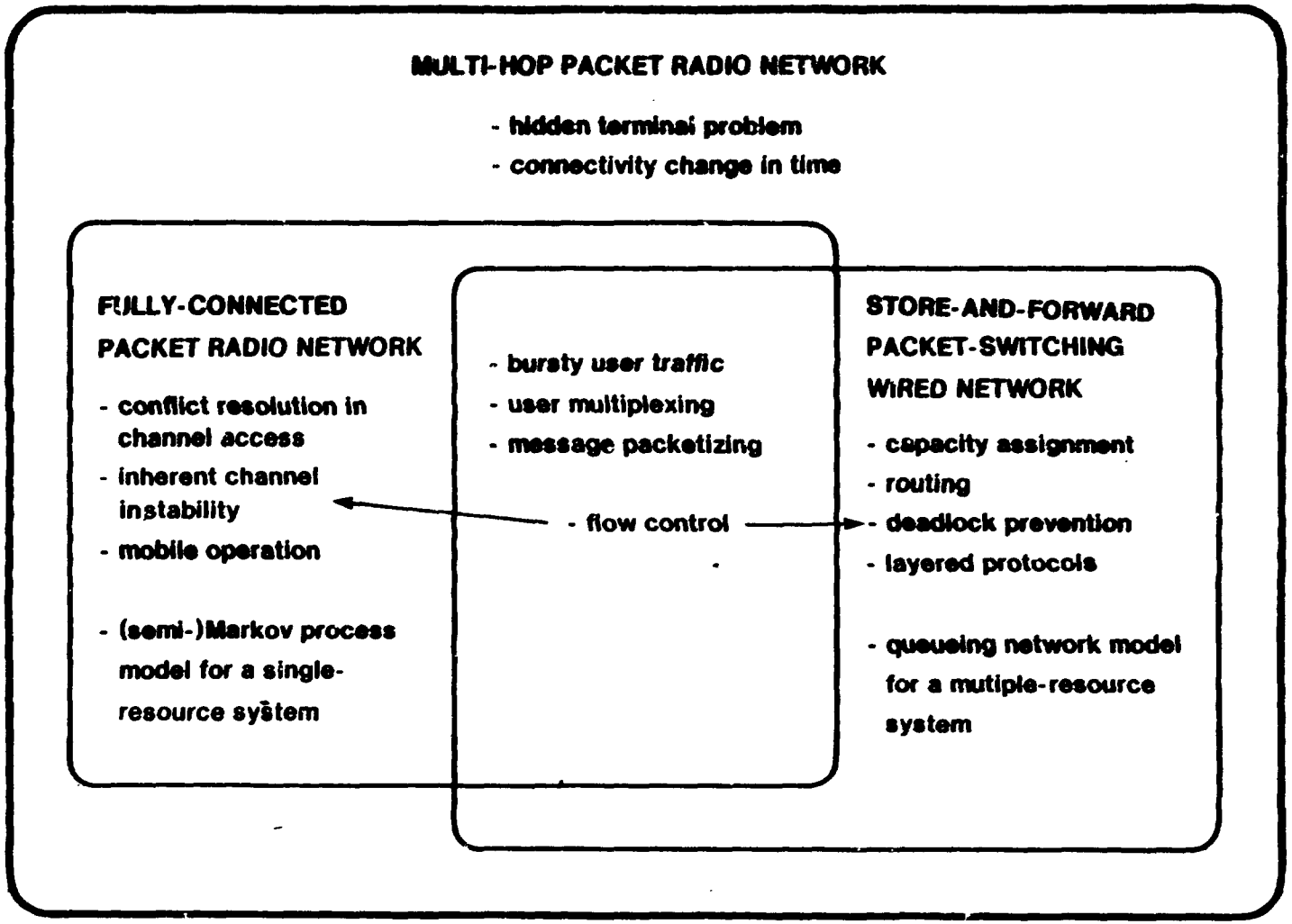


Figure 1.1 Characteristics of the multi-hop packet radio network.

1.1.2 System Functional Characteristics

Here we describe the key components and operations in single- and multi-hop packet radio networks. It is the purpose of this section to demonstrate important system aspects from the viewpoint of their effect on network performance. The discussion aims at providing preliminary knowledge for reading the following chapters in this dissertation.

Channel Access Protocols

A basic characteristic of a radio channel is its broadcast nature. Coupled with the multiple-access aspect (i.e., two or more users may attempt to use the channel at the same time), this represents a major cause of throughput degradation; namely, transmissions from different PRUs may interfere (*collide*) with one another. If all PRUs are centrally controlled or in line-of-sight of all others, there are a number of protocols for channel access that can handle the conflict resolution more or less efficiently (to name a few, TDMA (time-division multiple-access), FDMA (frequency-division multiple-access) and polling). However, in a general multi-hop environment, such cooperation of PRUs is not simple because a transmitter may not be able to exchange information directly with other potential transmitters which have a common intended receiver. (See [Toba80c] for a detailed classification of protocols.)

Protocols which are prone to collision are called *random-access* or *contention-type*. They are relatively simple and insensitive to topological changes in the PRU configuration at the expense of low maximum-achievable throughput and instability when uncontrolled. *ALOHA* and *carrier-sense multiple-access* (CSMA) fall in this category. (ALOHA permits a user to transmit any time he desires. CSMA tries to avoid collisions by listening to the carrier due to other users' transmission before transmitting, and inhibiting transmission if the channel is sensed busy [Toba80c].) Note in multi-hop systems that even CSMA with the assumptions of zero signal propagation delay and zero carrier detection time is subject to collisions due to the contention between mutually hidden PRUs. (PRUs i and j are said to be *mutually hidden* when i and j cannot hear the transmission from each other. The conflict due to hidden PRUs is called the *hidden-terminal problem* after [Toba74,Toba75].)

Collision-free protocols in multi-hop environment are also conceivable. An obvious (but very inefficient for bursty traffic) way is to pre-assign distinct channels to each PRU pair. Various procedures to minimize the number of channels by means of duplicate assignment (for non-interfering pairs) form the optimization theory for *frequency assignment problems* [Hale80]. A collision-free scheme with a single channel may be realized by giving full transmission rights only to a set of non-interfering PRU pairs and varying such sets in a cyclic or random fashion. This is called *spatial TDMA* in [Nels82]. (The aforementioned frequency assignment problems may then be called *spatial FDMA*.) By the same token, we may think of *spatial CDMA* (code-division multiple-access) which, by use of pseudo-noise (PN), frequency-hopping (FH), or hybrid (PN/FH) *spread spectrum modulation*, assigns different PN chip patterns or different FH hopping patterns to interfering PRU pairs [Kahn78]. (The same patterns can be used for non-

interfering PRU pairs.) Here the assumption of zero propagation and detection times is necessary for perfect collision-freedom. Dynamic collision-free schemes utilizing *busy tone* (in a separate channel) emanating from PRUs which hear channel activity are proposed in [Toba74,Toba75] (where the term *BTMA* (busy-tone multiple-access) was introduced) and in [Sidi81]. Here again, the assumption of zero propagation and detection times is necessary for complete collision-freedom.

Capture

Capture is the ability of a receiver to successfully receive a packet even though part or all of the packet arrives at the receiver overlapped in time by other packets. This capability greatly enhances multiple access efficiency. The basic mechanism of capture is the ability of the receiver to synchronize with and lock on to one packet and subsequently reject other overlapping packets as noise. We consider two types of capture.

When *frequency modulation* (FM) is employed, we may have the *FM capture* effect [Abra77]. That is, a receiver correctly receives a packet from a transmitter located at a distance r from the receiver, if none of the other PRUs within a radius αr of the receiver transmit simultaneously ($1 \leqslant \alpha$). The case $\alpha = 1$ is called *perfect capture*, whereas the case $\alpha \rightarrow \infty$ corresponds to *no capture*.

Networks which use spread spectrum modulation may exhibit a *delay capture* phenomenon [Davi80]. If the modulation pattern does not repeat within a packet duration, then two packets which use the same pattern would be strongly correlated over each data bit if they arrived at a receiver simultaneously, but would be pseudo-orthogonal if they arrived with a certain time offset. This property allows the first arriving packet at an idle receiver to be captured and successfully received while the later packets are rejected as noise. *Perfect capture* occurs when the first packet is always captured even if another packet arrives with a vanishingly small delay. In reality, this cannot be achieved because of a vulnerable period at the beginning of the reception and the interference rejection margin such that the received signal strength of rejected packets may exceed that of the captured packet.

Acknowledgment Schemes

In order to delete a copy of a packet from a transmitter's buffer, it must be sure that the packet has been successfully received at the intended receiver. Explicit or implicit acknowledgments are used to notify transmitters of such information. Explicit acknowledgments from receivers are not necessary in satellite packet switching where each transmitter can listen to its own transmission from a satellite transponder after a long (about a quarter second) round-trip propagation delay (*automatic acknowledgment*). However, in the terrestrial multi-hop packet radio networks of our concern, there are no such transponders. (Even if there were, transmitters could not hear the reflection because too short a propagation delay causes overlapping of the transmitted and reflected packets.)

In a store-and-forward operation of a multi-hop packet radio network, two types of acknowledgments are considered. In one case, a *positive acknowledgment* is returned by the receiver either as a stand-alone packet or as piggybacked on some other packet. A separate channel may be used for acknowledgment traffic. In the other case where CDMA is not used, the relaying of a packet by omnidirectional broadcasting from the receiver can serve as an implicit acknowledgment to the previous transmitter. This is called *echo acknowledgment.* In both cases, each packet is acknowledged one by one. If an acknowledgment is not received before a timeout occurs, the packet is retransmitted. Thus, these acknowledgment schemes correspond to the *stop-and-wait* data link control in wire-based networks. Grouped-acknowledgment schemes such as *go-back-n* will be inefficient in the ground radio environment where transmissions are very likely to be erroneous due to collisions.

Network Initialization

Consider a collection of geographically-distributed PRUs, each of which is powered on and capable of communicating packets to some subset of PRUs within its transmission range. To enable point-to-point packet transportation, each PRU must be identified by its relative location and be given routing information (e.g., the connectivity and quality of each link). The process of assigning such information (with *labels*) to PRUs is referred to as *network initialization.* Network initialization must be performed whenever the network resumes its operation from cold, or the network topology changes for various reasons (mobility of PRUs, decrease in transmission power, variation in received signal strength, etc.). Thus, the initialization algorithm and its efficiency are particularly significant in dynamic network operation.

There are several ways for network initialization. One way implemented in PRNET [Kahn78] is that each PRU periodically broadcasts a *radio-on packet* (ROP) announcing its existence and containing its identification and status. A subset of PRUs within the transmission range will hear this ROP and note the event in their tables along with the measured strength of the received signal. The quality of a radio link may be determined if each ROP contains the number of packets that the PRU has received from every other PRU it can hear. Upon receipt of such an ROP, a PRU can determine its connectivity and the percentage of packets successfully communicated on each link.

Some Markov chain models for initialization of a single-hop network have been proposed and analyzed in [Mino79b,Mino79c].

Routing

Routing algorithms for multi-hop packet radio networks may be classified in two categories; broadcast and point-to-point [Kahn78]. In *broadcast routing*, packets are transmitted without specifying the receiver address. Here, every repeater keeps a list of *unique packet identifiers* (UPIs) for previously broadcast packets in a certain period. If a repeater receives a broadcast packet whose UPI is already on its list, it will discard the packet. Otherwise, it will

accept the packet if correct, retransmit it and update the UPI list. Thus, copies of each packet radiate away from the source as in a wave-front type propagation. Broadcasting is not a particularly efficient mode for two-party communication, but it is a very robust way to distribute packets to all parts of the network.

In *point-to-point routing*, a packet originating at one PRU proceeds through a series of one or more repeaters until it reaches its final destination. The point-to-point route may be determined by a station which knows the correct connectivity of the whole network, or, in case there is no such a station, it may be discovered by the source PRU itself through broadcasting a *route-finding packet*. We have three modes of point-to-point routing; they are stationless, single-station and multiple-station operations.

In *stationless* operation, after a number of broadcast copies of the route-finding packet (through various paths with their delays recorded) have arrived at the final destination PRU, it will select the route with minimum delay. A *route-setup packet* is sent by the destination PRU which traverses the selected route in the reverse direction back to the source PRU while updating the routing tables at intermediate repeaters.

In a *single-station* network, hierarchical routing is proposed in [Gitm76]. In *hierarchical routing*, during network initialization, routing information called a *label* is assigned to every repeater by the station. The set of labels forms a hierarchical tree structure of repeaters rooted at the station. By use of these hierarchical labels, all traffic between any two repeaters is routed via the station. In another single-station routing, such as implemented in PRNET [West80], the station only provides labels for repeaters, and the repeater-to-repeater traffic does not go through the station, rather taking the shortest path between the two repeaters.

In a *multiple-station* network [Kahn78,Perl80], some repeaters may be assigned labels by several stations during initialization. Two stations which have labeled a common repeater are called *neighbors*. The route selection process takes place as follows. A PRU which generates packets for a destination outside the control of its local station sends a *route-request packet* to the station. The station then sends a copy of this packet, adding station ID and a list of traversed repeaters, to each neighboring station via some common repeater. When the packet is received by the neighboring station, it checks to see if the specified destination PRU is under its control. If not, it again relays the route-request packet in the same fashion. In this way, one or more route-request packets will eventually arrive at a station which has labeled the final destination PRU. The destination station then passes a complete list of repeaters to the destination PRU which initiates the route setup procedure.

Delay

The time that a packet in transit sojourns (i.e., occupies a position in a store-and-forward buffer) begins at the instant of its successful reception and ends when its transmission is acknowledged (either explicitly or via echo acknowledgment). This *forwarding delay* comprises the following times each associated with different actions which may occur

overlapped in time and an indefinite number of times: (i) the time to return an acknowledgment to the upstream PRU, (ii) the time to process the packet header for routing, (iii) queueing time, (iv) the packet transmission time and the time required for its propagation, (v) the randomized initial and retransmission delays, and (vi) the time to wait for and process an acknowledgment from the downstream PRU.

Picking an example from PRNET (see Section 1.1.3), we have packets with length 2,000 bits transmitted on a 400 Kbps channel over a line-of-sight distance of 32 miles (50 km). Then the packet transmission time is 5 msec., while the propagation delay is 0.17 msec. In IPR CAP6 (a protocol used for PRNET; see Section 1.1.3) [Jubi81], the random retransmission delay is multiples of 640 μsec., and the minimum header processing time is approximately 5 msec. The time until successful transmission (for ALOHA) or the time to wait for the idle channel (for CSMA) depends on the congestion around the PRU.

The *end-to-end delay* of a packet is the time required for the packet to move from its source PRU to its final destination PRU. The mean end-to-end delay over all packets is an important performance measure of the network. By use of Little's result [Litt61], this can be given as the sum of the average queue lengths over all PRUs divided by the mean number of delivered packets per unit time.

Flow Control

In [Gerl80], four main functions of flow control in wire-based packet-switching networks are given: they are (i) prevention of throughput degradation and loss of efficiency due to overhead, (ii) deadlock avoidance, (iii) fair allocation of resources among competing users, and (iv) speed matching between the network and its attached users. When we consider flow control in the context of a packet radio network, item (iv) may not be the case since each PRU's transmission/reception speed itself constitutes the network speed. Instead, the control of *channel stability* inherent in random access protocols must be taken into account. It is well known in single-hop, infinite population systems that an improper packet retransmission strategy can saturate the channel with continually retransmitted packets (the number of backlogged users grows indefinitely). For a finite number of users, the queue length in the buffer becomes infinite. A similar phenomenon can be expected in a multi-hop environment. Thus, some channel-stabilizing policy is needed.

In the study of a two-hop centralized network [Toba80a], it is the found the (optimized) system is mostly *channel bound* as opposed to *storage bound* (in terms of PRU buffer capacity). This observation suggests that the utilization of each PRU's buffer can be a good indicator of congestion in the network. Thus, some local (i.e., exercised by each PRU's decision) flow control strategy (at the network access level) based on the PRU buffer utilization appears to be useful in multi-hop networks. A qualitative discussion on symptoms of congestion is given in [West79].

Mobile Operation

A *mobile packet radio network* is capable of providing communications between mobile terminals (autos, hand-held devices for infantry, etc.) via mobile repeaters on vans, ships or airplanes to a (potentially mobile) computer center or monitoring station. Broadcast routing may be used to communicate with high-speed users to bypass the need for control of rapidly changing routes. A significant problem associated with the use of point-to-point routing in a mobile network is that the radio link between a mobile terminal and the repeater currently servicing it can disappear and reappear according to their relative locations and the terrain. This connectivity change will be announced via ROPs so that the station can keep track of mobile units. Thus arises a design question: what is the optimal rate of ROP emission as a function of topology and mobility characteristics? In order to approach this problem, one must first quantify the notion of connectivity and then establish the procedure to compute it given the (deterministic or stochastic) motion of PRUs. Analysis of such connectivity is addressed in [Mino78,Mino79a].

1.1.3 Some Existing Ground Radio Networks

Let us look at some practical ground radio networks developed in several places. Our primary concern includes structure of the network, conflict resolution capability in channel access protocol, handling of mobile environment, routing, etc. The contents of this section are heavily quoted from referenced literature.

Advanced Mobile Phone Service (AMPS) [Ehrl79]

The service trial of AMPS, developed by Bell Laboratories, was conducted in the Chicago area covering approximately 2,100 square miles with about 2,000 subscribers from 1977. AMPS is a cellular system controlled centrally by the *Mobile Telephone Switching Office* (MTSO) which is connected with wire to land radio stations, called *cellsites,* located at the center or corner of each hexagonal cell. Radio paths, employing frequency modulation (FM), are provided between a cellsite and radio-telephones (called *mobiles*) in the cell (thus, single-hop in radio). Transmission from mobiles to the cellsite uses channel frequencies between 825 and 845 MHz; from cellsites to the mobiles, frequencies between 870 and 890 MHz are used. By taking advantage of the FM capture effect, the same frequencies can be allocated to several distant cells (spatial reuse of the radio spectrum).

Two types of radio channels are provided for each cell. The *setup channel,* common to all mobiles in a cell, is used for initiating or setting up phone calls. The *voice channel,* providing the talking path, is assigned to each call by MTSO from unused channels within the cell (thus, circuit-switching and collision-free). At intervals, a mobile scans the setup channels to determine whether its movement has made the setup channel of another cell more appropriate for its use. When the mobile drives out of the coverage area of one cell, a *hand-off procedure*

(returning to a new setup channel and reassignment of a new voice channel) takes place and the telephone connection is rerouted through the cell it has just entered. These monitoring and controlling signals via the setup channel are interleaved with conversation, but they are so brief that most customers are unaware of their occurrence. The cellular structure is expandable in coverage area by adding cells and is capable of handling high traffic densities by dividing the cells.

Ptarmigan [Laws80]

Ptarmigan is a military communications system designed to replace the Bruin network currently in use by the British Forces assigned to NATO. The network components are frequently moving, and therefore a major aspect of the system is techniques to achieve consistency of service in the face of mobility.

The communication network of Ptarmigan consists of static and mobile *subscribers* and *switching nodes*. Two types of switching nodes are distinguished: *access nodes* for interfacing the subscribers, and *trunk nodes* inside the network. Both are mobile.

The routing algorithm is distributed and adaptive. First, a called subscriber is located by use of a *flood search message*. The path taken by the search message or its reply does not necessarily become the route of the call. Each node maintains an updated record of network connectivity from which it derives a *Route Preference Table* giving the preferred and alternative exit links to every other node in the network. From this table and taking into account the current traffic loading, the originating node selects a link to the next node on which to route the call. The node receiving the call message recomputes the preferred link in the same way. This procedure, called *delegated routing*, is followed by each node until the route is established between the originating and destination nodes. When the called subscriber is wanted again within a reasonable period, the flood search will occur only if it has moved to a new node (thus, mitigating the overhead due to flooding).

The radio access of mobile subscribers to the network is maintained by the *affiliation/re-affiliation procedure*. In order to obtain access, a subscriber must be affiliated, that is, identified and accepted by the system. Afterwards, he is given a continuous visual indication that he is in proper contact. When contact is lost the subscriber initiates an automatic affiliation search. To this end, the receiver tunes to the three stored frequencies, locking first for a strong signal, then for a good signal, and finally for the complete frequency band. This re-affiliation step completes whenever an appropriate signal is found.

Design considerations for another military network, the **U.S. Army Battlefield Information Distribution (BID)** are given in [Nils80].

High Frequency Intra-Task Force (HF ITF) Network [Bake81]

The HF ITF network is intended to be a general purpose network providing extended line-of-sight (50-1,000 km) communications for Naval task force units. The nodes in this network will be linked via radio waves from the HF (2-30 MHz) band. In the referenced paper, the architectural organization of mobile nodes is shown, and a distributed algorithm, called the *linked cluster algorithm*, that can establish (from any initial node configuration) and maintain (for any node motion and/or failure patterns) the connectivity is proposed.

The HF ITF network is organized into a set of *node clusters*, each node belonging to at least one cluster. Every cluster has its own *cluster head* which acts as a station for the nodes in that cluster. The cluster heads are linked via *gateways* (if needed) to provide paths for inter-cluster communication and global network connectivity. (The cluster heads and gateways form a *backbone network*.) Thus, a node becomes either an ordinary node, a cluster head, or a gateway at the completion of the linked cluster algorithm. A critical assumption of this algorithm, which uses TDMA for transmission of control messages, is that each node must know the number of nodes in the network.

Packet Radio Network (PRNET) [Kahn78,Quil79,Kunz81]

The development, implementation and operation of PRNET has been taking place under the sponsorship of Defense Advanced Research Projects Agency since 1973. Experimental testbeds are in the San Francisco Bay Area and at Fort Bragg, North Carolina. Involved in the working group are SRI, Rockwell International, and Bolt Beranek and Newman, Inc. PRNET provides for geographically separated fixed and mobile digital terminals to reliably communicate with each other.

Each user is connected into PRNET through a local packet radio. The initial radio equipment was designated the *Experimental Packet Radio* (EPR). It was upgraded to the *Improved Packet Radio* (IPR) for increased performance and enhanced *electronic counter countermeasure* (ECCM). An EPR consists of a radio unit which transmits and receives packets, and a digital unit which controls the radio unit and provides packet header processing.

The EPR *radio unit* operates with a fixed PN spread spectrum pattern which is identical for each transmitted bit. Thus, delay capture and CDMA are not available with EPR. (The IPR radio unit uses a direct sequence PN spread spectrum waveform which changes from bit to bit for antijamming and antispoofing purposes. The PN sequences with low cross-correlation provide a capture mechanism and support CDMA.) Two transmission data rates are available, 100 Kbps and 400 Kbps, with corresponding PN spread spectrum patterns of 128 and 32 chips per bit, respectively. The PN-modulated stream is then applied to a minimum-shift keying (MSK) modulator, and its output is up-converted to a selected 20 MHz portion of the 1,710-1,850 MHz band, power amplified, and transmitted through an azimuthally omnidirectional antenna. The EPR radio unit operates in a half duplex mode. When not transmitting, it remains in receive mode. With normal antenna height of 6 feet and line-of-sight condition,

EPRs can communicate over the radio horizon of about 32 miles.

The EPR *digital unit* uses a National IMP-16 microprocessor with 4,096 16-bit words of RAM and 1,024 words of PROM. (The IPR employs two TI-9900 CPUs.) Implemented protocols are the channel access protocol (CAP) for inter-EPR traffic, the end-to-end station-to-PR protocol (SPP) for monitor and control packets, a statistics gathering feature (CUMSTATS), and a debugging package (X-RAY).

CAP is responsible for the primary function of transferring packets to and from the adjacent EPRs on a route, monitoring the hop-by-hop echo acknowledgment, retransmission of unacknowledged packets, alternate routing, and determining packet disposition. CAP implements pure ALOHA, CSMA, and a variant of pure ALOHA (called *disciplined ALOHA*) in which random transmission is deferred until the end of an ongoing reception. The CAP 4 for a single-station network was revised to CAP 5 in order to handle mobile environments more efficiently. The centralized routing algorithm used by CAP 5.6 Labeler [West80] resident in the station is a modified *Floyd's algorithm* for the shortest path with a two-level distance measure (*hop count* and a multiplication of *link qualities*).

CAP 6 implementation for multiple-station routing has the following design objectives [Perl80]. (i) It should accommodate a fairly mobile network of up to thousands of PRUs. — The network is divided into subnets, each being under control of a station. (ii) Station failure must not disrupt the network. — Many stations are provided so that each PRU should be in at least two (on the average three) subnets at all times. (iii) Control traffic must be minimized. — Within a subnet, the station controls routing centrally. (iv) Stations do not have the capability to routinely forward traffic. — Their job is to receive routing information and hand out the routes. (v) Connectivity is changing so quickly that the subnet membership is only stable on the order of time for a route setup procedure to work. Stationless routing is implemented in CAP 7. The IPR Level 6 Protocols are documented in [Jubi81].

An EPR packet consists of a 48-bit *preamble* followed by a maximum of 2,048 bits of *header* and *text*, and a 32-bit *cyclic redundancy checksum*. The preamble is used by the radio unit of the receiving EPR to detect the carrier energy, to set the *automatic gain control* (AGC) to compensate for differing signal strengths of arriving packets, and to acquire packet timing. Correct reception of the packet is totally dependent upon acquisition of the preamble. If the preamble portions of two or more arriving packets overlap, the AGC and packet synchronization are likely to fail. A variable-length header (typically 96-144 bits) contains the traversed hop count, header length, packet length, source ID, destination ID, packet sequence No., SPP transmit count, packet type, acknowledgment flag, and so on. The checksum is appended by the transmitter and checked by the receiver. The IPR is provided with a *forward error correction* mechanism which operates in combination with error detection and retransmission techniques.

1.2 Outline of the Dissertation

1.2.1 Models, Measures and Methodologies

This dissertation is devoted to the theoretical performance evaluation for packet radio networks. Throughout the dissertation, we assume that all the packets are of constant length, and choose their transmission time as the unit of time. Acknowledgments (in response to transmission) are always assumed to be given for free, correctly and immediately. We also assume a noiseless channel so that the only reason for an unsuccessful transmission is the more or less simultaneous (depending on the capture characteristics) arrivals of multiple packets at the receiver.

We focus our study on random-access protocols such as ALOHA and CSMA in both slotted (i.e., discrete-time) and unslotted (or pure) versions. We also consider some variations including the effects of FM and delay capture and collision detection on the performance. In CSMA, we usually take into account the signal propagation delay. Also in CSMA, we distinguish the protocol in which the packets that arrive to find the channel busy are buffered until the next transmission chance (persistent CSMA) from the protocol in which they are discarded (nonpersistent CSMA). Thus we can think of the combinations of all theses features in a specific protocol (for example, unslotted persistent CSMA with collision detection). As for the hearing topologies, we consider the fully-connected case and the hidden-user case in single-hop systems, and the case of general multi-hop systems.

Our performance measures are mainly the throughput (system-wide, per-user and per-route) and the average delay (one-hop and end-to-end). We are also interested in the higher moments of the distribution of the packet interdeparture times (i.e., the intervals between two successive successful transmissions). In the multi-hop environment, we are concerned with the optimal transmission radius which maximizes the expected per-hop progress of packets towards their final destinations.

Our major tools to analyze the performance for the above-mentioned models are primarily borrowed from the theory of probability and stochastic processes. In particular, we use renewal and regenerative arguments, sums of independent random variables, superposition of independent renewal processes, semi-Markov process, the diffusion process approximation, and Markov chains. Simulations have been conducted to check the validity of our approximations.

1.2.2 Summary of the Results

Channel Throughput for Single-Hop, Finite-Population Systems [Chapter 2]

Most of past research for the throughput evaluation of single-hop systems has assumed an infinite population of users. This assumption may be inappropriate in a multi-hop environment since each user has only a limited number of communicating neighbors due to the finite transmission radius. We have studied the throughput for a finite population model, and expressed the results as explicitly as possible. Channel access protocols we have considered here include pure ALOHA, pure ALOHA with delay capture, slotted persistent CSMA, slotted persistent CSMA with collision detection, unslotted persistent CSMA, and unslotted persistent CSMA with collision detection. (These are the protocols for which the packet interdeparture time is either difficult to find (pure ALOHA) or not independent and identically distributed (persistent CSMA) so that we could not treat them using the models in Chapter 3.)

In our models, due to the assumption of exponentially or geometrically distributed idle (i.e., non-transmitting) periods at each user, the intervals between two successive epochs at which the system enters the idle period are shown to be independent and identically distributed. Therefore, the system state can be modeled as a regenerative process. We have calculated the expected number of successful transmissions in a regenerative cycle which is the channel throughput.

Packet Output Processes for Single-Hop, Fully-Connected Systems [Chapter 3]

For ALOHA and nonpersistent CSMA (and their variations), we have analyzed the packet interdeparture time in a single-hop, fully-connected environment for both finite and infinite-population models. (Taking the reciprocal of the mean interdeparture time, we get the channel throughput.)

Through the analysis of channel activity cycles alternating between idle and (successfully or unsuccessfully) transmitting states, we have exactly derived the distribution (in terms of the pdf's Laplace transform) for the packet interdeparture time X. Then we have calculated the channel throughput ($S = 1/\bar{X}$) and the coefficient of variation of X ($C^2 = Var[X]/\bar{X}^2$) explicitly. It has been shown that in efficient CSMA systems with collision detection or with delay capture, $C^2 \simeq 0$ while $S \simeq 1$. The cases where users have different transmission parameter values have also been analyzed. Using $\bar{X}$ and C^2 together with the elementary renewal theorem, we have obtained the asymptotic behavior of the number of successful transmissions for individual users.

Approximate Packet Output Processes for Single-Hop, Hidden-User Systems [Chapter 4]

We have given an approximate analysis for the packet departure processes in a hidden-user environment of single-hop packet broadcasting systems. The channel access protocols considered include pure ALOHA, and unslotted and slotted CSMA. The effect of (perfect) delay capture on unslotted CSMA has also been evaluated.

An exact stochastic analysis has been given for the duration of a channel idle period, a successful transmission period, and an unsuccessful transmission period consisting only of those packets from the users who can hear the initiating transmission. An approximate analysis has been developed for the duration of ar unsuccessful transmission period involving hidden users' packets. Our approximation is based on the theory of superposition of independent renewal processes, together with a proper reduction of transmission start rates to take care of carrier-sense effects.

The channel throughput and the coefficient of variation of the packet interdeparture time calculated by use of our approximation have been compared with simulation results in symmetric and wall configurations for a variety of degrees of hiddenness. The agreement between them is excellent in the symmetric hidden-user configurations (without delay capture) for almost the whole range of offered channel traffic and all reasonable values of propagation delay. For wall configurations and symmetric hidden-user configurations with perfect delay capture, the agreement is good until the offered traffic value exceeds its optimum which gives maximal channel throughput.

Packet Queueing Delay for Single-Hop Systems [Chapter 5]

For a finite number of users each with an independent packet arrival stream and an infinite buffer, we have studied exact results, bounds and approximations for the average packet delay (including queueing and (re)transmission(s)).

First, the exact analysis for two identical slotted ALOHA users in [Sidi83] has been applied to slotted CSMA with collision detection to find explicitly the mean packet delay. Then, for the case of more than two users of slotted ALOHA with Bernoulli arrivals, some upper bounds on the mean delay have been obtained.

For a general random-access system, we have formulated a diffusion process approximation to the joint queue length distribution, and have solved the stationary diffusion equation with reflecting boundary conditions. Here, the first two moments of the distribution for the packet interdeparture times (found in Chapters 3 and 4) are used to determine the coefficients in the diffusion equation. Based on the expression for the mean queue length, we have proposed two formulas for the mean packet delay, and have shown their good agreement with simulation results in several cases.

Optimal Transmission Ranges for Randomly Distributed Terminals [Chapter 6]

We have solved for the maximum expected progress per hop (Z), provided by the optimal transmission probability (p) and transmission radius (expressed in terms of the number of terminals in the range, N), for some models of randomly distributed packet radio terminals (with average density λ) under the assumption of heavy traffic (all terminals always have ready packets). The quantity $Z\sqrt{\lambda}$ has been used consistently as the dimensionless objective function for optimization problems with respect to p and N. Major conclusions about the performance of each model are as follows:

The optimal transmission with slotted ALOHA without capture is attained by $N=7.72$ and $p=0.113$ which gives $Z\sqrt{\lambda}=0.0431$. Therefore, each terminal transmits once in every nine slots on the average with the transmission radius covering just about eight nearest neighbors in the direction of the packet's final destination. The probability of success of such a transmission is nearly equal to $1/e$. The expected progress per transmission is about two thirds of R/e, where R is the optimal transmission radius ($N=\lambda\pi R^2$).

FM capture improves the performance of slotted ALOHA systems due to the more limited area of possibly interfering terminals around the receiver. The expected progress in a system with perfect capture (optimized with $N=7.1$ and $p=0.17$) is about 36% greater than that in the system without capture. The probability of successful transmission is also higher than $1/e$. A model which is more amenable to implementation (each terminal knows the positions of only a fixed number of its neighbors) has shown similar results.

Slotted nonpersistent CSMA provides a nominal improvement in performance over the ALOHA system (16% improvement in the optimized expected progress for zero propagation delay), which is not as large an improvement as we have obtained in the single-hop case. The reason for this is the large area of 'hidden' terminals (about half of the interfering area) which cannot hear the transmission, and the long vulnerable period (twice as long as the packet transmission time) due to those terminals. The performance of (slotted nonpersistent) CSMA is comparable to that of ALOHA with good FM capture (capture ratio about 1.5 *dB*). The degradation occurs as the ratio of propagation delay to the transmission time increases.

As an example of an inhomogeneous terminal distribution, the effect of a gap of width b in an otherwise uniformly Poisson-distributed terminal population on the optimal transmission has been considered. The expected progress of a packet residing in the terminal on the bank and destined to cross the gap is evaluated with parameter $\beta=\lambda b^2$, called the *gap intensity*. For fixed λ, the existence of the gap helps the progress for $\beta<2$, because some of the possibly interfering terminals are removed by the gap. The maximum in the optimized expected progress occurs at about $\beta=1$. Thus, to cross most gaps wider than the average inter-terminal distance, one had better not use a large transmission radius, but should more sensibly use a separate channel or wire.

End-to-End Packet Delay in Multi-Hop, Slotted-ALOHA Networks [Chapter 7]

We have analyzed the throughput-delay characteristics for slotted-ALOHA multi-hop packet radio networks where the hearing configuration of packet radio units (terminals and repeaters) and source-to-sink paths of packets are given and fixed. The problems are formulated as discrete-time Markov chains and then solved numerically.

Besides the basic model — characterized by isolated transmission behavior and single-buffered repeaters — three ways to improve the throughput-delay performance have been exploited. They include (i) transmission suppression when the destination's buffer is occupied, (ii) transmission acceleration when the buffers of all neighbors of the destination are empty, and (iii) multiple buffers for repeaters.

It has been shown that the transmission suppression scheme provides a natural flow control at the network access level to prevent packets from entering the 'communication subnet.' This brings about significantly lower delay for a given throughput, and achieves a much higher maximum throughput. Transmission acceleration combined with appropriate suppression gives further improvement in the throughput-delay trade-offs at the cost of necessitating more information about the network state.

With more than one buffer for repeaters we have fewer chances of failure of transmission due to a buffer shortage at destinations. It has been shown that increasing the number of buffers from 1 to 2 offers more performance enhancement than going from 2 buffers to 3 buffers. The effect of transmission suppression/acceleration in the multibuffer case was also demonstrated.

1.3 Survey of Related Work

This section provides a short survey of previous and current theoretical work which the study of this dissertation is based on and/or related with. Exhaustive reference is not intended.

1.3.1 Single-Hop Systems

Throughput for ALOHA

The throughput for an infinite population model of pure ALOHA was first presented in [Abra70]. The throughput for an infinite population model of slotted ALOHA with and without FM capture is in [Robe72]. [Abra73b] considered the channel and user throughput for a finite population model of slotted ALOHA. [Ferg77b] showed the throughput analysis for a finite population model of pure ALOHA. The effect of FM capture on the throughput at the central receiver surrounded by a continuum of slotted ALOHA users was addressed in

The coverage of each chapter in the following is (conceptually) tabulated in Table 1.1. Here, we also show whether the treatment is exact or approximate. Each chapter has been written as self-sufficient to allow the reader to selectively pick his favorite.

Table 1.1 Coverage of each chapter

<table>
<tr><th></th><th>THROUGHPUT/OUTPUT PROCESS</th><th>DELAY</th></tr>
<tr><td>single-hop</td><td>Chapter 2 (exact)
Chapter 3 (exact)</td><td rowspan="2">Chapter 5
(exact, bounds and approximations)</td></tr>
<tr><td>hidden-user</td><td>Chapter 4 (approx.)</td></tr>
<tr><td>multi-hop</td><td>Chapter 6 (approx.)</td><td>Chapter 7 (exact)</td></tr>
</table>

In this dissertation, we study a number of channel access protocols with their variations. Table 1.2 below indicates chapter numbers referring to each protocol model.

Table 1.2 Chapter reference to protocol model

Protocols	No Option	Collision Detection	Delay Capture ('FM Capture)
Pure ALOHA	2,3,4,5		2
Slotted ALOHA	3,5,6,7		6*
Slotted CSMA (nonpersistent)	3,4,6	3	
Slotted CSMA (persistent)	2	2,5	
Unslotted CSMA (nonpersistent)	3,4,5	3	3,4
Unslotted CSMA (persistent)	2	2	

[Abra77], where the existence of *Sisyphus distance* (the users beyond that distance can never get to the receiver) was shown in the case of perfect capture.

Throughput for CSMA

The introduction of various CSMA protocols and their throughput analysis for an infinite population of fully-connected users is in [Toba74(Chapter 3),Klei75b]. This is followed by the approximate throughput analysis in the hidden-user case in [Toba74(Chapter 4),Toba75]. Later, the throughput of CSMA with collision detection for an infinite population model (slotted nonpersistent and 1-persistent cases) was given in [Toba80d]. A finite-population model was considered recently by [Jenq82] and [Arth82].

Mean Retransmission Delay for Slotted ALOHA

The first throughput versus mean packet delay curves for an infinite population model of slotted was drawn in [Lam74,Klei73]. Recently, this analysis was extended to the case involving FM capture [Shac82].

Linear Feedback Models

For a finite number of users each with one-packet buffer, it is possible to formulate a finite Markov or semi-Markov process where the state is identified as the number of nonempty users. This is called a *linear feedback model,* and has often been used to get the mean packet delay versus throughput relationship (obtained from the mean number of nonempty users by use of Little's result [Litt61]) as well as to discuss the channel stability. Examples are: slotted and pure ALOHA in [Carl75], slotted ALOHA in [Lam74,Lam75], slotted ALOHA with delay capture [Davi80], slotted CSMA in [Toba74,Toba77], assigned-slot CSMA in [Hans79], slotted CSMA with collision detection in [Toba80d] and unslotted CSMA in [Heym82b].

Packet Queueing Delay in Slotted ALOHA

The problem of finding the average packet delay (including queueing and retransmission delays) for a finite population of slotted ALOHA users each of whom has infinite packet buffer has been attacked in [Toba76,Yemi80,Saad81]. Only recently, [Sidi83] found an explicit mean delay in the case of two identical users.

Dynamics on the Number of Nonempty Users

For an infinite number of users each with one-packet buffer, one can consider the dynamic behavior of the number of nonempty users. [Koba77] developed the diffusion approximation to this system, while [Fuku78,Tasa82] showed that the steady state can be approximated by the intersection of the load line and the throughput curves (EPA — *equilibrium point analysis*). [Lam80] considered the case of CSMA with collision detection by an approximate M/G/1-like model.

1.3.2 Multi-Hop Systems

Mixing ALOHA and Dedicated Channels

In [Akav79], the packet traffic is characterized by the dimensionless quantity ST, where S is the traffic to be carried and T is the acceptable delay; when $ST >> 1$ the traffic is called *steady*, and when $ST << 1$ it is called *bursty*. It is shown that a two-level mixed channel access protocol, using ALOHA for intra-group communications and dedicated channels for inter-group traffic, can be optimized to give the maximal throughput for medium burstiness.

Optimum Transmission Radius for Slotted ALOHA Networks

[Klei78,Silv80] made the following observation. Choosing the optimal transmission range for geographically distributed PRUs involves the trade-off: a long range enables packets to reach their destinations in a few (successful) hops, but increases the probability of collision at every hop. They showed for a planar Poisson distribution of PRUs that the optimal radius is such that the average number of PRUs covered in a transmission radius is about 6. Their analysis was extended in [Nels82] for a system with FM capture.

Capacity of the Tandem Network

In [Yemi80], the capacities (maximum-achievable throughput values) of the two versions of slotted ALOHA in an infinite series of PRUs are determined. The *polite policy* (the service rate at each PRU is guaranteed to be faster than the arrival rate) gives 4/27, while the synchronized *rude policy* (the transmission probability 1 is given to each PRU every third slot) yields 1/3 without any collisions.

Capacities for Various Channel Access Protocols

The network capacities with the following channel access protocols are studied in [Nels82]: slotted ALOHA with FM capture, rude CSMA and spatial TDMA. In the study of the FM capture effect on the performance of slotted ALOHA, the transmission randomizing parameter is optimized to maximize the throughput. This analysis is extended to find a tight bound on the performance of all protocols in a mobile environment. In conventional CSMA, a PRU transmits only when it senses the channel idle. In a multi-hop environment, this is sometimes wasteful because the channel might be idle at an intended receiver. Thus, in *rude CSMA*, a PRU may transmit even when it senses a busy channel, depending on the estimate of the number of the neighborhood transmitters. Transmission policies to maximize throughput are studied. *Spatial TDMA* is the assignment of time slots to geographically distributed PRUs so as to avoid collisions. Existing graph coloring algorithms are utilized. The criteria of optimization are maximal throughput as well as minimizing the number of distinct 'colors'.

Comparison of Acknowledgment Schemes

In [Elsa82], the effect of acknowledgment schemes on the performance of a star-shaped broadcast network (centered at the single repeater) is studied. The network capacity and the average delay at the repeater are evaluated (by use of a Markov model) for the following four schemes: (i) free and instantaneous acknowledgments, (ii) common channel for data and acknowledgments, (iii) split channel for data and acknowledgments, and (iv) common channel with high-power acknowledgments. It is shown that the split channel achieves the highest capacity, and that the high-power acknowledgment scheme provides a good alternative when data packets are short.

Throughput-Delay Analysis of Two-Hop Centralized Network with Slotted-ALOHA and CSMA

In [Toba80a,Toba80b], the two-hop centralized configuration is such that traffic originating at terminals is relayed by store-and-forward repeaters to the central station. For the slotted-ALOHA protocol, the improvement of throughput-delay characteristics by increasing the number of repeater buffers beyond 2 is shown to be minor. Thus, the system is said to be *channel bound* as opposed to *storage bound.* The nonpersistent CSMA is shown to provide an improved capacity over slotted-ALOHA especially when the number of repeaters increases.

Approximate Evaluation of Mean Packet Delay

[Lein80] proposed a numerical procedure to approximately compute the mean end-to-end packet transportation delay for given source-to-destination traffic rates in a general multi-hop packet radio network. This procedure utilizes the delay versus throughput curves obtained for single-hop systems for a variety of protocols (e.g., the ZAP approximation in [Klei77]), and iteratively adjust the traffic level.

Product-Form Solution

Assume in unslotted CSMA the following: (i) zero propagation delay, (ii) perfect delay capture, (iii) exponentially distributed packet length, and (iv) Poisson arrival of new and retransmitted packet transmissions whose durations are sampled each time afresh. Under these conditions, [Boor80] showed the existence of product-form solution to the joint probability of buffer-occupancy for all users in multi-hop networks. This model was extended in [Toba82b] to other protocols such as pure ALOHA, BTMA and CDMA.

Iterative Assessment of Queue Interference

A series of papers [Lee82,Silv82a,Silv82b,Silv82c] show an iterative approach to assess the interaction among users in single- and multi-hop environment. The iteration consists of finding the parameterized probability of buffer occupancy at each user and computing the interactions among nonempty users. Pure and slotted ALOHA and CSMA have been considered.

CHAPTER 2
Throughput Analysis for Finite-Population Packet Broadcasting Systems

The channel throughput for a finite number of packet broadcasting users is analyzed newly for six random access protocols including, for example, pure ALOHA slotted persistent carrier-sense-multiple-access (CSMA) with and without collision detection, and unslotted persistent CSMA with and without collision detection. We consider both h- and 1-persistent CSMA. The effect of delay capture on the throughput in pure ALOHA is also studied.

2.1 Introduction

One class of multi-access protocols for packet communication systems is the random access (or contention) technique where the entire bandwidth is provided to the users as a single channel to be accessed randomly. (For a classification of multi-access protocols, see for example [Toba80c].) Channel access protocols such as ALOHA [Abra70] and carrier-sense-multiple-access [Toba74,Klei75b] are included in this category. Here one of the basic measures for the efficiency of protocol is the throughput, i.e., the average fraction of time that the channel is used for useful data communication. (Three factors accounting for the throughput degradation are propagation delay, user's idle (not transmitting) period, and packet collision (overlapping of transmissions from multiple users) inherent in the random access.)

This chapter focuses on the throughput analysis for random access protocols for a finite number of users (transmitters) and a single receiver in line-of-sight of all users. Thus, our study is applicable to ground packet radio systems as well as local-area computer networks which consist of a relatively small number of users. (A system with a relatively large number of users can be approximated by an infinite population model.) The analysis for finite-population systems may also serve as the first step to approach the performance modeling of multi-hop packet radio networks (such as PRNET described in [Kahn78]) where each user has only a limited number of communicating neighbors.

A number of studies on the throughput analysis for random access protocols have already appeared in the literature. However, most of them have been based on the assumption that there are infinitely many users such that the collective channel traffic forms a Poisson process with a finite rate. One exception was a paper by Abramson [Abra73b] in which he studied a finite population model of slotted ALOHA. In Chapter 3, we will derive not only the channel throughput but also the distribution of packet interdeparture times (i.e., the intervals between two consecutive successful transmissions) for the following random access protocols (in both infinite and finite population models): slotted ALOHA, slotted nonpersistent CSMA with and

without collision detection, unslotted nonpersistent CSMA with and without collision detection and delay capture. In these protocols, the successive interdeparture times are independent and identically distributed, and the channel throughput is simply the reciprocal of the mean interdeparture time.

The channel access protocols we consider here are:

a. pure ALOHA [Abra70]

b. pure ALOHA with delay capture

c. slotted h- and 1-persistent CSMA

d. slotted h- and 1-persistent CSMA with collision detection

e. unslotted h- and 1-persistent CSMA

f. unslotted h- and 1-persistent CSMA with collision detection

(Each protocol model is described individually below. We adopt h to denote persistence in CSMA since the usual p is used for the arrival rate.) Previous work based on the infinite population model for the subset of the above-mentioned protocols includes [Abra70] (pure ALOHA), [Toba74,Klei75b] (slotted h- and 1-persistent CSMA, and unslotted 1-persistent CSMA), and [Toba80d] (slotted 1-persistent CSMA with collision detection). In this chapter, we give an exact throughput analysis in the case of a finite population for all of the above-listed protocols. Arthurs and Stuck [Arth82] also provide an analysis of throughput for slotted and unslotted persistent CSMA with collision detection based on models different from ours. (A difference between our model and theirs involves the way in which one disposes of a packet, the transmission of which is suppressed as a result of sensing a busy channel. In our model, similar to [Toba74,Klei75b], they are dismissed from the system at the beginning of the next transmission period, while in [Arth82] they are retained in the buffer until successfully transmitted; so 1-persistent CSMA is impossible in the [Arth82] model.)

2.2 Throughput Analysis

In dealing with the case of a finite population, we *assume* that each user has periods which are independent and exponentially or geometrically (depending on whether the time is continuous or slotted) distributed in which he has no packets. By superimposing these idle periods over all users, the system idle period in which no users have a packet (denoted by I) is easily seen to also be exponentially or geometrically distributed. This assumption makes analysis tractable by taking advantage of the memoryless property [Klei75a,p.66]. The case of an infinite population does not need a specific assumption on the distribution of each user's idle period because the Palm-Khinchine theorem (see [Heym82a]) guarantees that the collective

idle period is always independent of one another and exponentially or geometrically distributed.

Due to the above assumption, we recognize that each epoch in the system idle period is a regenerative point in the sense that the system state after any such epoch is a probabilistic replica of the system state beginning at the previous such epoch. Thus the system state alternates between idle periods, I, and busy periods, B, in which at least one user has a packet. We call a consecutive pair B and I a *regeneration cycle*. Let U be the time spent in useful transmission during a regeneration cycle. Then, the channel throughput, S, is generally expressed as

$$S = \frac{\overline{U}}{\overline{B} + \overline{I}} \tag{2.1}$$

where $\overline{X}$ denotes the expectation of a random variable X, i.e., $\overline{X} \triangleq E[X]$.

Throughout this and other chapters we assume a constant packet length whose transmission time is chosen as the unit of time. We denote by M the number of users, and by G the total packet arrival rate (in units of packets per packet transmission time). In CSMA we take into account the signal propagation delay denoted by a. (In ALOHA, we assume that the propagation delay is incorporated in the transmission time, and so we let $a = 0$.)

Each section below begins with a description of the protocol model followed by the definition of protocol-dependent parameters. The condition for a successful transmission is stated. Then we evaluate $\overline{I}$, $\overline{B}$ and $\overline{U}$ for the finite population model. Our result in the limit $M \to \infty$ is shown to be in agreement with the known results for an infinite population.

2.2.1 Pure ALOHA

In pure (or unslotted) ALOHA, any packet starts its transmission as soon as it arrives at a user, and is discarded by that user after the transmission. Thus, each of M users alternates between the idle (not transmitting) and busy (transmitting) states independently of all other users. Note that transmissions may be started at any time by each user. Let the duration of an idle state at user i be exponentially distributed with mean $1/g_i$ ($i = 1,2,\cdots,M$). The transmission duration is of length unity for all users. Without delay capture, the condition for a successful transmission is that there be no other overlapped transmissions during the entire transmission time of a packet. The derivation of the throughput expression here follows Ferguson [Ferg77b] although he does not give the result in Eq.(2.8) below explicitly.

We first look at an idle-busy cycle at user j. Since each busy period (of length 1) at user j is followed by an idle time of mean length $1/g_j$, we have

$$Prob\,[\text{user } j \text{ is idle}] = \frac{1/g_j}{1 + 1/g_j} = \frac{1}{1 + g_j} \tag{2.2}$$

Since all users behave independently, it follows that

$$P_0 \triangleq Prob\,[\text{common channel is idle}] = \prod_{j=1}^{M}\left(\frac{1}{1+g_j}\right) \tag{2.3}$$

Now, due to the renewal property of channel state, P_0 must equal the ratio of the average duration of channel idle period $\bar{I}$ to the average cycle time $\bar{B}+\bar{I}$, where $\bar{B}$ is the average duration of the common channel busy period:

$$P_0 = \frac{\bar{I}}{\bar{B}+\bar{I}} \tag{2.4}$$

Note that the duration of channel idle period is exponentially distributed with mean

$$\bar{I} = \frac{1}{\sum_{j=1}^{M} g_j} \tag{2.5}$$

Thus we have the average cycle length

$$\bar{B}+\bar{I} = \frac{\bar{I}}{P_0} = \frac{\prod_{j=1}^{M}(1+g_j)}{\sum_{j=1}^{M} g_j} \tag{2.6}$$

User i has a successful transmission when it breaks the channel idle period among other users and there are no transmissions from any others during its transmission time of length 1. Thus the probability of a successful transmission for user i is given by

$$U_i = \frac{g_i}{\sum_{j=1}^{M} g_j}\exp\left(-\sum_{\substack{j=1\\(j\neq i)}}^{M} g_j\right) \tag{2.7}$$

Therefore, by definition the throughput for user i, denoted by S_i, is given by

$$S_i = \frac{U_i}{\bar{B}+\bar{I}} = \frac{g_i \exp\left(-\sum_{\substack{j=1\\(j\neq i)}}^{M} g_j\right)}{\prod_{j=1}^{M}(1+g_j)} \qquad i=1,2,\cdots,M \tag{2.8}$$

The channel throughput S is given by

$$S = \sum_{i=1}^{M} S_i \tag{2.9}$$

Let us determine the boundaries of the region of maximum-allowable throughputs in the M-dimensional space $(S_1,S_2,\cdots,S_M)$. The boundary which defines the maximum for S_i while all other S_j $(j\neq i)$ are held fixed at allowable values is found (for all i simultaneously) by setting the Jacobian $J(S_1,S_2,\cdots,S_M;g_1,g_2,\cdots,g_M)$ equal to zero [Brve70]. (This Jacobian is the determinant of the $M\times M$ matrix whose i,j element is $\partial S_i/\partial g_j$.) After some algebra, we find that this condition reduces to

$$\sum_{i=1}^{M} \frac{1}{(1+g_i)^2} = M - 1 \tag{2.10}$$

When we assume that the M users form m groups (indexed as $1,2,\cdots,m$) such that the users in group k (group size $= m_k$) have the transmission rate g_k, the condition in Eq.(2.10) takes on the form

$$\sum_{k=1}^{m} \frac{m_k}{(1+g_k)^2} = \left(\sum_{k=1}^{m} m_k\right) - 1 \tag{2.11}$$

where $\sum_{k=1}^{m} m_k = M$. By some manipulation on Eq.(2.11), it can be shown that

$$g_k \leq \sqrt{\frac{m_k}{m_k - 1}} - 1 \qquad k = 1,2,\cdots,m \tag{2.12}$$

Further, if we define $G_k \triangleq m_k g_k$ and make $m_k \to \infty$ with fixed G_k ($k = 1,2,\cdots,m$), then since

$$\frac{m_k}{(1+g_k)^2} - m_k = \frac{m_k}{(1+G_k/m_k)^2} - m_k \simeq m_k\left(1 - \frac{2G_k}{m_k}\right) - m_k = -2G_k,$$

Eq.(2.11) reduces to

$$\sum_{k=1}^{m} G_k = \tfrac{1}{2} \tag{2.13}$$

which allows throughput of G_k/e for group k ($k = 1,2,\cdots,m$).

In Figure 2.1, we have plotted some of optimal throughput contours based on Eq.(2.11) when we divide the users into two groups. We note that the shapes of these contours are similar to their counterparts for slotted ALOHA shown in [Abra73b], although our contours are much more squeezed into the lower throughput region.

In the case of identical users ($g_1 = g_2 = \cdots = g_M \triangleq g$), Eqs.(2.8) and (2.9) give

$$S = \frac{gM\, e^{-g(M-1)}}{(1+g)^M} \tag{2.14}$$

which is optimized by

$$g = \sqrt{\frac{M}{M-1}} - 1 \tag{2.15}$$

For $M \to \infty$ while keeping $G = gM$ at a finite value, we have the well-known result by Abramson [Abra70]:

$$S = G\, e^{-2G} \tag{2.16}$$

whose maximal value is $1/(2e)$ at $G = \tfrac{1}{2}$.

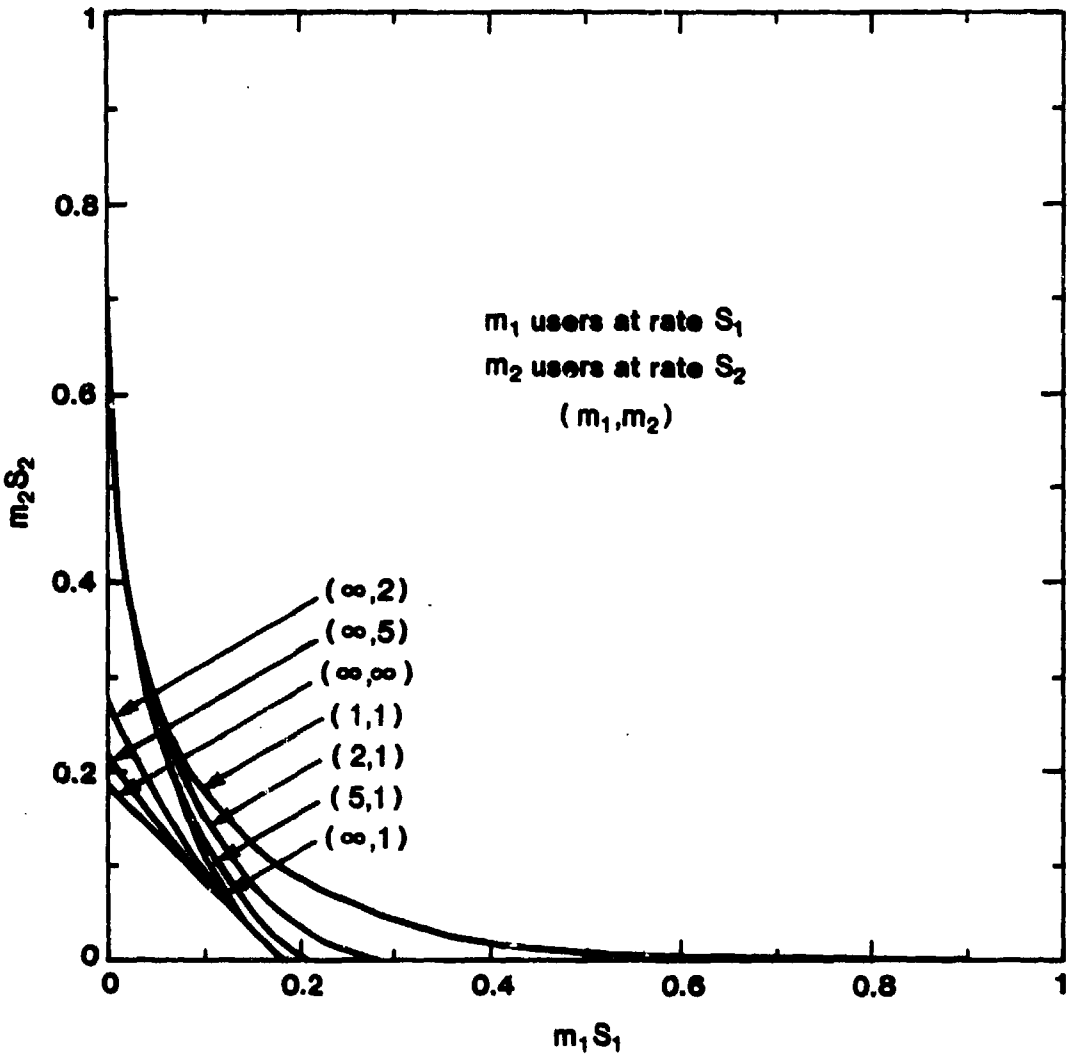


Figure 2.1 Allowable source rates for pure ALOHA.

We have mentioned that the channel cycle alternates between idle and busy periods. For the case of identical users, the average duration of an idle period is given by $\bar{I} = 1/(gM)$. A busy period containing a successful transmission is of length 1. We are now interested in the average duration of a busy period containing unsuccessful transmissions, denoted by $\bar{F}$. This can be found through Eq.(2.6) and the relationship:

$$\bar{B} = e^{-g(M-1)} \cdot 1 + [1 - e^{-g(M-1)}] \cdot \bar{F} \tag{2.17}$$

Thus, we get

$$\bar{F} = \frac{(1+g)^M - 1 - gM\, e^{-g(M-1)}}{gM\,[1 - e^{-g(M-1)}]} \tag{2.18}$$

For $g \to 0$, $\bar{F} \to 3/2$ (which consists of 1 for a transmission time and ½ for the average interval between the arrival times of two overlapping packets). For $M \to \infty$ with $G = gM$ fixed, we have

$$\bar{F} = \frac{e^G - 1 - G\, e^{-G}}{G(1 - e^{-G})} \tag{2.19}$$

(This is also derived directly in Appendix B.) At the optimal $G = ½$, we have $\bar{F} = 1.756$, $\bar{I} = 2$, $\bar{B} = (e^G - 1)/G = 1.297$, and $P_0 = e^{-G} = 0.6065$.

2.2.2 Pure ALOHA with Delay Capture

Delay capture, available with spread spectrum modulation, is the capability of a receiver to successfully receive a packet even when part of the packet arrives at the receiver overlapped in time by other packet transmissions [Kahn78]. To evaluate the capture effect, let us introduce a parameter c such that the transmission of a packet started by breaking the channel idle period is successful if there are no other transmissions during a time c. Therefore, the case $c = 1$ corresponds to pure ALOHA without capture, while the case $c = 0$ may be called *perfect capture*. In [Davi80], it is stated for a typical ground-based, spread-spectrum system that the capture time is 10 μsec and that the packet transmission time is 20 msec. In such a case, we have $c = 1/2000$.

In analyzing pure ALOHA with delay capture, we notice that the cycle of channel usage is identical to that of pure ALOHA without capture; only the probability of successful transmission is different. (Note when $c = 0$ that every packet started by breaking a channel idle period is successful.) Thus, similar to Eq.(2.8), the throughput for user i is given by

$$S_i = \frac{g_i \exp\left(-c \sum_{\substack{j=1 \\ (j \neq i)}}^{M} g_j\right)}{\prod_{j=1}^{M} (1 + g_j)} \qquad i = 1,2,\cdots,M \tag{2.20}$$

Corresponding to Eq.(2.10) we have

$$\sum_{i=1}^{M} \frac{1}{(1+g_i)(1+cg_i)} = M-1 \tag{2.21}$$

For the case of statistically identical users, the channel throughput is given by

$$S = \frac{gM\,e^{-cg(M-1)}}{(1+g)^M} \tag{2.22}$$

which, for $M \to \infty$ with fixed $G \triangleq gM$, reduces to

$$S = G\,e^{-(1+c)G} \tag{2.23}$$

It is interesting (amazing) to note that the throughput for the case of perfect capture ($c=0$) is identical to that of slotted ALOHA. Even with perfect capture, pure ALOHA cannot prevent users from starting a transmission in an already busy channel which elongates the period of unsuccessful transmission. In Figure 2.2, we show the values of S given by Eq.(2.22) when g is optimally chosen according to Eq.(2.21). We can here recognize the throughput enhancement brought by the capture effect. Roughly speaking, a factor of 2 difference between the cases $c=0$ and $c=1$ can be noted for $M \geq 2$.

2.2.3 Slotted Persistent CSMA

In slotted CSMA we assume the time is slotted with slot size a (the propagation delay), and all users are synchronized to start transmission only at slot boundaries. (For convenience, we assume that $1/a$ is an integer.) The duration of a transmission period (whether successful or not) is $1+1/a$ slots. We only consider the case of statistically identical users. In persistent CSMA [Toba74,Klei75b], all the packets that arrive during any ongoing transmission are buffered until some next transmission is started in the channel, and then they are discarded or rescheduled. Let each empty user (who does not have a buffered packet) have an arrival with probability p (and does not with probability $1-p$) in any slot, where $0 < p < 1$. (This is a Bernoulli arrival process.) Assume that each nonempty user (who has a buffered packet) starts transmission with probability h (and does not with probability $1-h$) at the slot boundaries following any idle slot. (This is an h-persistent protocol, where $0 < h \leq 1$.) An attempted transmission is successful if none of the other users start a transmission at the same time. Our derivation of channel throughput for a finite population model follows the approach in [Toba74,Klei75b] for an infinite population model.

To analyze the throughput, let us introduce some notation which defines the channel states as illustrated in Figure 2.3. Let a channel idle period (denoted by I) be the time in which the channel is idle and no packets are awaiting transmission. When any packet arrives, the next slot is said to begin a busy period (denoted by B) which ends if no packets have accumulated at the end of transmission. Let U be the time spent for successful transmission(s) in a busy period B. Then, we have the channel throughput as in Eq.(2.1).

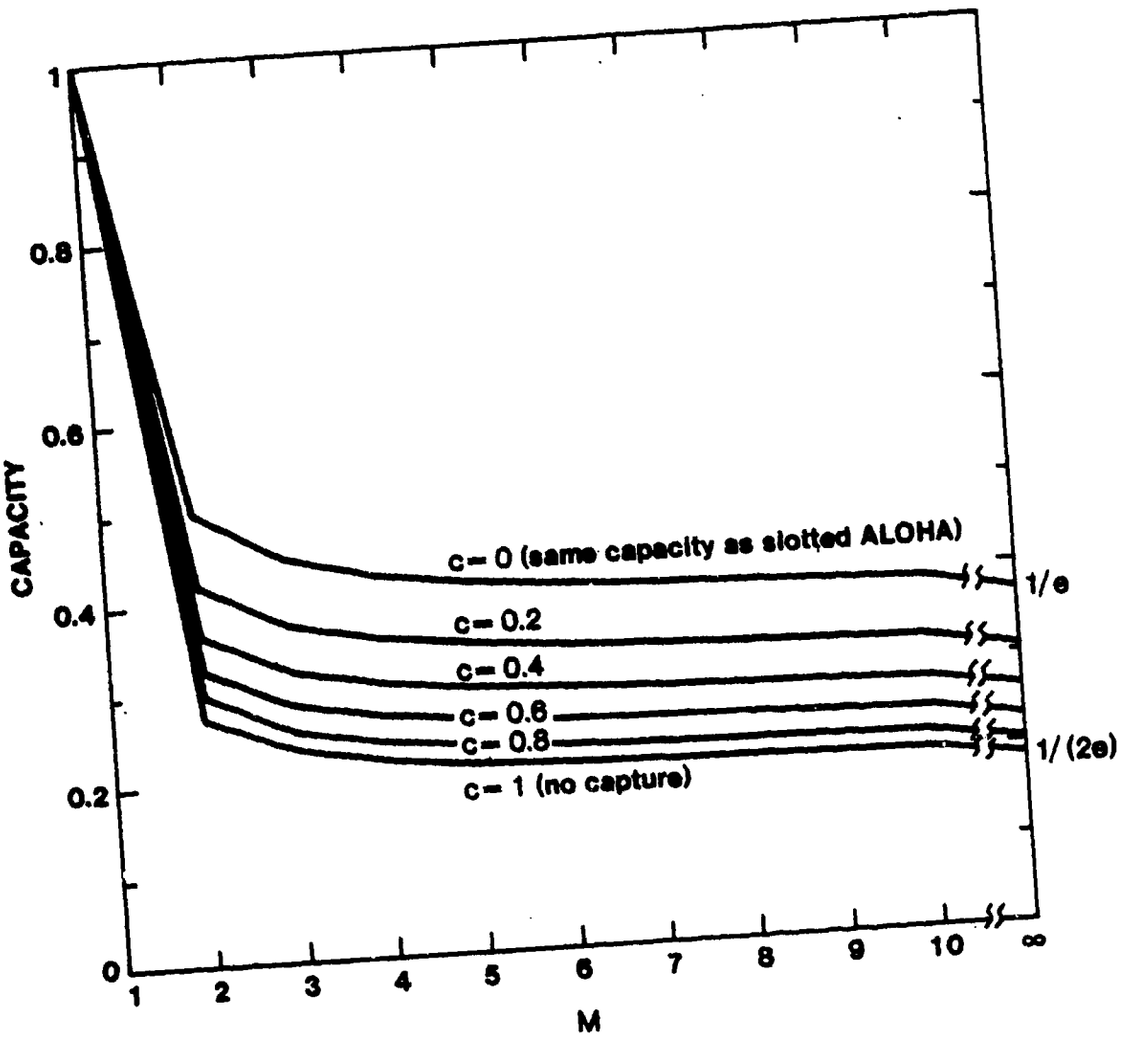
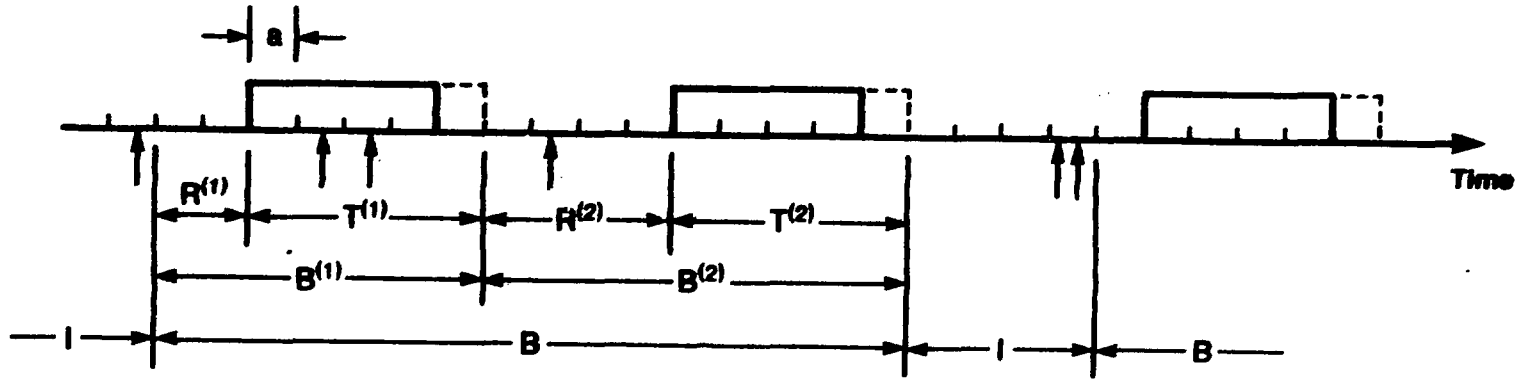
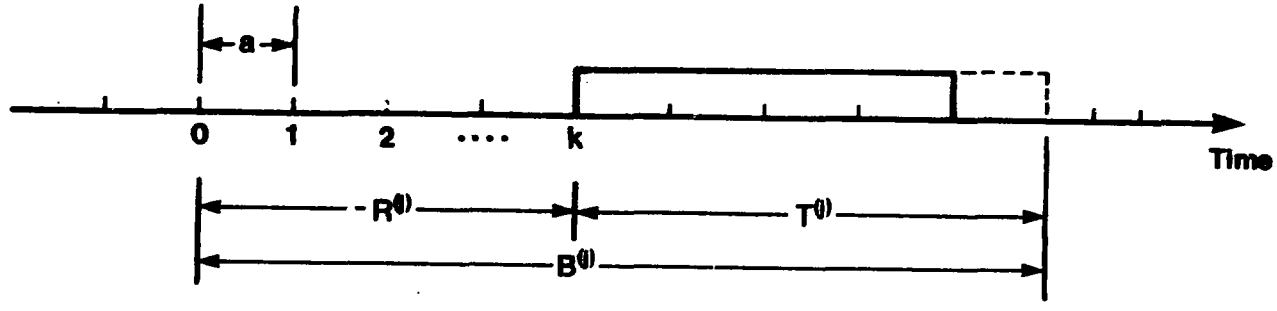


Figure 2.2 Capacity of optimally-controlled pure ALOHA with delay capture.



(a) Overview of channel state.



(b) Close look at the j-th sub(busy) period.

Figure 2.3 Channel state in slotted persistent CSMA.

We divide a channel busy period B into several sub(busy)periods such that the jth subperiod (denoted by $B^{(j)}$) consists of a transmission delay (denoted by $R^{(j)}$) followed by a transmission time (denoted by $T^{(j)}$). A transmission delay is the time in which the channel is idle and packets are awaiting transmission; in the 1-persistent protocol, $R^{(j)}$ is always zero since packets start transmission as soon as they arrive. In CSMA without collision detection, we have $T^{(j)} = 1 + a$ whether the transmission is successful or not. Thus we have

$$B^{(j)} = R^{(j)} + 1 + a \qquad j = 1,2,\cdots \tag{2.24}$$

Lastly, let $U^{(j)}$ be the useful transmission time in the jth subperiod:

$$U^{(j)} = \begin{cases} 1 & \text{if } T^{(j)} \text{ is successful} \\ 0 & \text{if } T^{(j)} \text{ is unsuccessful} \end{cases} \tag{2.25}$$

Next, let J be the number of sub(busy)periods included in a busy period B. Then we have

$$B = \sum_{j=1}^{J} B^{(j)} \; ; \qquad U = \sum_{j=1}^{J} U^{(j)} \tag{2.26}$$

Since the busy period continues as long as there is at least one arrival during the last transmission time (such an event occurs with probability $1-(1-p)^{(1+1/a)M}$), J is geometrically distributed as

$$Prob\,[J = j] = [1-(1-p)^{(1+1/a)M}]^{j-1}(1-p)^{(1+1/a)M} \qquad j = 1,2,\cdots$$

$$\bar{J} = \frac{1}{(1-p)^{(1+1/a)M}} \tag{2.27}$$

Note in Eq.(2.26) that $\{B^{(j)}; j = 1,2,\cdots,J\}$ are independent and $\{B^{(j)}; j = 2,3,\cdots,J\}$ are identically distributed. Also, J is independent of each $B^{(j)}$. The same thing can be said for $\{U^{(j)}\}$. Thus we have

$$\bar{B} = E[B^{(1)}] + (\bar{J}-1)E[B^{(2)}] \; ; \; \bar{U} = E[U^{(1)}] + (\bar{J}-1)E[U^{(2)}] \tag{2.28}$$

It is clear that the duration of an idle period I is geometrically distributed as

$$Prob\,[I = ka] = (1-p)^{M(k-1)}[1-(1-p)^{M}] \qquad k = 1,2,\cdots$$

$$\bar{I} = a/[1-(1-p)^{M}] \tag{2.29}$$

Thus, from Eqs.(2.1), (2.24) and (2.27)-(2.29), we can calculate S if we know $E[R^{(j)}]$ and $E[U^{(j)}]$ for $j = 1,2$.

Let $\pi_n(X)$ be the probability that we have n arrivals among M users in X slots given that $n \geq 1$. Using the Bernoulli arrival rate p at each of M users in a slot, we have

$$\pi_n(X) = \frac{1}{1-(1-p)^{MX}} \binom{M}{n} [1-(1-p)^{X}]^{n}(1-p)^{X(M-n)} \qquad n = 1,2,\cdots,M \tag{2.30}$$

Then the distribution of the number of packets awaiting transmission at the beginning of $B^{(j)}$, denoted by $N_0^{(j)}$, is given by

$$Prob\,[\,N_0^{(j)} = n\,] = \begin{cases} \pi_n(1) & j=1 \\ \pi_n(1+1/a) & j=2,3,\cdots \end{cases} \tag{2.31}$$

(Note in Eq.(2.31) that $N_0^{(j)}$, $j \geqslant 2$, is assumed to be independent of the number of nonempty users at the beginning of $T^{(j-1)}$, which may not be true depending on when a user having a packet at the beginning of $T^{(j-1)}$ discards it and can thus accept the next arrival. However, for simplicity we assume that all users make room for new arrivals from the beginning of $T^{(j-1)}$ putting aside the already buffered packets (including the packets being transmitted in $T^{(j-1)}$) which are discarded; this is the same assumption as in [Toba74,Klei75b], and leads to Eq.(2.31).)

The distribution of $R^{(j)}$ given $N_0^{(j)} = n$, $j \geqslant 1$, can be found as follows. Let us number the slot boundaries as $k = 0,1,2,\cdots$ from the beginning of $R^{(j)}$ as depicted in Figure 2.3(b), and denote by $N_k^{(j)}$ the number of awaiting packets at the kth boundary. Then, according to the h-persistent protocol and Bernoulli arrivals, we have (let $A_i^{(j)}$ be the number of arrivals during the ith slot in $R^{(j)}$)

$$\begin{aligned} &Prob\,[\,R^{(j)} \geqslant ka, A_i^{(j)} = m_i\,(1 \leqslant i \leqslant k),\,|\,N_0^{(j)} = n\,] \\ &= (1-h)^n \binom{M-n}{m_1} p^{m_1}(1-p)^{M-n-m_1} \cdot (1-h)^{n+m_1} \binom{M-n-m_1}{m_2} p^{m_2}(1-p)^{M-n-m_1-m_2} \\ &\quad \cdots \cdot (1-h)^{n+m_1+\cdots+m_{k-1}} \binom{M-n-m_1-\cdots-m_{k-1}}{m_k} p^{m_k}(1-p)^{M-n-m_1-\cdots-m_k} \\ &= (1-h)^{kn}(1-p)^{k(M-n)}(M-n)! \prod_{i=1}^{k} \frac{1}{(m_i)!}\left[\frac{p(1-h)^{k-i}}{(1-p)^{k+1-i}}\right]^{m_i} \end{aligned} \tag{2.32}$$

It follows that

$$\begin{aligned} &Prob\,[\,R^{(j)} \geqslant ka, A_i^{(j)} = m_i\,(1 \leqslant i \leqslant k)\,|\,N_0^{(j)} = n\,] \\ &= \sum_{\sum_{i=1}^{k} m_i = m} Prob\,[\,R^{(j)} \geqslant ka, N_k^{(j)} = n + \sum_{i=1}^{k} m_i\,|\,N_0^{(j)} = n\,] \\ &= (1-h)^{kn}(1-p)^{k(M-n)} \binom{M-n}{m} \left[\sum_{i=1}^{k} \frac{p(1-h)^{k-i}}{(1-p)^{k+1-i}}\right]^m \\ &= (1-h)^{kn}(1-p)^{k(M-n)} \binom{M-n}{m} \left(\frac{p}{h-p}\right)^m \left[1-\left(\frac{1-h}{1-p}\right)^k\right]^m \qquad m = 0,1,2,\cdots,M-n \end{aligned} \tag{2.33}$$

From Eq.(2.33), the expected value of $R^{(j)}$ given $N_0^{(j)} = n$, $j \geqslant 1$, is evaluated as

$$E[R^{(j)} \mid N_0^{(j)} = n] = a \sum_{k=1}^{\infty} \sum_{m=0}^{M-n} Prob\,[R^{(j)} \geq k, N_k^{(j)} = n+m \mid N_0^{(j)} = n]$$

$$= a \sum_{k=1}^{\infty} (1-h)^{kn} \left[\frac{h(1-p)^k - p(1-h)^k}{h-p} \right]^{M-n} \tag{2.34}$$

Unconditioning Eq.(2.34) by using Eqs.(2.30) and (2.31), we get

$$E[R^{(j)}] = \begin{cases} r(1) & j=1 \\ r(1+1/a) & j=2,3,\cdots \end{cases} \tag{2.35}$$

where

$$r(X) \triangleq \frac{a}{1-(1-p)^{XM}} \sum_{k=1}^{\infty} \left\{ (1-h)^k - (1-p)^X h \left[\frac{(1-h)^k - (1-p)^k}{h-p} \right] \right\}^M$$

$$- \frac{a(1-p)^{XM}}{1-(1-p)^{XM}} \sum_{k=1}^{\infty} \left[\frac{h(1-p)^k - p(1-h)^k}{h-p} \right]^M \tag{2.36}$$

Thus, from Eqs.(2.24), (2.27)-(2.29), (2.35) and (3.36), we get

$$\bar{B} + \bar{I} = E[R^{(1)}] + 1 + a + [(1-p)^{-(1+1/a)M} - 1](E[R^{(2)}] + 1 + a) + \bar{I}$$

$$= \frac{1+a}{(1-p)^{(1+1/a)M}} + \frac{a}{(1-p)^{(1+1/a)M}} \sum_{k=1}^{\infty} \left\{ (1-h)^k - (1-p)^{1+1/a} h \left[\frac{(1-h)^k - (1-p)^k}{h-p} \right] \right\}^M \tag{2.37}$$

We proceed to calculate $E[U^{(j)}]$. From the definition of $U^{(j)}$ in Eq.(2.25), we have

$$E[U^{(j)} \mid R^{(j)} \geq k, N_k^{(j)} = n+m, N_0^{(j)} = n] = (n+m)h(1-h)^{n+m-1} \tag{2.38}$$

(i.e., the probability of 1 transmission out of $n+m$ nonempty users). Unconditioning Eq.(2.38) on $R^{(j)}$ and $N_k^{(j)}$ by using Eq.(2.33), we have

$$E[U^{(j)} \mid N_0^{(j)} = n] = nh \sum_{k=0}^{\infty} (1-h)^{(k+1)n-1} \left[\frac{h(1-p)^{k+1} - p(1-h)^{k+1}}{h-p} \right]^{M-n}$$

$$+ (M-n)ph \sum_{k=1}^{\infty} (1-h)^{(k+1)n} \left[\frac{(1-p)^k - (1-h)^k}{h-p} \right] \left[\frac{h(1-p)^{k+1} - p(1-h)^{k+1}}{h-p} \right]^{M-n-1} \tag{2.39}$$

Further unconditioning Eq.(2.39) on $N_0^{(j)}$ by using Eqs.(2.30) and (2.31), we get

$$E[U^{(j)}] = \begin{cases} u(1) & j=1 \\ u(1+1/a) & j=2,3,\cdots \end{cases} \tag{2.40}$$

where

$$u(X) \triangleq \frac{hM}{1-(1-p)^{XM}} \sum_{k=0}^{\infty} \left\{ (1-h)^k - (1-p)^X \left[\frac{h(1-h)^k - p(1-p)^k}{h-p} \right] \right\}$$

$$\cdot \left\{ (1-h)^{k+1} - (1-p)^X h \left[\frac{(1-h)^{k+1} - (1-p)^{k+1}}{h-p} \right] \right\}^{M-1}$$

$$- \frac{pMh(1-p)^{XM}}{1-(1-p)^{XM}} \sum_{k=1}^{\infty} \left[\frac{(1-p)^k - (1-h)^k}{h-p} \right] \left[\frac{h(1-p)^{k-1} - p(1-h)^{k+1}}{h-p} \right]^{M-1} \qquad (2.41)$$

From Eqs.(2.27), (2.28), (2.40) and (2.41), we obtain

$$\bar{U} = \frac{hM}{(1-p)^{(1+1/a)M}} \sum_{k=0}^{\infty} \left\{ (1-h)^k - (1-p)^{1+1/a} \left[\frac{h(1-h)^k - p(1-p)^k}{h-p} \right] \right\}$$

$$\cdot \left\{ (1-h)^{k+1} - (1-p)^{1+1/a} h \left[\frac{(1-h)^{k+1} - (1-p)^{k+1}}{h-p} \right] \right\}^{M-1} \qquad (2.42)$$

Substituting Eqs.(2.37) and (2.42) into Eq.(2.1), we get the channel throughput of a slotted h-persistent CSMA system (with propagation delay a) consisting of M identical users with each Bernoulli arrival rate p:

$$S = \frac{hM \sum_{k=0}^{\infty} \left\{ (1-h)^k - (1-p)^{1+1/a} \left[\frac{h(1-h)^k - p(1-p)^k}{h-p} \right] \right\}}{1+a+a \sum_{k=1}^{\infty} \left\{ (1-h)^k - (1-p)^{1+1/a} h \left[\frac{(1-h)^k - (1-p)^k}{h-p} \right] \right\}^M} \cdot \left\{ (1-h)^{k+1} - (1-p)^{1+1/a} h \left[\frac{(1-h)^{k+1} - (1-p)^{k+1}}{h-p} \right] \right\}^{M-1} \qquad (2.43)$$

By letting $h = 1$ in Eq.(2.43), we have the channel throughput of slotted 1-persistent CSMA as

$$S = \frac{M(1-p)^{(M-1)(1+1/a)} \{ [1-(1-p)^{1+1/a}][1-(1-p)^M] + p(1-p)^{M+1/a} \}}{(1+a)[1-(1-p)^M] + a(1-p)^{(1+1/a)M}} \qquad (2.44)$$

In the limit $M \to \infty$ with $aG = pM$ held at a finite value, Eqs.(2.43) and (2.44) become

$$S = \frac{G \sum_{k=0}^{\infty} \left\{ h(1-h)^k + a[1-(1-h)^{k+1}] \right\} \exp\left\{ G(1-h)^{k+1} + aG \left[-(k+1) + \frac{1-(1-h)^{k+2}}{h} \right] \right\}}{(1+a)\, e^{(1+a)G} + a \sum_{k=1}^{\infty} \exp\left\{ G(1-h)^k + aG \left[-k + \frac{1-(1-h)^{k+1}}{h} \right] \right\}} \qquad (2.45)$$

and

$$S = \frac{Ge^{-(1+a)G}(1+a-e^{-aG})}{(1+a)(1-e^{-aG}) + ae^{-(1+a)G}} \qquad (2.46)$$

respectively. Eq.(2.46) (an infinite population model of slotted 1-persistent CSMA) is derived in [Toba74,Klei75b], where only a procedure to obtain Eq.(2.45) (an infinite population model of slotted h-persistent CSMA) is also given. An explicit expression in Eq.(2.45) is newly provided in this chapter. So are Eqs.(2.43) and (2.44) for finite population models.

In Figure 2.4, we show the throughput of slotted $h\,(=0.03)$-persistent CSMA for M users in terms of the total offered traffic rate $G=pM/a$ $(a=0.01)$. We let $p=\min[1, aG/M]$. A few interesting observations here are: (i) the maximum throughput values are not very dependent on M so long as $M \geqslant 5$, (ii) for a finite M, the throughput does not degrade to zero as G becomes large (this is the case where the busy period is very long but it steadily pushes packets out), (iii) the curves we see in increasing M from 1 to ∞ resemble those for the throughput-load relationship in flow-controlled systems (see for example [Gerl80]). Elaborating on (iii) above, we see that the maximal throughput of an uncontrolled system $(M=\infty)$ is higher than that of an excessively controlled system $(M=1)$. The throughput of less controlled systems (large M) quickly degrades with congestion (large G) while that of controlled systems does not. The highest maximum throughput and sustained behavior in cases of congestion are achieved in moderately controlled systems (around $M=5$ here). We note, however, that for $h=1$ we have only the feature (i) above; except for $M=1$ in this case, the throughput equally degrades as G gets large for any number of users.

2.2.4 Slotted Persistent CSMA with Collision Detection

We consider a slotted h-persistent CSMA protocol with collision detection and a finite population. Here the assumptions and parameters are the same as in the previous section, except that the duration of an unsuccessful transmission is now $b+a$ where $a \leqslant b \leqslant 1$. Our treatment follows that in [Toba80d] for an infinite population model of 1-persistent CSMA.

The channel throughput S is still expressed as in Eq.(2.1) where $\bar{I}$ is given by Eq.(2.29). To find $\bar{B}$ and $\bar{U}$, let us denote by $B(X)$ the mean duration of the busy period following the packet-accumulation time of X slots. Similarly, we denote by $U(X)$ the mean useful transmission time during the same busy period. Since a busy period is induced by those packets which have arrived in the preceding slot, we have

$$\bar{B} = B(1) \;\; ; \quad \bar{U} = U(1) \tag{2.47}$$

It follows from Eqs.(2.1) and (2.29) that

$$S = \frac{U(1)}{B(1) + a/[1-(1-p)^M]} \tag{2.48}$$

where $U(1)$ and $B(1)$ are determined below.

Since the duration of a successful transmission $(1+1/a$ slots) is different from that of an unsuccessful transmission $(1+b/a$ slots), the distribution of $N_0^{(j)}$, $j \geqslant 2$, depends on whether $T^{(j-1)}$ is successful or not. From the recursive consideration similar to [Toba80d], we have

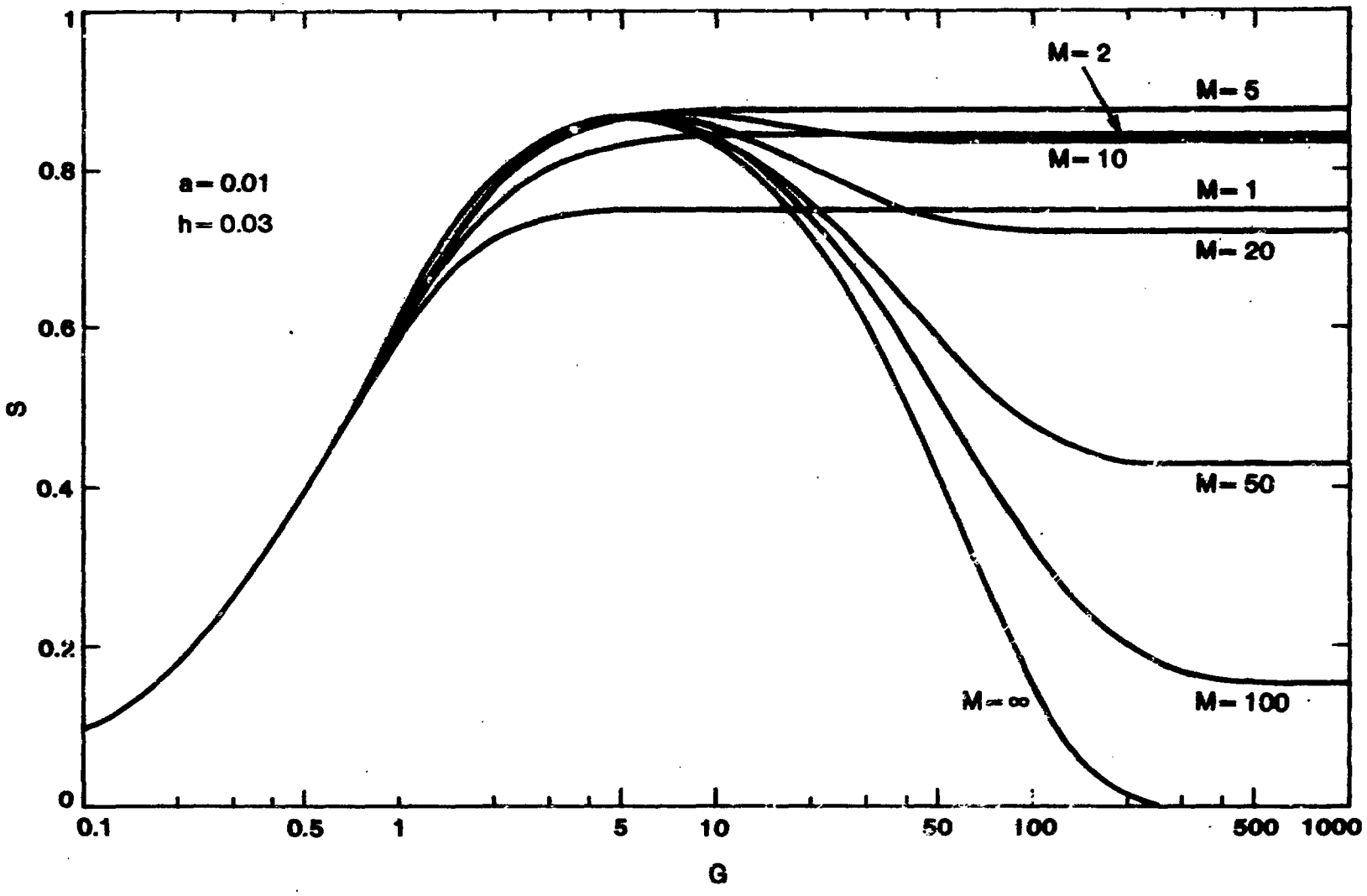


Figure 2.4 Throughput of slotted h-persistent CSMA.

$$B(X) = r(X) + \{1 + a + [1-(1-p)^{(1+1/a)M}]\, B(1+1/a)\}\, u(X)$$

$$+ \{b + a + [1-(1-p)^{(1+b/a)M}]\, B(1+b/a)\}\, [1-u(X)], \tag{2.49}$$

$$U(X) = \{1 + [1-(1-p)^{(1+1/a)M}]\, U(1+1/a)\}\, u(X)$$

$$+ [1-(1-p)^{(1+b/a)M}]\, U(1+b/a)\, [1-u(X)] \tag{2.50}$$

where $r(X)$ and $u(X)$ are defined by Eqs.(2.36) and (2.41), respectively. Writing Eq.(2.49) for $X=1+1/a$ and $X=1+b/a$, we obtain two equations in the two unknowns, $B(1+1/a)$ and $B(1+b/a)$. Solving for them, and using them in Eq.(2.49), we can calculate $B(1)$. Similarly $U(1)$ is obtained from Eq.(2.50). Thus S is found by Eq.(2.48). We note that in 1-persistent CSMA ($h=1$), we have simply $r(X) \equiv 0$ and

$$u(X) = \frac{M(1-p)^{X(M-1)}[1-(1-p)^X]}{1-(1-p)^{XM}} \tag{2.51}$$

The limit $M \to \infty$ with $aG = pM$ held at a fixed value gives the throughput for an infinite population model. In this case, we use the following expressions for $r(X)$ and $u(X)$ (obtained from Eqs.(2.36) and (2.41) by taking the limit) in Eqs.(2.49) and (2.50) with obvious modifications of their coefficients.

$$r(X) = \frac{a}{1-e^{-aGX}} \sum_{k=1}^{\infty} \left\{ \exp[aGX(1-h)^k] - 1 \right\} \exp\left\{ aG\left[-X-k+\frac{1-(1-h)^k}{h} \right] \right\} \tag{2.52}$$

$$u(X) = \frac{aGh}{1-e^{-aGX}} \sum_{k=0}^{\infty} \left\{ \left[(1-h)^k X + \frac{1-(1-h)^k}{h} \right] \exp[aGX(1-h)^{k+1}] - \frac{1-(1-h)^k}{h} \right\}$$

$$\cdot \exp\left\{ aG\left[-X-(k+1)+\frac{1-(1-h)^{k+1}}{h} \right] \right\} \tag{2.53}$$

We note that in the 1-persistent case ($h=1$) we have $r(X) \equiv 0$ and $u(X) = aGX\, e^{-aGX}/(1-e^{-aGX})$ which reduce Eqs.(2.49) and (2.50) to the same form as in [Toba80d].

Figures 2.5 and 2.6 display the throughput values of slotted h-persistent CSMA with collision detection in terms of the total offered traffic value $G = pM/a$ ($a = 0.01$). In Figure 2.5, we see again little dependence of the maximum throughput on the number of users. Also, in 1-persistent CSMA the throughput degrades rapidly as G exceeds some critical value. Figure 2.6 demonstrates the throughput values for different values of the collision detection parameter b in the case of $M = 100$ users for 0.03-persistent CSMA. A good collision detection (small b) is seen to contribute to sustaining throughput when G is large.

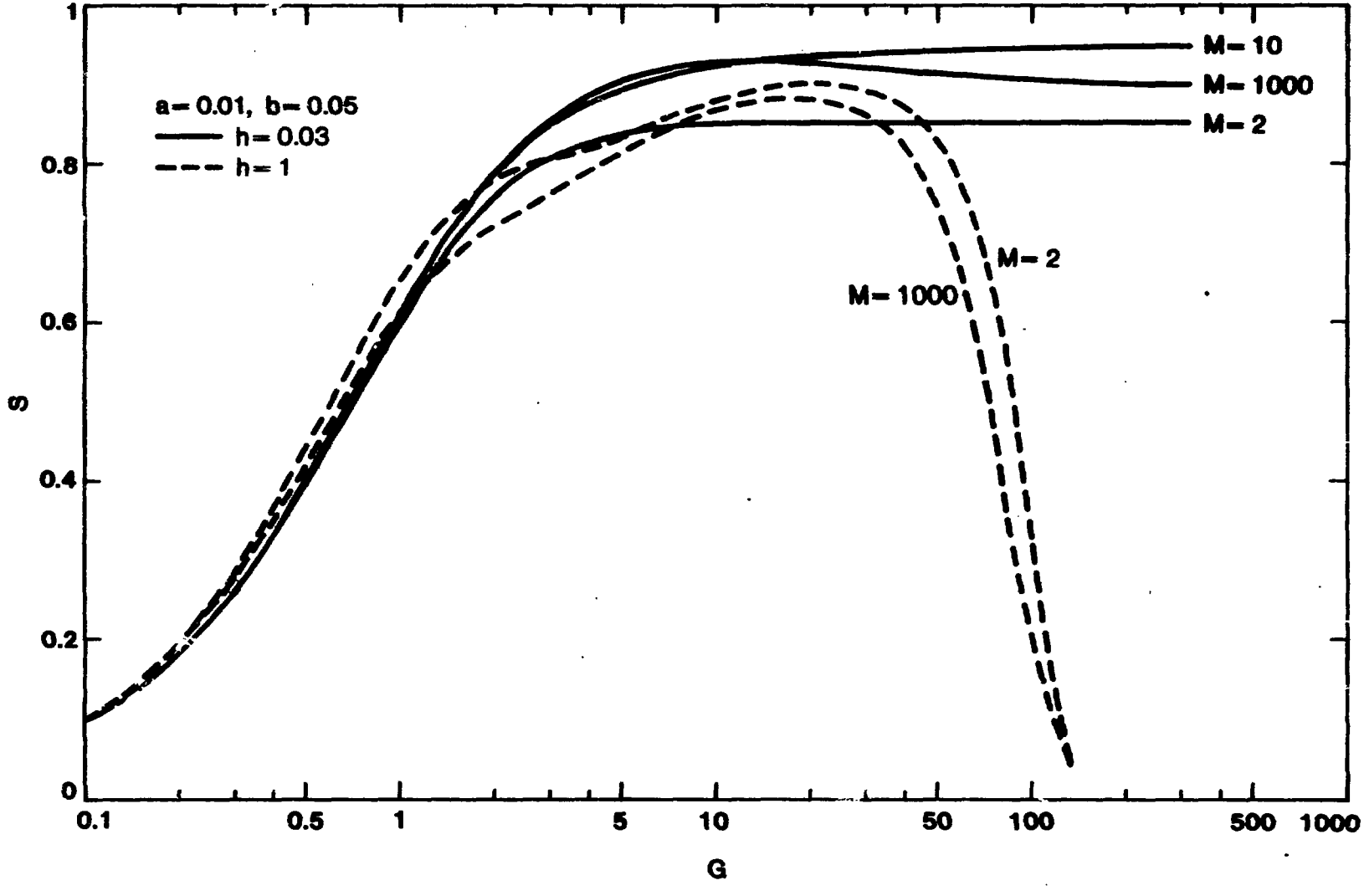


Figure 2.5 *Throughput of slotted h-persistent CSMA with collision detection.*

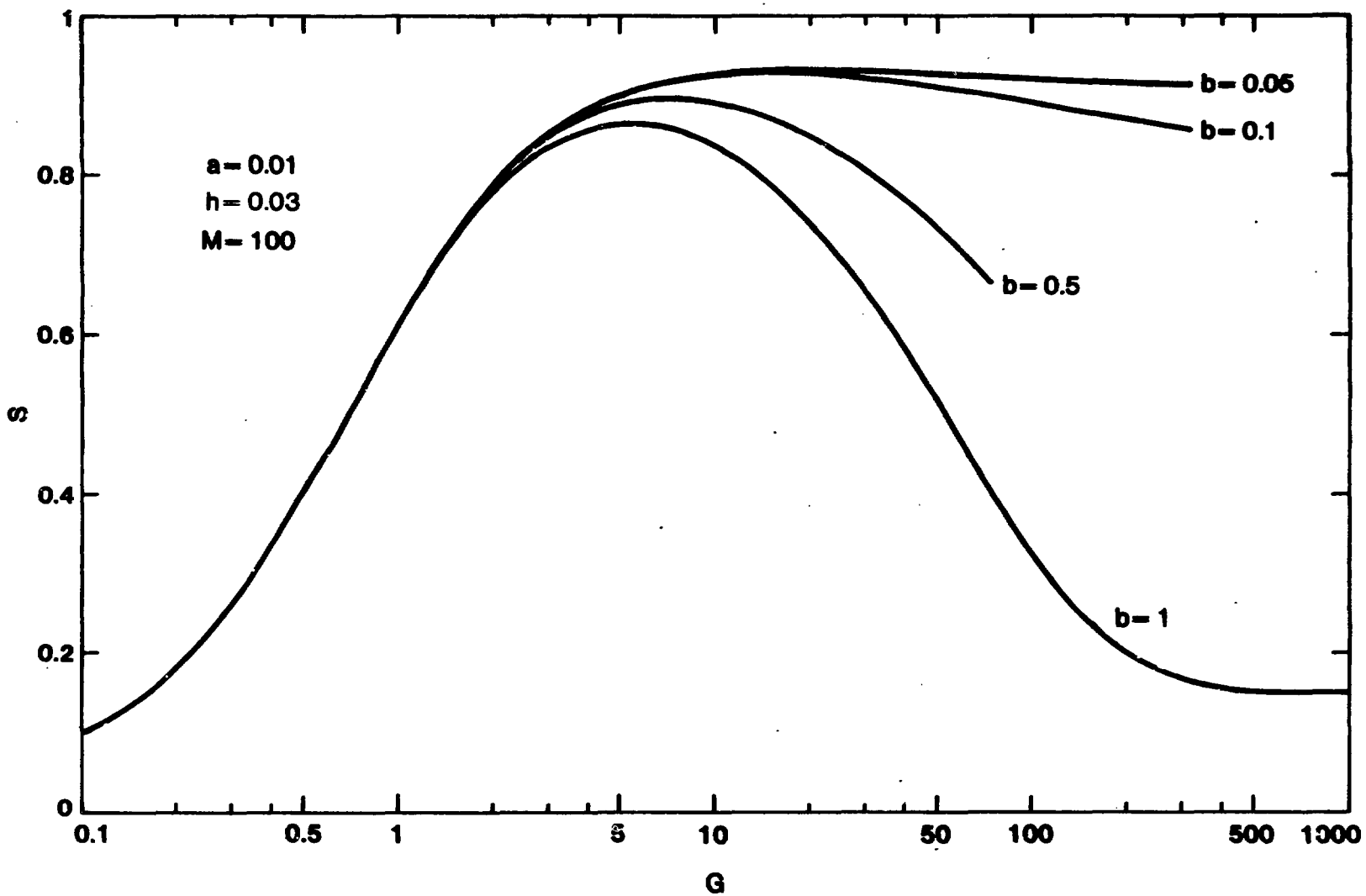


Figure 2.6 Throughput of slotted 0.03-persistent CSMA with collision detection.

2.2.5 Unslotted Persistent CSMA

We now proceed to study unslotted persistent CSMA. Here the unit of time is still the constant packet transmission time and a denotes the signal propagation delay (in that time unit) such that all users recognize what happened in the system a time units before. Let M be the number of users (we consider only the case of identical users), and let the time until a packet arrives at each of the empty users be independent and exponentially distributed with mean $1/g$. If a packet arrives at a user when the channel is sensed idle, he schedules start of transmission a random amount of time later. We assume that this random time is exponentially distributed with mean $1/h$, where $0 < h < \infty$. In time a after any transmission has started, it is recognized by all users who then discard old packets (if any) and make room for new arrivals. If a packet arrives at a user when the channel is sensed busy, the start of transmission is scheduled also at an exponentially distributed time (with mean $1/h$) after the end of the transmission period. An attempted transmission is successful if it is started by breaking the idle channel and if no other transmissions take place within a time a after the start. We call this protocol unslotted h-persistent CSMA. (The case $h = \infty$ corresponds to 1-persistent CSMA, and is treated separately in Section 2.2.6 below.) (Note that the meaning of the parameter h in unslotted CSMA is different from that in slotted CSMA.)

The channel throughput of this system can be analyzed similar to slotted persistent CSMA. Referring to Figure 2.7(a), let B and I be the durations of system busy and idle periods, respectively, as before. That is, the idle period is defined as the time in which the channel is idle and no packets are awaiting transmission. Upon an arrival of packet, the system enters the busy period which terminates at the end of such a transmission period that no packets have arrived therein. If U is the number of successful transmissions achieved in a busy period, then by the renewal argument Eq.(2.1) again gives the channel throughput.

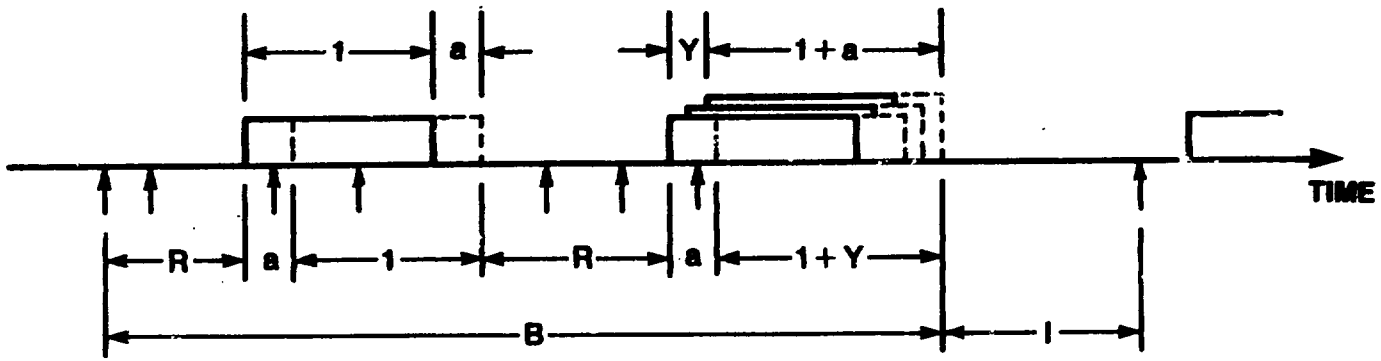
A channel busy period is divided into a number of successive sub(busy)periods each consisting of a transmission delay (denoted by R) followed by a transmission time (of duration $1+a+Y$ where Y is defined shortly). The transmission delay is the time in which the channel is idle but some packets are awaiting transmission. If a transmission is successful, its duration is $1+a$ ($Y=0$); the duration of an unsuccessful transmission period is $1+a+Y$, where Y is the transmission start time of the last colliding packet. Note that the jth subperiod, $j \geq 2$, is generated by the packets which arrive during $1+Y$ in the $(j-1)$st subperiod, whereas the first subperiod is generated by one packet. Let B_n be the mean duration of the busy period initiated by n packets, and U_n be the mean number of successful transmissions in the same busy period. Then, clearly

$$\bar{B} = B_1 \; ; \quad \bar{U} = U_1 \tag{2.54}$$

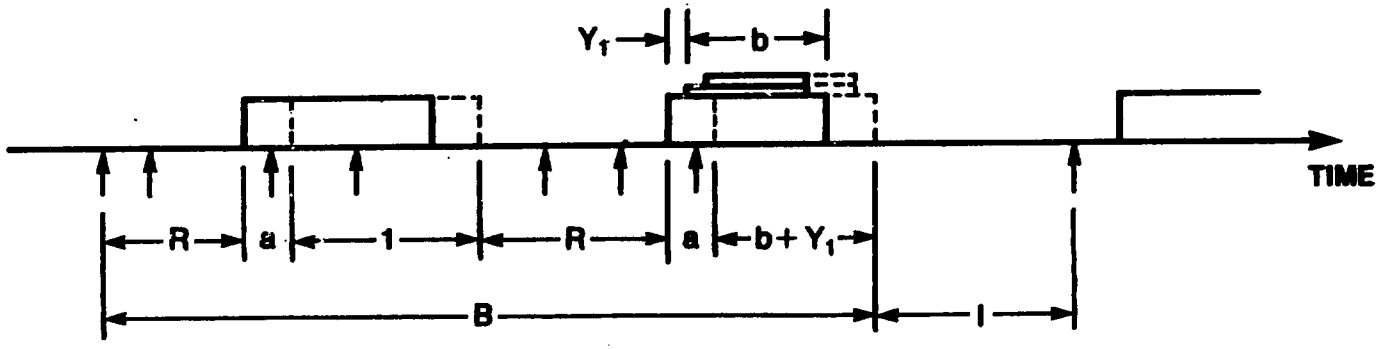
Since $\bar{I} = 1/(gM)$, Eq.(2.1) gives

$$S = \frac{U_1}{B_1 + 1/(gM)} \tag{2.55}$$

In the remainder of this section, we derive a system of linear equations for $\{B_n ; n = 1,2,\cdots,M\}$ and $\{U_n ; n = 1,2,\cdots,M\}$.



(a) Unslotted persistent CSMA.



(b) Unslotted persistent CSMA with collision detection.

Figure 2.7 Channel state in unslotted persistent CSMA.

Let us focus our attention on a sub(busy)period which begins with n packets. Taking the origin of the time at the start of this subperiod, let $N(x)$ be the number of packets present in the system (i.e., the number of nonempty users) at time x. We first consider the distribution of R on the condition that $N(0)=n$. Note that during the transmission delay R, each user behaves independently of others. Each of n busy users schedules his transmission after time x with probability e^{-hx}. Consider the event $\{R>x,\ N(x)=n+m\}$. Each of $M-n-m$ users does not have an arrival before x with probability e^{-gx}. For each of m users, the time until arrival plus start of transmission is less than x with probability

$$\int_0^x ge^{-gz}\cdot e^{-h(x-z)}dz = \frac{g}{h-g}\left(e^{-gx}-e^{-hx}\right) \tag{2.56}$$

Thus we have

$$Prob\left[R>x,\ N(x)=n+m \mid N(0)=n\right] = e^{-hxn}\binom{M-n}{m}e^{-gx(M-n-m)}\left[\frac{g(e^{-gx}-e^{-hx})}{h-g}\right]^m$$

$$x\geq 0 \qquad m=0,1,2,\cdots,M-n \tag{2.57}$$

Adding Eq.(2.57) over all m, the number of arrivals during R, we have

$$Prob\left[R>x \mid N(0)=n\right] = e^{-hxn}\left(\frac{he^{-gx}-ge^{-hx}}{h-g}\right)^{M-n} \qquad x\geq 0 \tag{2.58}$$

and so

$$E\left[R_{(n)}\right] \triangleq E\left[R \mid N(0)=n\right] = \int_0^\infty e^{-hnx}\left(\frac{he^{-gx}-ge^{-hx}}{h-g}\right)^{M-n}dx$$

$$n=1,2,\cdots,M \tag{2.59}$$

We next consider the behavior of $M-1$ users in the transmission period following the transmission delay such that $R=x$ and $N(x)=n+m$. (Since $g,h<\infty$, at most one event happens at a time; one user begins the transmission period.) On this condition, the event $\{Y\leq y\}$ occurs in the following cases. Again note that during the first a time units, each user behaves independently. Each of $n+m-1$ nonempty users either does not start transmission before a with probability e^{-ha}, or starts transmission before y with probability $1-e^{-hy}$. There are three cases of behavior for each of $M-n-m$ users who were empty at the end of R: (i) no arrival during a with probability e^{-ga}, (ii) arrival before a but transmission after a with probability $g(e^{-ga}-e^{-ha})/(h-g)$ (similar to Eq.(2.56)), (iii) arrival and transmission before y with probability

$$\int_0^y ge^{-gz}\left[1-e^{-h(y-z)}\right]dz = 1-\frac{he^{-gy}-ge^{-hy}}{h-g} \tag{2.60}$$

Therefore, we have

$$Prob\left[Y\leq y \mid R=x,\ N(x)=n+m,\ N(0)=n\right]$$

$$= (1-e^{-hy}+e^{-ha})^{n+m-1}\left[1-\frac{he^{-gy}-ge^{-hy}}{h-g}+e^{-ga}+\frac{g(e^{-ga}-e^{-ha})}{h-g}\right]^{M-n-m}$$

$$= (1-e^{-hy}+e^{-ha})^{n+m-1}\left[\frac{h(1-e^{-gy}+e^{-ga})-g(1-e^{-hy}+e^{-ha})}{h-g}\right]^{M-n-m} \qquad 0 \leqslant y \leqslant a \tag{2.61}$$

Note that

$$Prob\,[\,Y=0 \mid R=x, N(x)=n+m, N(0)=n\,] = e^{-ha(n+m-1)}\left(\frac{he^{-ga}-ge^{-ha}}{h-g}\right)^{M-n-m} \tag{2.62}$$

is the probability of a successful transmission. Unconditioning Eq.(2.61) on $N(R)$ and R using Eqs.(2.57) and (2.58) successively gives

$$f(y;n) \triangleq Prob\,[\,Y \leqslant y \mid N(0)=n\,]$$

$$= (1-e^{-hy}+e^{-ha})^{n-1}\left[\frac{h(1-e^{-gy}+e^{-ga})-g(1-e^{-hy}+e^{-ha})}{h-g}\right]^{M-n} \qquad 0 \leqslant y \leqslant a \tag{2.63}$$

and similar unconditioning of Eq.(2.62) yields

$$\gamma_{(n)} \triangleq f(0;n) = e^{-ha(n-1)}\left(\frac{he^{-ga}-ge^{-ha}}{h-g}\right)^{M-n} \qquad n=1,2,\cdots,M \tag{2.64}$$

Note that $\gamma_{(n)}$ is the probability of success in the subperiod begun with n packets. The mean of Y in the similar subperiod is given by

$$E\,[\,Y_{(n)}\,] \triangleq E\,[\,Y \mid N(0)=n\,] = a - \int_0^a f(y;n)\,dy \qquad n=1,2,\cdots,M \tag{2.65}$$

Lastly we consider the condition that we have k accumulated packets at the end of the transmission period. If the duration of the transmission period is $1+a+y$, those packets which arrive during $1+y$ are buffered. So the probability of having k packets in the transmission period of duration $1+a+y$ is given by $g_k(1+y)$, where

$$g_k(y) \triangleq \binom{M}{k}(1-e^{-gy})^k\, e^{-gy(M-k)} \qquad k=1,2,\cdots,M \tag{2.66}$$

By similar unconditioning, we have the probability p_{nk} that the next subperiod begins with k packets:

$$p_{nk} = g_k(1)f(0;n) + \int_0^a g_k(1+y)\,d_y f(y;n) \tag{2.67}$$

where $f(y;n)$ is given by Eq.(2.63). (We have here assumed that $g_k(1+y)$ is independent of n, the number of nonempty users at the beginning of the transmission period. This is based on the same assumption as that stated after Eq.(2.31).)

Now we are in a position to write down a system of equations for $\{B_n\}$ and $\{U_n\}$. By renewal considerations, they are given by

$$B_n = E[R_{(n)}] + 1 + a + E[Y_{(n)}] + \sum_{k=1}^{M} B_k p_{nk} \qquad n = 1,2,\cdots,M \tag{2.68}$$

and

$$U_n = \gamma_{(n)} + \sum_{k=1}^{M} U_k p_{nk} \qquad n = 1,2,\cdots,M \tag{2.69}$$

where $E[R_{(n)}]$, $E[Y_{(n)}]$, $\gamma_{(n)}$ and p_{nk} are given by Eqs.(2.59), (2.65), (2.64) and (2.67), respectively. Thus all we have to do to compute the throughput is to use the solution B_1 and U_1 to Eqs.(2.68) and (2.69) in Eq.(2.55).

In Figure 2.8, we plot the throughput of unslotted h-persistent CSMA for $a = 0.01$ and $M = 10$ in terms of the total offered traffic rate $G = gM$. (Note that here, the scale of G is different from Figures 2.4, 2.5, 2.6 and 2.9.) For small h, a long transmission delay R suppresses the start of actual transmission too much for small G, thus causing a reduced throughput. When h is increased, so is the probability of collision. The maximum throughput is achieved on the balance of idle period, transmission delay and probability of collision. The smaller h is, the larger the optimal G is.

2.2.6 Unslotted 1-Persistent CSMA

In the case of 1-persistent CSMA ($h = \infty$), all the packets accumulated by the end of a transmission period are started immediately at the beginning of the next subperiod. Therefore the duration of transmission delay R is always zero, and the fact used in Section 2.2.5 that each transmission period is initiated by one user is not valid. However a similar analysis is possible, and the channel throughput can still be calculated by using the solutions to Eqs.(2.68) and (2.69), where the following replacement is made:

$$E[R_{(n)}] = 0 \;; \quad \gamma_{(n)} = \delta_{n,1} e^{-ga(M-1)} \tag{2.70}$$

($\delta_{n,1} = 1$ if $n = 1$, and $= 0$ otherwise)

$$E[Y_{(n)}] = a - \int_0^a (1 - e^{-gy} + e^{-ga})^{M-n} dy \tag{2.71}$$

$$p_{nk} = g_k(1) e^{-ga(M-n)} + (M-n) g \int_0^a g_k(1+y) e^{-gy} (1 - e^{-gy} + e^{-ga})^{M-n-1} dy \tag{2.72}$$

In the limit $M \to \infty$ with $G = gM$ held at a finite value, $E[Y_{(n)}]$ and p_{nk} become independent of n as

$$\bar{Y} = a - \frac{1 - e^{-aG}}{G} \tag{2.73}$$

$$p_k = e^{-G(1+a)} \left\{ \frac{G^{k+1}}{(k+1)!} [(1+a)^{k+1} - 1] + \frac{G^k}{k!} \right\} \qquad k = 0,1,2,\cdots \tag{2.74}$$

Thus we have

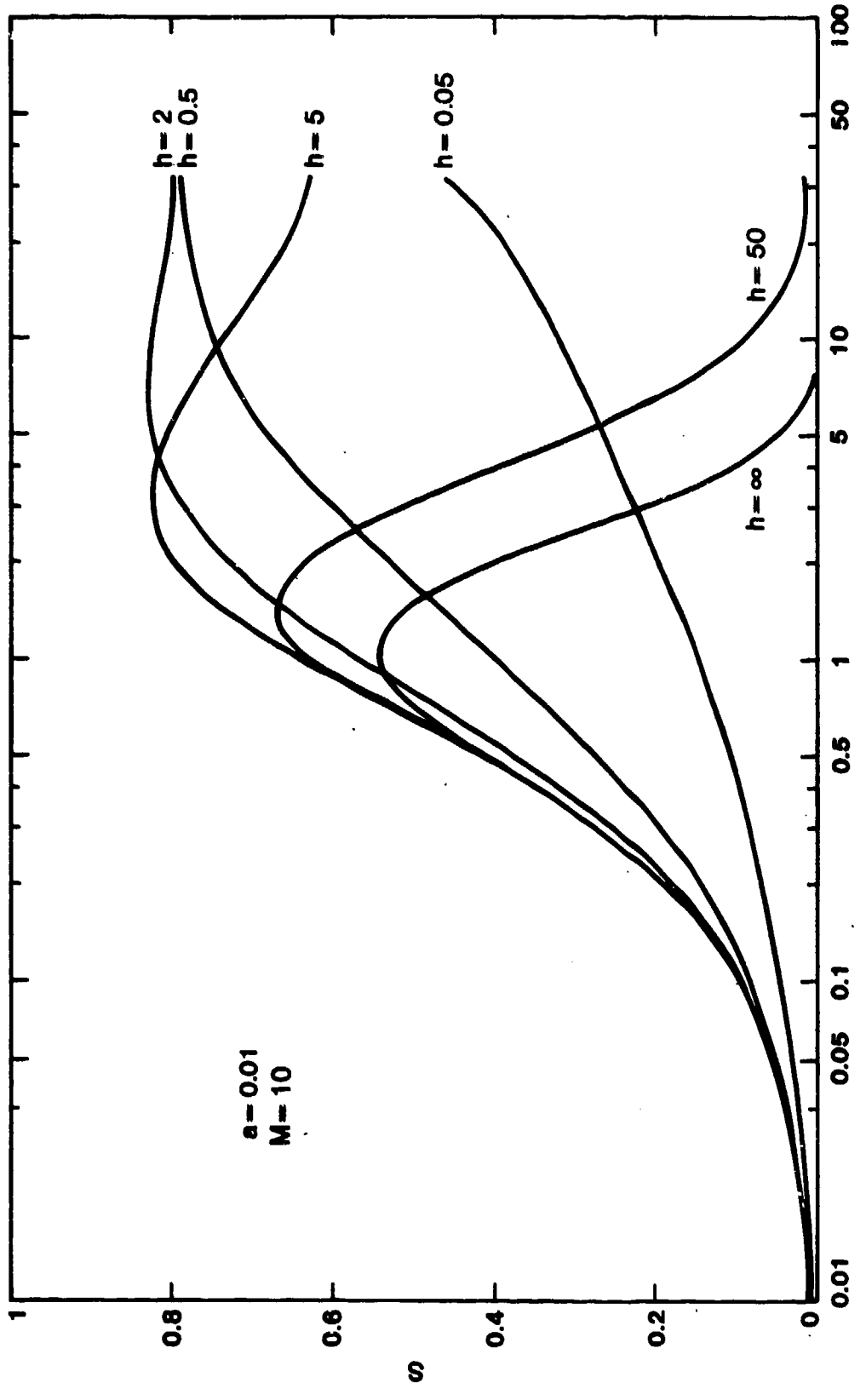


Figure 2.8 Throughput of unslotted h-persistent CSMA.

$$\bar{B} = \frac{1+a+\bar{Y}}{p_0}; \quad \bar{I} = \frac{1}{G}; \quad \bar{U} = e^{-aG} + \frac{p_1}{p_0} e^{-aG} \tag{2.75}$$

Then we recover the result in [Toba74,Klei75b] for an infinite population model:

$$S = \frac{Ge^{-G(1+2a)}[1+G+aG(1+G+aG/2)]}{G(1+2a)-(1-e^{-aG})+(1+aG)e^{-G(1+a)}} \tag{2.76}$$

The values of in Eq.(2.76) when $a = 0.01$ are plotted in Figure 2.8 with a label $h = \infty$.

2.2.7 Unslotted Persistent CSMA with Collision Detection

CSMA with collision detection can be treated similarly. Let the collision detection in an unslotted system be such that the duration of an unsuccessful transmission is $b + a + Y_1$, where Y_1 is the transmission start time of the *first* colliding packet. (Unlike CSMA without collision detection, note here that it is the first colliding packet that stops the transmission of the leading packet which lingers till the last moment.) The distribution of transmission delay R and the probability of success in a subperiod given that $N(0) = n$ are the same as before; they are given by Eqs.(2.57)-(2.59) and Eq.(2.64), respectively. An illustration of channel state is given in Figure 2.7(b).

Let us find the probability of the event $\{ Y_1 > y \}$ conditioned on the event $\{ R = x, N(x) = n+m, N(0) = n \}$ and that the transmission is unsuccessful. This event occurs when each of $n+m-1$ nonempty users does not start transmission before y with probability $e^{-hy(n+m-1)}$, and when the time of transmission start following an arrival at each of $M-n-m$ empty users is after y with probability $(he^{-Ry} - ge^{-hy})/(h-g)$ (calculated similar to Eq.(2.60)). Since each user behaves independently during $0 \leq y \leq a$, we have

$$Prob\,[\, Y_1 > y \mid \text{collision}, R = x, N(x) = n+m, N(0) = n \,]$$

$$= \frac{e^{-hy(n+m-1)}\left(\frac{he^{-Ry} - ge^{-hy}}{h-g}\right)^{M-n-m} - e^{-ha(n+m-1)}\left(\frac{he^{-Ra} - ge^{-ha}}{h-g}\right)^{M-n-m}}{1 - e^{-ha(n+m-1)}\left(\frac{he^{-Ra} - ge^{-ha}}{h-g}\right)^{M-n-m}} \tag{2.77}$$

Unconditioning Eq.(2.77) on R and $N(R)$ (by using Eq.(2.57)) and taking the average, we get

$$E\,[\, Y_{1(n)} \,] \triangleq E\,[\, Y_1 \mid \text{collision}, N(0) = n \,] = \frac{\int_0^a f_1(y;n)\,dy - a\gamma_{(n)}}{1 - \gamma_{(n)}} \tag{2.78}$$

where $\gamma_{(n)}$ is given by Eq.(2.64), and

$$f_1(y;n) \triangleq Prob\,[\, Y_1 > y \mid N(0) = n \,] = e^{-hy(n-1)}\left(\frac{he^{-Ry} - ge^{-hy}}{h-g}\right)^{M-n} \tag{2.79}$$

Note also $\gamma_{(n)} = f_1(a;n)$.

The probability that a successful subperiod begun with n nonempty users ends up with k nonempty users is obviously given by

$$p_{nk}^{(s)} = g_k(1) \tag{2.80}$$

where $g_k(y)$ is defined in Eq.(2.66). In the case of an unsuccessful transmission, since the packets are accumulated over the duration $b + Y_1$, the corresponding probability is given by

$$p_{nk}^{(c)} = \frac{-1}{1-\gamma_{(n)}} \int_0^a g_k(b+y)\, d_y f_1(y;n) \tag{2.81}$$

We are now able to write a system of equations for $\{B_n\}$. By renewal consideration as in Section 2.2.4, we have

$$B_n = E[R_{(n)}] + \gamma_{(n)}\{1 + a + \sum_{k=1}^{M} B_k p_{nk}^{(s)}\} + [1-\gamma_{(n)}]\{b + a + E[Y_{1(n)}] + \sum_{k=1}^{M} B_k p_{nk}^{(c)}\} \tag{2.82}$$

which may be rewritten as

$$B_n = E[R_{(n)}] + \gamma_{(n)} + [1-\gamma_{(n)}](b+a) + \int_0^a f_1(y;n)\, dy + \sum_{k=1}^{M} p_{nk} B_k$$

$$n = 1,2,\cdots,M \tag{2.83}$$

where

$$p_{nk} \triangleq \gamma_{(n)} p_{nk}^{(s)} + [1-\gamma_{(n)}]\, p_{nk}^{(c)} = \gamma_{(n)} g_k(1) - \int_0^a g_k(b+y)\, d_y f_1(y;n) \tag{2.84}$$

is the probability that the next subperiod begins with k packets. Similarly, for $\{U_n\}$ we have

$$U_n = \gamma_{(n)} + \sum_{k=1}^{M} p_{nk} U_k \qquad n = 1,2,\cdots,M \tag{2.85}$$

Thus, by solving Eq.(2.83) for $\{B_n\}$ and Eq.(2.85) for $\{U_n\}$, we can find B_1 and U_1 which are to be used in Eq.(2.55) to compute S.

Figure 2.9 shows the throughput of unslotted h-persistent CSMA with collision detection for some combinations of h and b in the case of $a = 0.01$ and $M = 10$. Here the throughput values appear to depend little on b, particularly for large values of h unless b is close to a. The reason for this is explained below by using the explicit expressions for $\bar{B}$ and $\bar{U}$ for an infinite population model of 1-persistent CSMA.

In unslotted 1-persistent CSMA with collision detection, the system of equations for $\{B_n\}$ and $\{U_n\}$ is given by

$$B_n = e^{-Ga(M-n)} + [1 - e^{-Ga(M-n)}]\,[b + a + \frac{1}{g(M-n)}] + \sum_{k=1}^{M} p_{nk} B_k \tag{2.86}$$

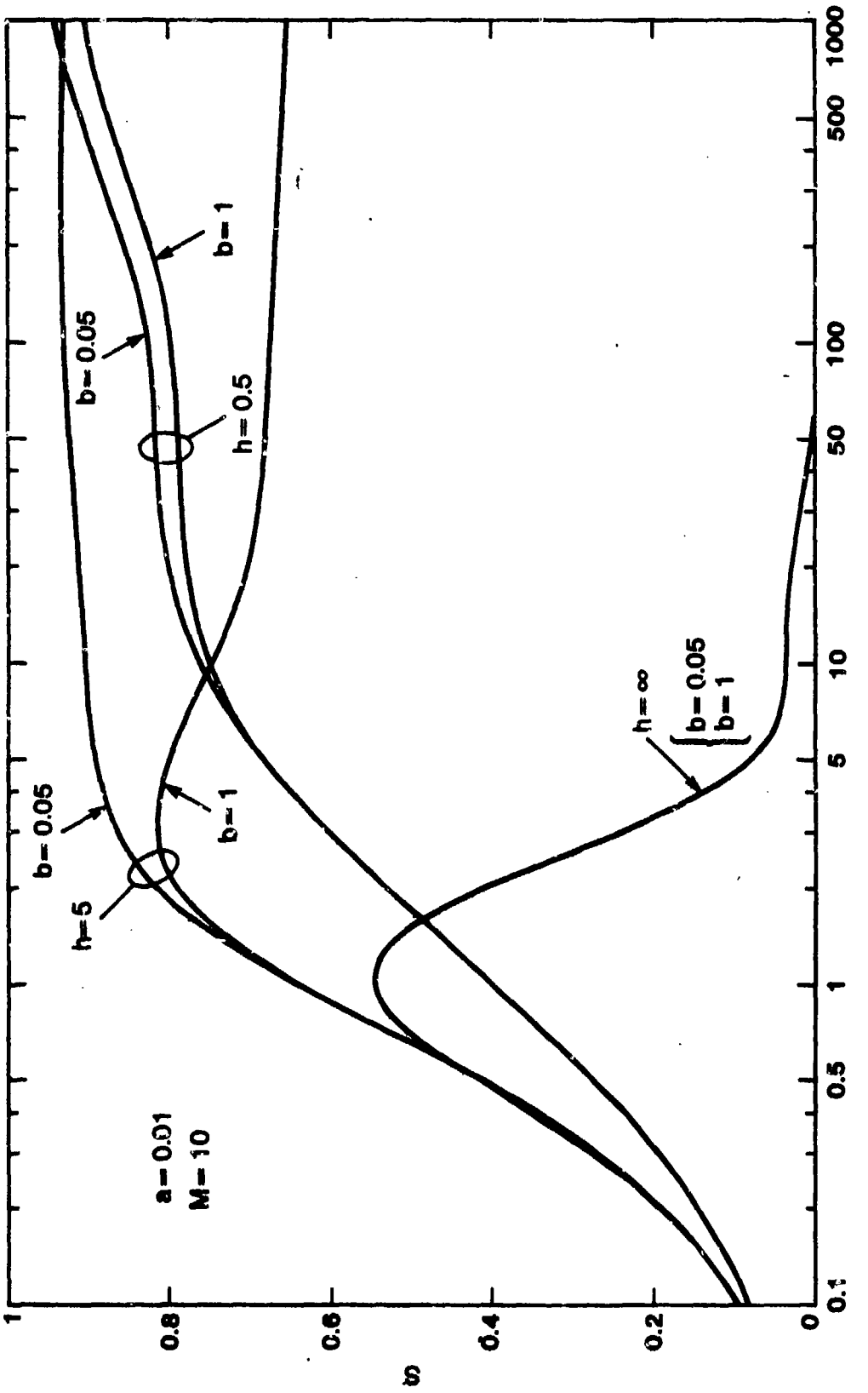


Figure 2.9 Throughput of unslotted persistent CSMA with collision detection.

$$U_n = \delta_{n,1} e^{-ga(M-1)} + \sum_{k=1}^{M} p_{nk} U_k \qquad n = 1,2,\cdots,M \tag{2.87}$$

where

$$p_{nk} = g_k(1) e^{-ga(M-n)} + (M-n) g \int_0^a g_k(b+y) e^{-gy(M-n)} dy \tag{2.88}$$

In the limit $M \to \infty$ with $G = gM$ fixed at a finite value, p_{nk} becomes independent of n and we obtain explicitly

$$p_0 = e^{-G(1+a)} + e^{-bG}(1 - e^{-2aG})/2 \tag{2.89}$$

$$p_1 = Ge^{-G(1+a)} + Ge^{-bG}[(\frac{b}{2} + \frac{1}{4G})(1 - e^{-2aG}) - \frac{a}{2} e^{-2aG}] \tag{2.90}$$

$$B_1 = \frac{1}{p_0}[e^{-aG} + (1 - e^{-aG})(b + a + \frac{1}{G})] \tag{2.91}$$

$$U_1 = (1 + \frac{p_0}{p_1}) e^{-aG} \tag{2.92}$$

Substituting these expressions as well as $gM = G$ into Eq.(2.55), we obtain

$$S = \frac{G(1+G)e^{-G(1+2a)} + Ge^{-G(b+a)}[(bG/2 + \frac{1}{4})(1 - e^{-2aG}) - aGe^{-2aG}/2]}{e^{-G(1+a)} + Ge^{-aG} + (1 - e^{-aG})[1 + (b+a)G] + e^{-bG}(1 - e^{-2aG})/2} \tag{2.93}$$

Figure 2.10 plots S in Eq.(2.93) in the case of $a = 0.01$. As in Figure 2.9, S depends little on b unless b is small. This comes from the following behavior of B_1 and U_1 as G changes. When G is small we are likely to have successful transmissions, so that the effects of collision detection are nominal. When G gets large and the collision detection becomes effective, the increase in the number of subperiods in a busy period (as e^{bG} in $1/p_0$) outweighs the decrease in the duration of each subperiod (linear in b). As a result, the duration of the whole busy period grows rapidly (as be^{bG}), while the growth in the number of successful transmissions in a busy period levels off and then decreases due to increased collisions. Thus the collision detection feature seems ineffective in these cases.

On the other hand, if b is very small the double peaks in Figure 2.10 are striking. As G increases, the first peak (near $G = 1$) corresponds to the point of balance between the durations of the idle period and the busy period which is most likely to contain one transmission period which is successful. The second peak (whose position and height depend on b) corresponds to the point of maximum number of successful transmissions per unit length of a busy period (the idle period has little effect here because of large G). The smaller b is, the faster the unsuccessful transmission periods are ended. This is why a smaller b brings about a larger throughput for a given value of G in this region. Furthermore, the optimal G which maximizes the throughput (in this region) is larger for smaller values of b since then more transmissions can be started for the same duration of the busy period.

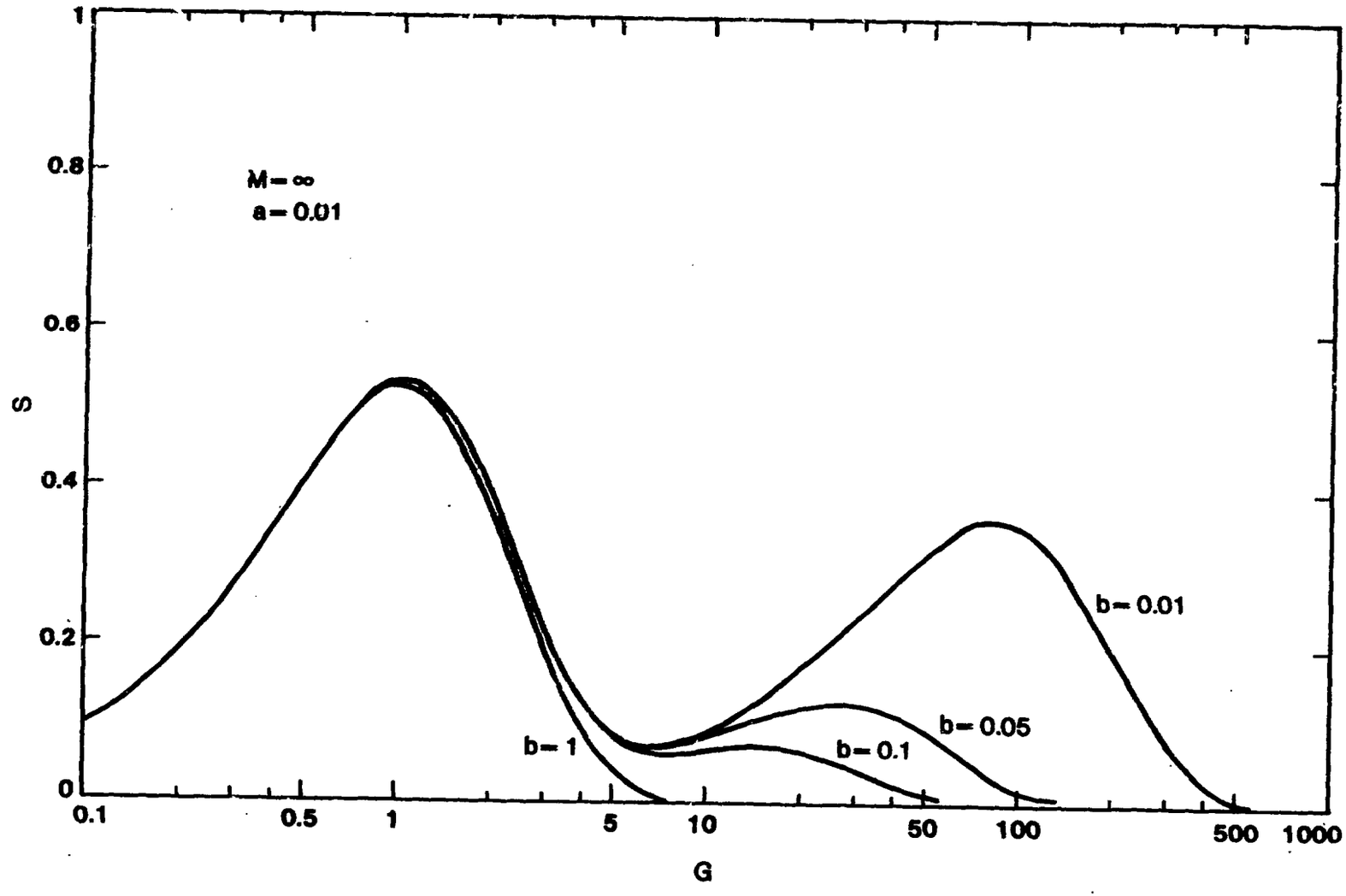


Figure 2.10 Throughput of unslotted 1-persistent CSMA with collision detection.

2.3 Conclusion

In this chapter, we have given the throughput analysis for pure ALOHA and slotted and unslotted persistent CSMA. Together with our study in Chapter 3 on the output processes for slotted ALOHA and slotted and unslotted nonpersistent CSMA (and their variations), we have clarified the underlying stochastic structures for a broad class of random channel-access protocols for finite and infinite populations of users. Due to the assumption of exponentially or geometrically distributed idle periods, the intervals between two successive epochs at which the system enters the idle period are independent and identically distributed. Therefore, the system state can be modeled as a regenerative process. Furthermore, in ALOHA and nonpersistent CSMA, since there is at most one successful transmission in a regenerative cycle, the packet interdeparture times (i.e., the intervals between two successive successful transmissions) are also independent and identically distributed. In persistent CSMA which we have studied in this chapter, they are generally neither independent nor identically distributed. However, through renewal arguments, we can still calculate the channel throughput as we have shown here. We expect that this chapter along with Chapter 3 provides a unified treatment of the throughput analysis for the traditional ALOHA and CSMA.

Two major assumptions we have adopted in this chapter are (i) the packet interarrival times at each user are independent and identically (exponentially or geometrically) distributed, and (ii) (in persistent CSMA) the number of packets accumulated at the end of a transmission period is simply the number of arrivals during that transmission period in disregard of the packets which were already buffered at the beginning of the transmission period (they are discarded). The assumption (i) is essential in making analysis tractable in virtue of the memoryless property. The assumption (ii) was used to be consistent with the previous treatment in [Toba74,Klei75b] (in fact, in Eq.(2.46), we have derived one of the previous results as a special case). One of the advantages drawn from this assumption is that, in the analysis of slotted persistent CSMA (Section 2.2.3) the subperiods $B^{(j)}$, $j \geq 2$, are statistically independent and identical, and this fact brings about the closed-form expression for throughput as Eq.(2.43). Instead of (ii) above, we could have assumed that all the accumulated packets are kept in the buffer until they are successfully transmitted. Then, the number of packets accumulated during a transmission period would depend on the number of packets buffered at the beginning of the transmission period. We note that the throughput analysis of (slotted and unslotted) CSMA based on the latter assumption is still possible; we then only would have a system of linear equations for $\{ B_n \}$ and $\{ U_n \}$ (see Section 2.2.5 for their definition) like Eqs.(2.68) and (2.69). Thus, the assumption (ii) is not essential for the tractability of analysis.

Since the major purpose of the present chapter is to explore the throughput analysis techniques for finite population models, we have not gone into the area of optimization (with respect to G and h) or the comparison of optimized throughput values (capacities) among protocols. These subjects as well as possible ramification of models (e.g., the above-mentioned alternative assumption to (ii)) remain to be elaborated.

CHAPTER 3
Output Processes in Contention Packet Broadcasting Systems

The processes consisting of the packet interdeparture times in contention-type packet broadcasting systems are studied. The channel access protocols considered include slotted and unslotted ALOHA and carrier-sense-multiple-access (CSMA) with collision detection or with delay capture effect. Through analysis of the channel activity cycle, the distribution, mean and coefficient of variation of the packet interdeparture times are explicitly derived. Taking the reciprocal of the mean interdeparture time, we obtain the channel throughput. Cases with dissimilar users are also considered. Application of the present results to the packet queueing processes is given in Chapter 5.

3.1 Introduction

This chapter presents an analysis of packet interdeparture times (i.e., intervals between two consecutive successful transmissions) for a number of channel access protocols encountered in packet broadcasting communication systems such as packet radio networks and local-area computer networks. Specifically, we are interested in the average $\bar{X}$ and the coefficient of variation $C^2 = Var[X]/\bar{X}^2$ of the packet interdeparture time X for given protocols. Throughout the chapter we assume the constant packet transmission time to be 1 as the unit of time. Then, $S = 1/\bar{X}$ is equivalent to the channel throughput which can alternatively be obtained as the ratio of the average time that the channel is used for successful transmission in a cycle of channel usage to the average cycle duration. Our approach is to calculate $\bar{X}$ and C^2 from the distribution of X. We also make the heavy-traffic assumption that all users contain packets all the time.

The channel access protocols we consider here include:

a. pure ALOHA [Abra70]

b. slotted ALOHA [Robe72]

c. slotted carrier-sense-multiple-access (CSMA) [Toba74,Klei75b]

d. slotted CSMA with collision detection (CSMA/CD) [Toba80d]

e. unslotted CSMA [Toba74,Klei75b]

f. unslotted CSMA with collision detection

g. unslotted CSMA with delay capture

(Each protocol model is described individually below.) In all models, we assume the memoryless property that whenever a user experiences an idle (non-transmitting) period he renews his action independently of the past happenings. Then we can find the distribution of X explicitly for all of the above-listed protocols. Unfortunately, we could not obtain the distribution of X for a finite population of pure ALOHA users. The reason for this clumsiness is that we were unable to derive the distribution of the duration of an unsuccessful transmission period for pure ALOHA. (We identify this duration with the busy period in a queueing system with a finite input population, constant service time and an infinite number of servers.)

We may assume that each user has a different value for its transmission parameter such as the probability of starting transmission in a slot. Such a case occurs, for example, in the priority-based ordering of users, or the adaptive self-adjustment of parameter values according to the imposed load.

Using the mean and variance of the packet interdeparture times from the system, we can get the average and variance of the number of successful transmissions in a given long interval. Furthermore, we can relate these system-wide quantities to the means and covariances of the numbers of successful transmissions from the individual users. They can then be used to determine the coefficients in the diffusion process approximation to the user's queue length distribution (see Chapter 5).

As related work, we note Tobagi's analysis of packet interdeparture time based on the 'linear feedback model' of slotted ALOHA and slotted CSMA [Toba82a]. For pure ALOHA, Ferguson [Ferg77a] gives an approximation to the packet interdeparture time for a randomly selected user.

3.2 The Number of Successful Transmissions

Let $\{X^{(n)};\ n=1,2,\cdots\}$ be a sequence of packet interdeparture times (from the entire system) beginning at the end of an arbitrarily chosen successful transmission (let this instant be the time origin $t=0$.) They are assumed to be independent and identically distributed; their generic representation is X. Then

$$S^{(n)} = X^{(1)} + X^{(2)} + \cdots + X^{(n)} \qquad n=1,2,\cdots \tag{3.1}$$

defines the time at which the nth successful transmission completes. By definition, $\{S^{(n)};\ n=1,2,\cdots\}$ is a renewal process; see, for example, [Karl75]. For time $t > 0$, let $D(t)$ be the number of successful transmissions completed during an interval $[0,t]$:

$$D(t) = \max\{n\,;\, S^{(n)} \leqslant t\} \tag{3.2}$$

Now renewal theory tells us

$$\lim_{t\to\infty} \frac{\overline{D(t)}}{t} = \frac{1}{\bar{X}}\;; \qquad \lim_{t\to\infty} \frac{Var[D(t)]}{t} = \frac{Var[X]}{\bar{X}^3} \tag{3.3}$$

Thus the asymptotic behavior of $\overline{D(t)}$ and $Var[D(t)]$ can be obtained from $\bar{X}$ and $Var[X]$.

Next, let us assume that $M < \infty$ is the number of users in the system whose transmission parameters are not necessarily identical. They are indexed as $1,2,\cdots,M$. Let $D_i(t)$ be the number of successful transmissions completed by user i during $[0,t]$ $(i=1,2,\cdots,M)$. Let q_i be the probability that a successful transmission is achieved by user i; $\sum_{i=1}^{M} q_i = 1$. In the case where all users have the same parameter value, we have $q_i = 1/M$. Then it can be readily shown (see Appendix A) that the means and covariances of $[D_1(t),\ D_2(t),\ \cdots,\ D_M(t)]$ are given by

$$\overline{D_i(t)} = q_i\,\overline{D(t)}\;,$$

$$Cov[D_i(t),D_j(t)] = q_i q_j\{Var[D(t)] - \overline{D(t)}\} + \delta_{ij} q_i\,\overline{D(t)} \qquad i,j = 1,2,\cdots,M \tag{3.4}$$

where

$$\delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i\neq j \end{cases}$$

We note that the dependence of $D_i(t)$ and $D_j(t)$ $(i\neq j)$ comes from the fact that when user i is successful user j is not and vice versa. Therefore, by use of Eqs.(3.3) and (3.4), we can obtain the asymptotic behavior of the numbers of successful transmissions by individual users.

We remark that if $\{X_i^{(n)};\ n=1,2,\cdots\}$ is a sequence of packet interdeparture times from user i $(i=1,2,\cdots,M)$ beginning at $t=0$ (which can be the completion time of some other user), then $\{S_i^{(n)};\ n=1,2,\cdots\}$ where $S_i^{(n)} = X_i^{(1)} + X_i^{(2)} + \cdots + X_i^{(n)}$ $(n=1,2,\cdots)$, is a delayed renewal process, i.e., the distribution of $X_i^{(1)}$ is not identical to those of $X_i^{(2)}$, $X_i^{(3)}$, $\cdots$ which are identical and generically denoted by X_i. However, for $D_i(t) = \max\{n\,;\, S_i^{(n)} \leqslant t\}$, we still have the asymptotes

$$\lim_{t\to\infty} \frac{\overline{D_i(t)}}{t} = \frac{1}{\bar{X}_i}\;; \qquad \lim_{t\to\infty} \frac{Var[D_i(t)]}{t} = \frac{Var[X_i]}{\bar{X}_i^3} \qquad i=1,2,\cdots,M \tag{3.5}$$

Substituting Eqs.(3.3) and (3.5) into Eq.(3.4) for $i=j$, we get the relationship

$$S_i = q_i S\;; \qquad 1 - C_i^2 = q_i(1 - C^2) \qquad i=1,2,\cdots,M \tag{3.6}$$

where $C_i^2 = Var[X_i]/\bar{X}_i^2$, and $S_i = 1/\bar{X}_i$ is the throughput of user i. In the case of identical users $(q_i = 1/M)$, Eq.(3.6) reduces to

$$S = M S_i\;; \qquad 1 - C^2 = M(1 - C_i^2) \qquad i=1,2,\cdots,M \tag{3.7}$$

We note that Eqs.(3.6) and (3.7) can also be derived by considering random splitting of a non-Poisson stream as shown in [Sevc77] for the case of two-way splitting, and in [Kueh79]

generally.

Except for the case of slotted ALOHA, it is very difficult to evaluate C_i^2 by considering the individual output process for user i. Therefore, we discuss the packet interdeparture times from the entire system to find $\bar{X}$ and $Var[X]$. We also note for perfect scheduling or for an M/D/1 queueing system that we have $X = 1$ so that $\bar{X} = 1$ and $C^2 = 0$. Then from Eq.(3.6), we have $\bar{X}_i = 1/q_i$ and $C_i^2 = 1 - q_i$. In this case the distribution of X_i is geometric as well as X.

3.3 Output Processes for Identical Users

In this section, we consider the packet interdeparture time X for a population of users with identical transmission parameters. Before looking at the individual protocol cases, let us discuss the distribution of X in a unified manner. To do so, we define the transmission period in channel as the state where at least one user is transmitting or any transmission is being sensed. Also, the channel idle period is defined as the state where no users are transmitting or no transmissions are being sensed. Thus, the channel state alternates between the transmission and idle periods. (There can be two consecutive transmission periods with an idle period of duration 0 between them.) Now, let K be the number of transmission periods included in X of which the last one is the only successful transmission. Let $I^{(k)}$ and $F^{(k)}$ be the durations of the kth idle period and kth unsuccessful transmission period, respectively, and T be the duration of the successful transmission period. Then, we have

$$X = \sum_{k=1}^{K-1} [\, I^{(k)} + F^{(k)} \,] + I^{(K)} + T \tag{3.8}$$

Since we have assumed the memoryless property in protocol, the beginning of each idle period is a system renewal point (i.e., the behavior of the system after that point does not depend on what happened before that point). Therefore, $\{ I^{(k)};\ k = 1,2,\cdots \}$ are independent and identically distributed random variables whether users are identical or not; let I be a generic representation of the $I^{(k)}$'s. For the same reason, the sequence of $\{ F^{(k)};\ k = 1,2,\cdots \}$, consisting of $F^{(k)}$ following $I^{(k)}$, are independent variables. Furthermore, in a system of users with identical transmission parameters, they are also identically distributed because each of them does not depend upon who has initiated the transmission period. So, let F be the generic representation of the $F^{(k)}$'s. Thus, a sequence of renewal cycle durations $\{ I^{(k)} + F^{(k)};\ k = 1,2,\cdots \}$ are independent and identically distributed as $I + F$. Also, $I^{(K)} + T$ is independent of the previous cycles and is distributed as $I + T$. By these arguments, we can compute the mean and variance of X directly as

$$\bar{X} = (\bar{K} - 1)(\bar{I} + \bar{F}) + \bar{I} + \bar{T},$$

$$Var[X] = (\bar{K}-1)\,Var[I+F] + (\bar{I}+\bar{F})^2\,Var[K] + Var[I+T]$$

$$= \bar{K}\,Var[I] + (\bar{F}-1)\,Var[F] + Var[T] + (\bar{I}+\bar{F})^2\,Var[K] \qquad (3.9)$$

where we have assumed that $F^{(k)}$ and T are independent of $I^{(k)}$.

If we denote by $I^*(s)$, $F^*(s)$ and $T^*(s)$ the Laplace transforms of the pdf's for I, F and T, respectively, and denote by $K^*(z)$ the z-transform of the distribution of K, then under the same assumptions the Laplace transform of the pdf for X, $X^*(s)$, is given by

$$X^*(s) = \sum_{k=1}^{\infty} [I^*(s)F^*(s)]^{k-1}[I^*(s)T^*(s)] \cdot Prob\,[K=k]$$

or

$$X^*(s) = \frac{T^*(s)}{F^*(s)}\,K^*[I^*(s)F^*(s)] \qquad (3.10)$$

Now, let γ be the probability of a successful transmission once it has been started by breaking the channel idle period (γ is a protocol-dependent function of M and other system parameters). Then, clearly, K has a geometric distribution

$$Prob\,[K=k] = (1-\gamma)^{k-1}\gamma \qquad k=1,2,\cdots$$

$$K^*(z) = \frac{\gamma z}{1-z(1-\gamma)},$$

$$\bar{K} = \frac{1}{\gamma} \;; \qquad Var[K] = \frac{1-\gamma}{\gamma^2} \qquad (3.11)$$

From Eqs.(3.10) and (3.11), we obtain the fundamental relationship

$$X^*(s) = \frac{\gamma\,T^*(s)}{\dfrac{1}{I^*(s)} - (1-\gamma)\,F^*(s)} \qquad (3.12)$$

Therefore, given a protocol, $\bar{X}$ and C^2 can be computed by Eqs.(3.9) and (3.11) if we obtain γ and the means and variances of I, F and T depending on the protocol. Also, by Eq.(3.12), $X^*(s)$ can be obtained from γ and the distributions of I, F and T.

3.3.1 Slotted ALOHA

Let us begin our study with slotted ALOHA where the slot size is unity (i.e., the constant packet transmission time). For an illustration of the packet interdeparture time, see Figure 3.1(a). Let each of M users transmit in any slot with probability p independently of all others. The condition for a successful transmission once it is transmitted is that there be no other simultaneous transmissions. Thus we have

$$\gamma = U/(1-E) \tag{3.13}$$

where

$$U = pM(1-p)^{M-1} \; ; \qquad E = (1-p)^M \tag{3.14}$$

The idle period is geometrically distributed as

$$Prob[I=n] = E^n(1-E) \qquad n=0,1,2,\cdots$$

$$I^*(s) = \frac{1-E}{1-e^{-s}E} \; ,$$

$$\bar{I} = \frac{E}{1-E} \; ; \qquad Var[I] = \frac{E}{(1-E)^2} \tag{3.15}$$

We have the constant transmission periods:

$$F = T = 1 \quad \text{or} \quad F^*(s) = T^*(s) = e^{-s} \tag{3.16}$$

Thus, from Eqs.(3.9), (3.11) and (3.12), we get

$$X^*(s) = \frac{e^{-s}U}{1-e^{-s}(1-U)} \; ,$$

$$S = U \; ; \qquad C^2 = 1-U \tag{3.17}$$

We note that this result could be obtained directly from the geometric distribution of X:

$$Prob[X=n] = (1-U)^{n-1}U \qquad n=1,2,\cdots \tag{3.18}$$

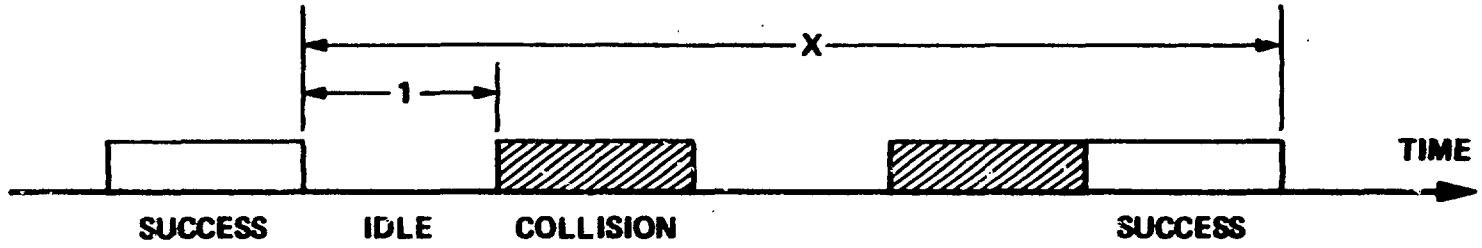
For $M \to \infty$ while holding $G = pM$ at a fixed finite value, we have $U \to Ge^{-G}$ to get

$$S = Ge^{-G} \; ; \qquad C^2 = 1-Ge^{-G} \tag{3.19}$$

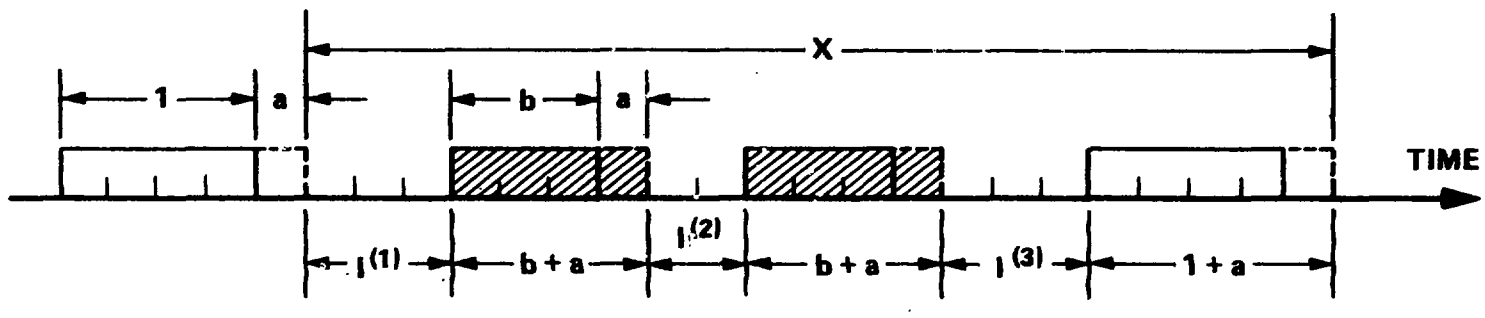
The equation for C^2 is a new result. For $G=1$ which maximizes S, we have $S = 1/e = 0.3679$ and $C^2 = 1-1/e = 0.6321$.

3.3.2 Pure ALOHA (Infinite Population)

For pure (or unslotted) ALOHA where all packets have the same length 1, we can analytically obtain the distribution of X (and C^2) only for an infinite population of users who collectively form a Poisson source of packet transmissions. (The throughput can be analytically obtained for a finite number of nonidentical users.) This is due to the difficulty in finding the distribution of F for a finite population of users.

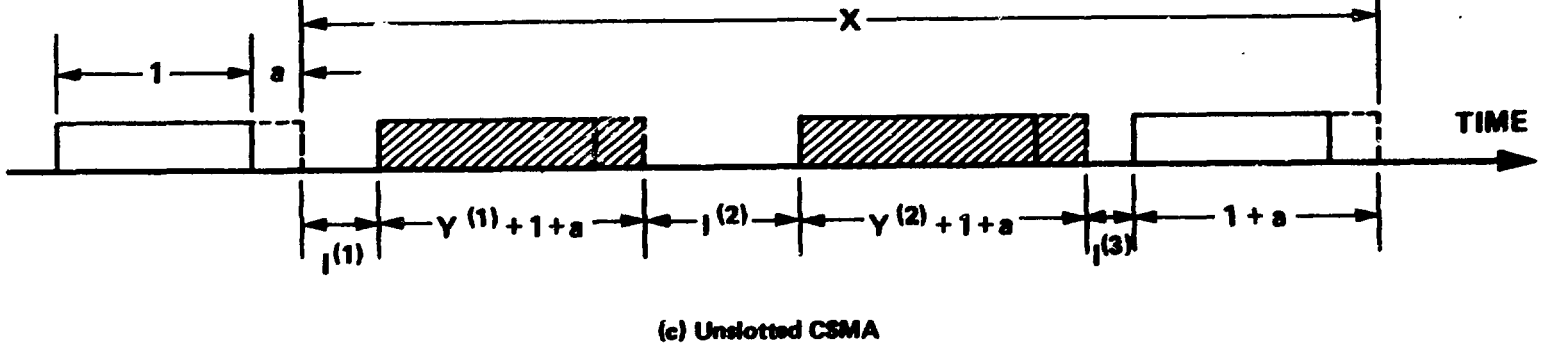


(a) Slotted ALOHA

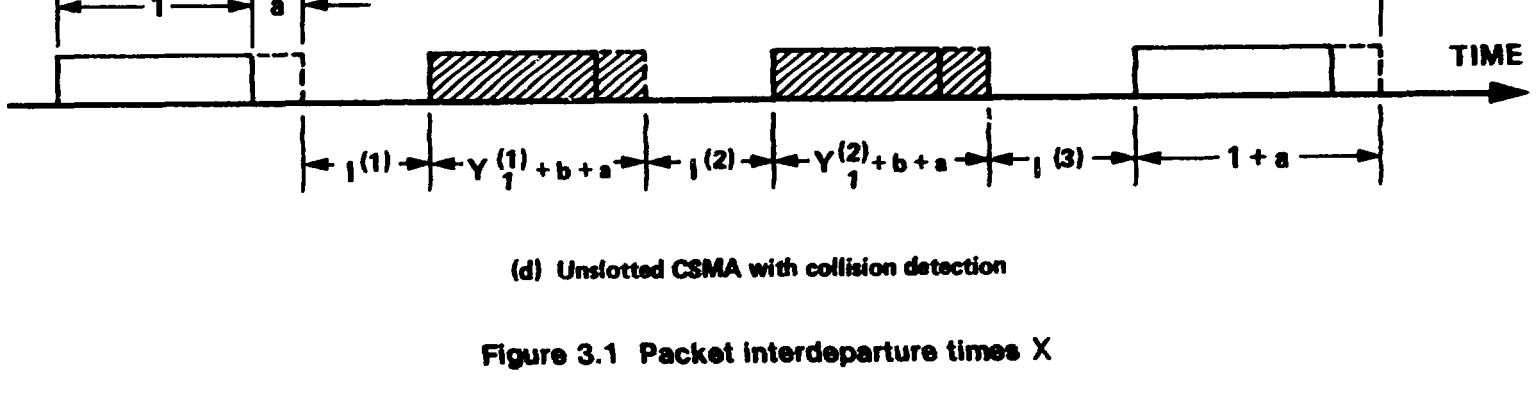


(b) Slotted CSMA with collision detection

Figure 3.1 Packet interdeparture times X



(c) Unslotted CSMA



(d) Unslotted CSMA with collision detection

Figure 3.1 Packet interdeparture times X

Let us consider an infinite population of users from which packets are transmitted such that the interarrival times are independent and exponentially (identically) distributed with mean $1/G$; see Figure 3.2(a). Then, from its memoryless property, the channel idle time is exponentially distributed in the same way:

$$Prob[I \leq y] = 1 - e^{-Gy} \qquad y \geq 0 \; ; \qquad I^*(s) = \frac{G}{G+s}$$

$$\bar{I} = \frac{1}{G} \; ; \qquad Var[I] = \frac{1}{G^2} \tag{3.20}$$

The probability of a successful transmission of an idle-period-breaking packet is given by

$$\gamma = e^{-G} \tag{3.21}$$

i.e., the probability of no other transmissions during its entire transmission period $T = 1$ ($T^*(s) = e^{-s}$). The duration of an unsuccessful transmission period F (see Figure 3.2(b)) is analyzed in Appendix B where we derive

$$F^*(s) = \frac{G\,e^{-(s+G)}[1-e^{-(s+G)}]}{(1-e^{-G})\,[s+G\,e^{-(s+G)}]} \; ,$$

$$\bar{F} = \frac{e^G - 1 - G\,e^{-G}}{G\,(1-e^{-G})} \; ; \quad Var[F] = \frac{e^{2G}}{G^2} - \frac{2e^G}{G} - \frac{e^{-G}}{(1-e^{-G})^2} \tag{3.22}$$

Substituting Eqs.(3.20)-(3.22) into Eqs.(3.9), (3.11) and (3.12), we get

$$X^*(s) = \frac{G\,e^{-(s+G)}\,[s+G\,e^{-(s+G)}]}{s^2 + sG\,[1+e^{-(s+G)}] + G^2 e^{-2(s+G)}} \; ,$$

$$S = G\,e^{-2G} \; ; \quad C^2 = 1 + 2\,e^{-G} - 2\,e^{-2G} - 4\,G\,e^{-2G} \tag{3.23}$$

We note that the result for C^2 is new while the expression for S is given in [Abra70]. For $G = \frac{1}{2}$, which maximizes S, we have $S = 1/(2e) = 0.1839$ and $C^2 = 0.7415$. In Figure 3.2(c), we show the pdf for X obtained by the numerical inversion (using a formula in [Dubn68]) of $X^*(s)$ in Eq.(3.23) with $G = \frac{1}{2}$. It appears to decrease monotonically from a peak near $X = 1$. However the rate of decrease is smaller than that of the probability mass for slotted ALOHA in Eq.(3.19) with $G = 1$.

3.3.3 Slotted CSMA and CSMA with Collision Detection

We now proceed to analyze slotted CSMA where the slot size is equal to a, the ratio of the signal propagation delay to the packet transmission time. We consider CSMA/CD such that an unsuccessful transmission period lasts $b + a$, where $a \leq b \leq 1$; X is illustrated in Figure 3.1(b). Thus the case $b = 1$ corresponds to CSMA without collision detection.

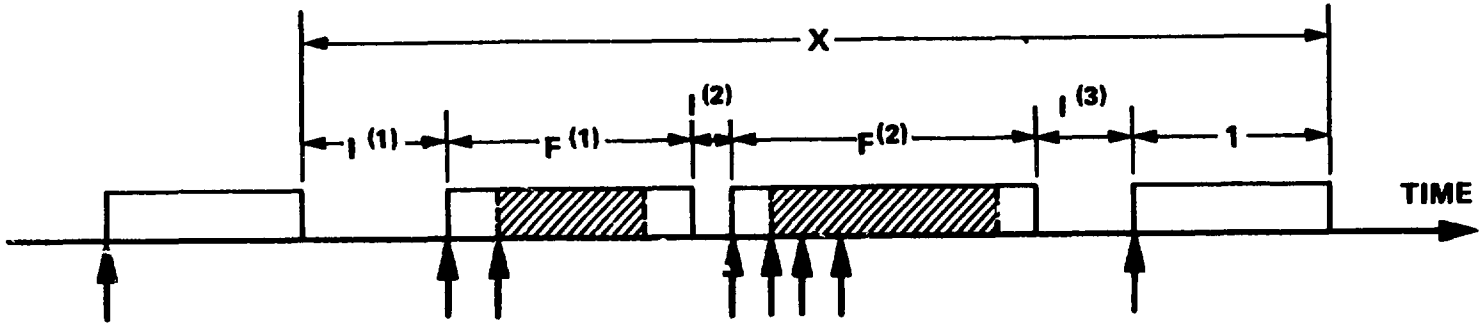


Figure 3.2(a) Packet interdeparture time X in pure ALOHA

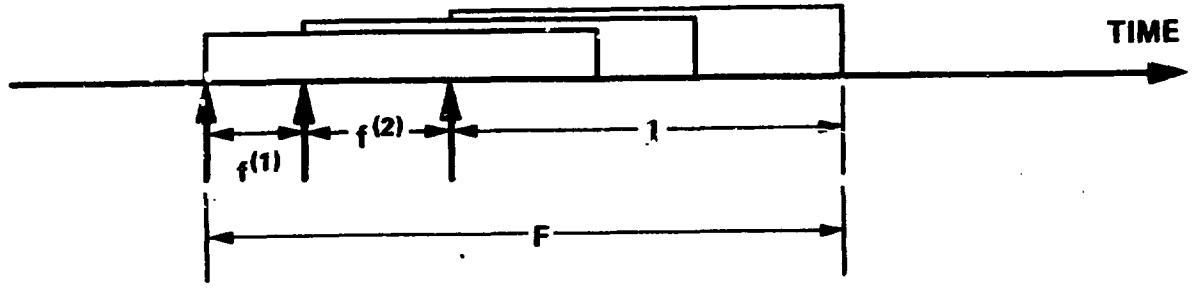


Figure 3.2(b) Unsuccessful transmission period in pure ALOHA

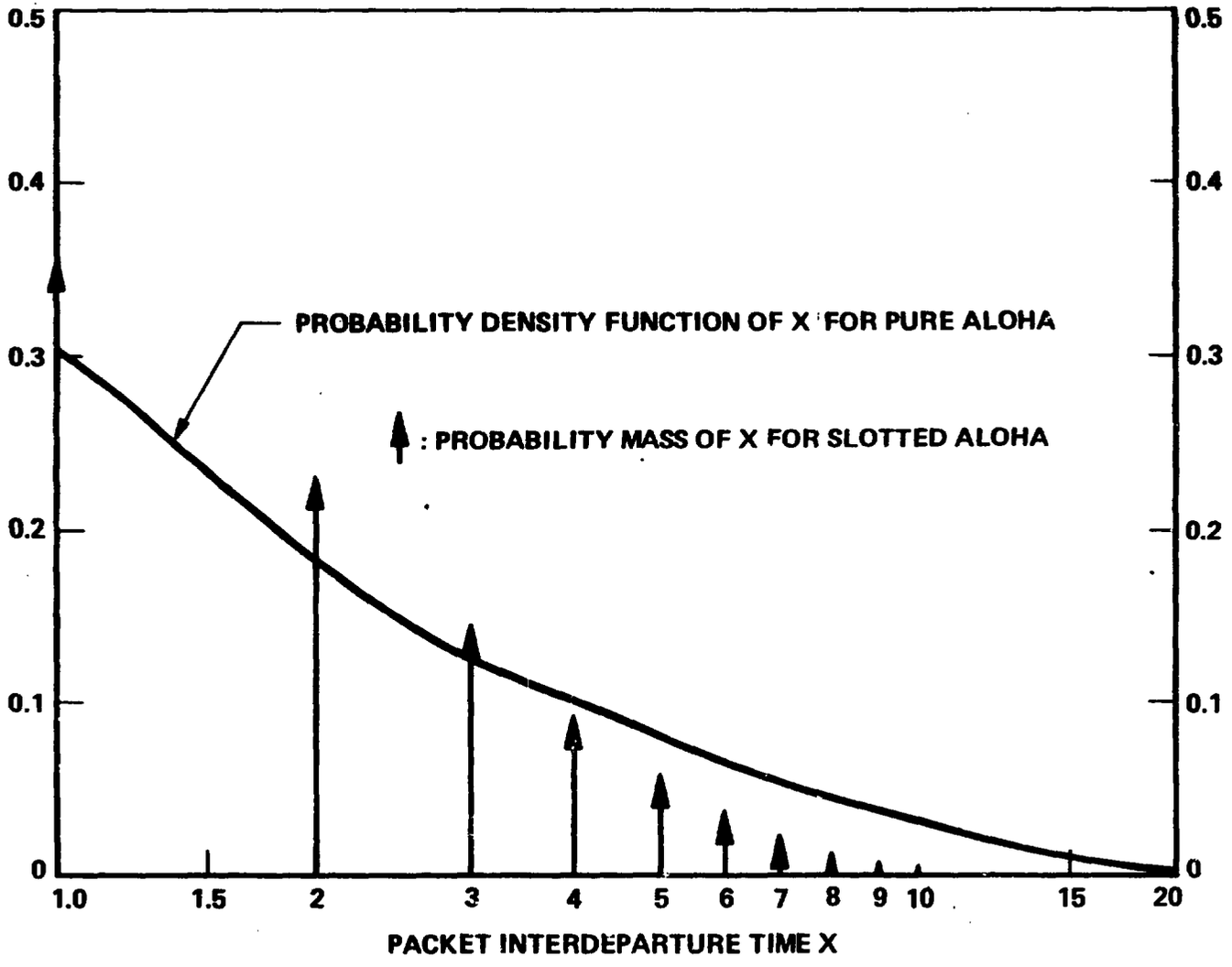


Figure 3.2(c) Probability density function of packet interdeparture time in pure ALOHA

Let each of M users start to transmit (after sensing any idle slot) with probability p independently of all others. Such a transmission is successful if none of other users have started transmission at the same time. (After the first slot of transmission period, no other users start transmission because they sense the channel busy.) Thus, we have the expression for γ as in Eqs.(3.13) and (3.14). The channel idle period is geometrically distributed as

$$Prob\,[\,I = na\,] = E^n(1-E) \qquad n = 0,1,2,\cdots$$

$$I^*(s) = \frac{1-E}{1-e^{-as}E} \tag{3.24}$$

The transmission periods are of constant length:

$$T = 1 + a \;; \qquad F = b + a$$

or

$$T^*(s) = e^{-s(1+a)} \;; \quad F^*(s) = e^{-s(b+a)} \tag{3.25}$$

Substituting Eqs.(3.13), (3.24) and (3.25) into Eqs.(3.9) and (3.11), we get

$$S = \frac{U}{a + U + b(1-U-E)},$$

$$Var[X] = \frac{[a + b(1-E)]^2}{U^2} + \frac{b^2E - (b+a)^2}{U} \tag{3.26}$$

where U and E are given by Eq.(3.14). From Eq.(3.12), we also get

$$X^*(s) = \frac{U e^{-s(1+a)}}{1 - e^{-as}E - e^{-s(b+a)}(1-U-E)}$$

from which we have

$$Prob\,[\,X = 1 + a + na + k(b+a)\,] = U\binom{n+k}{k} E^n (1-U-E)^k \qquad n,k = 0,1,2,\cdots \tag{3.27}$$

The implication of Eq.(3.27) should be clear since n and k are the numbers of idle slots and unsuccessful transmission periods, respectively, experienced until the time of a successful transmission.

We may note that the value of p which maximizes S can be obtained as a solution to the equation

$$(a+b)(1-pM) = b(1-p)^M \tag{3.28}$$

which was derived by Molle [Moli81] in his study of the 'local optimality' condition. It can be proved that with the value of p determined by Eq.(3.28), we have $C^2 < 1$. In the case of CSMA without collision detection ($b = 1$), we have

$$S = \frac{U}{1 + a - E} \;; \quad C^2 = 1 - U \cdot \frac{(1+a)^2 - E}{(1+a-E)^2} \tag{3.29}$$

The channel throughput S in Eq.(3.26) in the limit $M \to \infty$ with G fixed such that $aG = pM$:

$$S = \frac{aG\, e^{-aG}}{a + aG\, e^{-aG} + b\,(1 - aG\, e^{-aG} - e^{-aG})} \tag{3.30}$$

has been obtained in [Toba80d]. The result for the case of no collision detection ($b = 1$) was given incorrectly in [Toba74] and was corrected in [Klei75b]. (The plot in [Klei75b], however, was for the result in [Toba74]; the corrected plot was given by Molle [Moll81].)

3.3.4 Unslotted CSMA

We next consider three variants of unslotted CSMA where the propagation delay is again denoted by a. Let us begin with the basic unslotted nonpersistent CSMA (without collision detection and without capture effect); for an illustration of X, see Figure 3.1(c). We assume that each of M users schedules his next transmission at an exponentially distributed time after he has sensed the channel idle. Let $1/g$ be the mean of this exponential distribution. Since the channel idle time I is the minimum of all user's scheduling delays, its distribution is given by

$$Prob\,[\,I \leqslant y\,] = 1 - e^{-gMy} \qquad y \geqslant 0$$

$$\bar{I} = \frac{1}{gM}\ ; \qquad Var[I] = \frac{1}{(gM)^2} \tag{3.31}$$

A transmission started by breaking the channel idle period is successful if none of other users start transmission within time a. Therefore, the probability of its success is given by

$$\gamma = e^{-ga(M-1)} \tag{3.32}$$

The duration of a successful transmission period is constant:

$$T = 1 + a \tag{3.33}$$

while that of an unsuccessful transmission period is expressed as

$$F = 1 + a + Y \tag{3.34}$$

where Y is the transmission starting time of the last colliding packet. The distribution of Y can be calculated to yield

$$Prob\,[\,Y \leqslant y \mid \text{collision}\,] = \frac{(1 - e^{-gy} + e^{-ga})^{M-1} - e^{-ga(M-1)}}{1 - e^{-ga(M-1)}} \qquad 0 \leqslant y \leqslant a \tag{3.35}$$

where the factor $(1 - e^{-gy} + e^{-ga})$ accounts for the probability that each of the $M-1$ users (who behave independently until time a) either starts transmission within y or does not until a. From Eq.(3.35) $\bar{Y}$ and $Var[Y]$ are numerically evaluated.

Using Eqs.(3.31)-(3.35) in Eqs.(3.9) and (3.11), we have the channel throughput of unslotted CSMA as

$$S = \frac{e^{-ga(M-1)}}{\frac{1}{gM} + 1 + 2a - \int_0^a (1 - e^{-gy} + e^{-ga})^{M-1}\, dy} \tag{3.36}$$

We leave out the complicated expression for C^2.

For $M \to \infty$ with G fixed at $G = gM$, since $(1 - e^{-gy} + e^{-ga})^{M-1} \simeq [1 - g(a-y)]^{M-1} \simeq e^{-G(a-y)}$, Eq.(3.36) reduces to

$$S = \frac{G\, e^{-aG}}{G(1+2a) + e^{-aG}} \tag{3.37}$$

given in [Toba74,Klei75b]. In this limit, the distribution and variance of X are given by

$$X^*(s) = \frac{G(s-G)\, e^{-s(1+a)-aG}}{(s+G)(s-G) - G^2 e^{-aG-s(1+a)}[1 - e^{-(s-G)a}]},$$

$$Var[X] = \frac{2 - e^{-aG}}{G^2 e^{-aG}} + \frac{(1+2a)^2}{e^{-2aG}} - \frac{1+2a}{e^{-aG}} \tag{3.38}$$

Here we may note that as $G \to 0$, we have $S \simeq G$ and $Var[X] \simeq 1/G^2$ so that $C^2 \simeq 1$.

3.3.5 Unslotted CSMA with Collision Detection

The case of unslotted CSMA with collision detection can be treated similarly except for the duration of an unsuccessful transmission period which is now expressed as

$$F = b + a + Y_1 \tag{3.39}$$

where b is the time required for an idle-period-breaking user to abort its transmission after the the first colliding packet has started transmission, and Y_1 is the time offset of the first colliding transmission; see Figure 3.3 for the channel timing chart (adapted from [Toba80d] for the unslotted system) and see Figure 3.1(d) for the packet interdeparture process. We assume that $a \leqslant b \lesssim 1$; the case $b = 1$ in unslotted system is not equivalent to the one without collision detection. (Our proposition for the duration of F given in Eq.(3.39) differs from that by Molle [Moll81] who used $F = b + a + Y$ where Y is the transmission start time of the last colliding packet. As shown in Figure 3.3, it is the first colliding packet that stops the transmission of the leading packet lingering till the last; other transmissions have been aborted before by detecting the leading packet. Thus, Eq.(3.39) seems more reasonable although the difference is of order a.) The distribution of Y_1 is given by

$$Prob[\, Y_1 > y \,|\, \text{collision}] = \frac{e^{-gy(M-1)} - \gamma}{1 - \gamma} \qquad 0 \leqslant y \leqslant a$$

$$\bar{Y}_1 = \frac{1}{g(M-1)} - \frac{a\gamma}{1-\gamma}\ ;\ \ Var[Y_1] = \frac{1}{g^2(M-1)^2} - \frac{a^2\gamma}{(1-\gamma)^2} \tag{3.40}$$

where γ is given in Eq.(3.32).

Substituting Eqs.(3.31)-(3.33), (3.39) and (3.40) into Eqs.(3.9), (3.11) and (3.12), we get

$$X^*(s) = \frac{e^{-ga(M-1)-s(1+a)}}{\frac{s+gM}{gM} - g(M-1)\,e^{-s(b+a)}\cdot\frac{1-e^{-[s+g(M-1)]a}}{s+g(M-1)}},$$

$$S = \frac{e^{-ga(M-1)}}{\frac{1}{gM} + e^{-ga(M-1)} + [\,b+a+\frac{1}{g(M-1)}\,][\,1-e^{-ga(M-1)}\,]},$$

$$Var[X] = \frac{1}{\gamma(gM)^2} + (\frac{1}{\gamma}-1)\,Var[Y_1] + (\frac{1}{gM}+b+a+\bar{Y}_1)^2\cdot\frac{1-\gamma}{\gamma^2} \tag{3.41}$$

The value of g which maximizes S in Eq.(3.41) is given as a solution to the equation

$$(2M-1)[1-ga(M-1)] = (1+\frac{b}{a})(ga)^2M(M-1)^2 + M\,e^{-ga(M-1)}$$

For $M \to \infty$ with G fixed at $G = gM$, we have

$$X^*(s) = \frac{G(s+G)\,e^{-aG-s(1+a)}}{(s+G)^2 - G^2e^{-s(b+a)}[\,1-e^{-(s+G)a}\,]}$$

$$S = \frac{G\,e^{-aG}}{2+(G-1)\,e^{-aG}+(b+a)\,G(1-e^{-aG})} \tag{3.42}$$

The optimal G for Eq.(3.42) is similarly determined by

$$2(1-aG) = (1+\frac{b}{a})(aG)^2 + e^{-aG}$$

We note in this limit that $\bar{Y}_1 = a - \bar{Y}$ and $Var[Y_1] = Var[Y]$ for Y in Eq.(3.35).

3.3.6 Unslotted CSMA with Delay Capture

Lastly, we consider unslotted CSMA with delay capture effect. Delay capture (available with spread spectrum modulation) is the capability of a receiver to successfully receive an idle-period-breaking packet even through most part of it arrives at the receiver overlapped in time by other packets [Davi80]. In this case, the distribution of the channel idle time is still given by Eq.(3.31). However, a transmission started by breaking the channel idle period is successful if none of other users start transmission within time c ($\leqslant a$). (The case $c = a$ is equivalent to CSMA without delay capture while the case $c = 0$ may be called perfect capture.) Therefore, the probability of success for an idle-period-breaking packet is given by

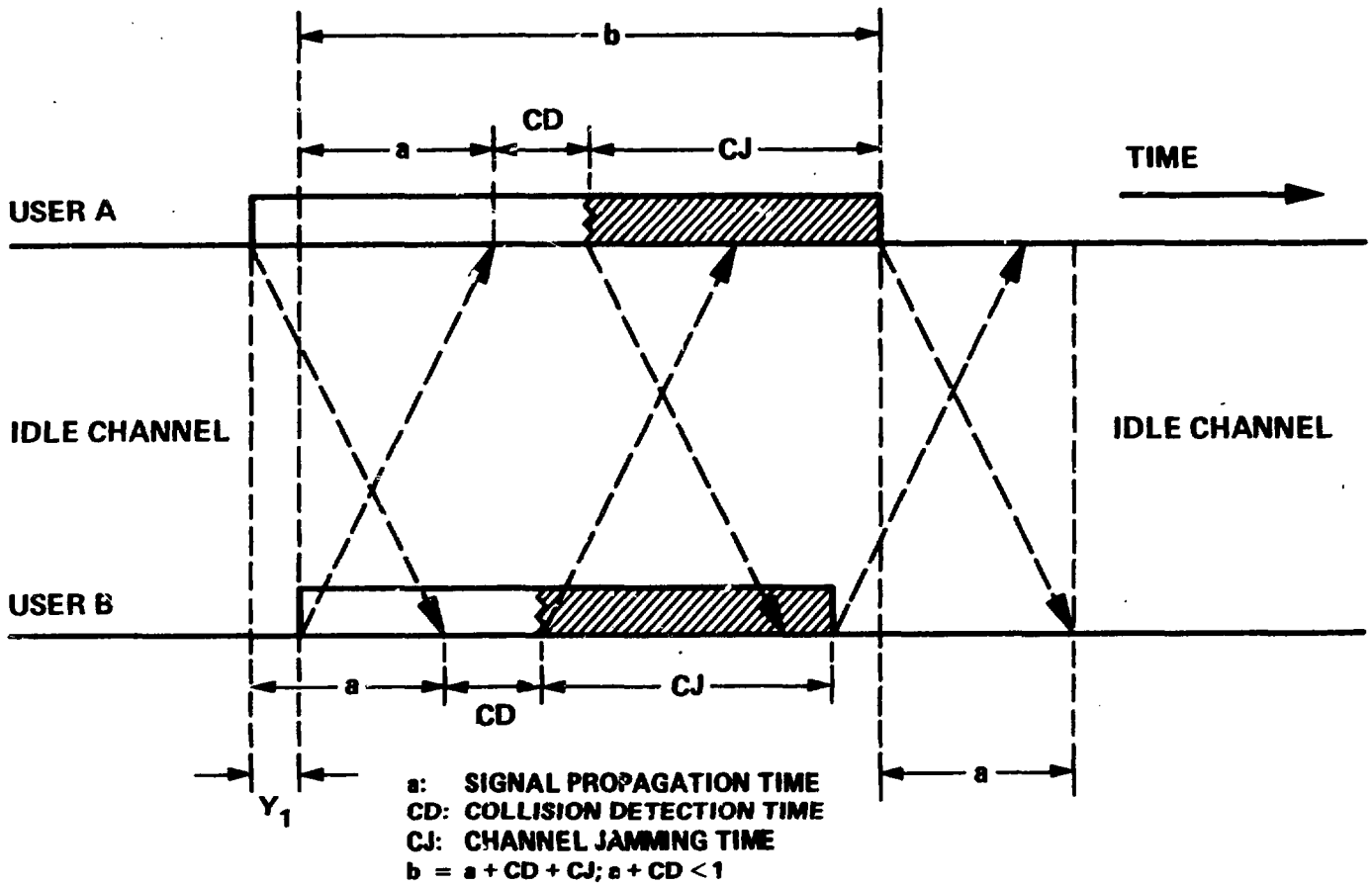


Figure 3.3 Collision detection timing in unslotted CSMA/CD

$$\gamma = e^{-gc(M-1)} \tag{3.43}$$

Note that the duration of a successful transmission period T is variable due to those transmissions which may occur after time c but before time a; see Figure 3.4(a). If Y_s denotes the transmission start time of the last overlapping packet in a successful transmission period, we have

$$T = 1 + a + Y_s \tag{3.44}$$

where Y_s is distributed as

$$Prob\,[\,Y_s \leqslant y\,] = \begin{cases} e^{-g(a-c)(M-1)} & 0 \leqslant y \leqslant c \\ [1 - e^{-g(y-c)} + e^{-g(a-c)}]^{M-1} & c \leqslant y \leqslant a \end{cases} \tag{3.45}$$

The duration of an unsuccessful transmission period F (see Figure 3.4(b)) is similarly expressed as

$$F = 1 + a + Y_c \tag{3.46}$$

where Y_c is the transmission start time of the last colliding packet, and is distributed as

$$Prob\,[\,Y_c \leqslant y\,] = \begin{cases} \dfrac{(1 - e^{-gy} + e^{-ga})^{M-1} - e^{-ga(M-1)}}{1 - \gamma} & 0 \leqslant y \leqslant c \\ \dfrac{(1 - e^{-gy} + e^{-ga})^{M-1} - (e^{-gc} - e^{-gy} + e^{-ga})^{M-1}}{1 - \gamma} & c \leqslant y \leqslant a \end{cases} \tag{3.47}$$

Eqs.(3.45) and (3.47) are derived in Appendix C.

From Eqs.(3.44)-(3.47), we can evaluate $\bar{T}$, $Var[T]$, $\bar{F}$ and $Var[F]$ numerically, and then through Eqs.(3.9) and (3.11) we can get $\bar{X}$ and $Var[X]$. In particular, we have

$$S = \frac{e^{-gc(M-1)}}{\dfrac{1}{gM} + 1 + 2a - \displaystyle\int_0^a (1 - e^{-gy} + e^{-gc})^{M-1}\,dy} \tag{3.48}$$

We note that the denominator in Eq.(3.48) is the same as that in Eq.(3.36) for unslotted CSMA without delay capture. This is because the presence of the capture phenomenon does not affect the channel cycle; it only changes the probability of successful transmission from $e^{-ga(M-1)}$ to $e^{-gc(M-1)}$.

For $M \to \infty$ with G fixed at $G = gM$, Eq.(3.48) becomes

$$S = \frac{G\,e^{-cG}}{G(1+2a) + e^{-aG}} \tag{3.49}$$

The optimal G which maximizes S in Eq.(3.49) is given as a solution to the equation

$$c\,G^2(1+2a) = [1 + (a-c)G]\,e^{-aG} \tag{3.50}$$

The distribution and the variance of X are given by

$$X^*(s) = \frac{G\,e^{-aG-s(1+a)}\{s-G[1-e^{-sc}(1-e^{-(a-c)(s-G)})]\}}{(s+G)(s-G)-G^2e^{-aG-s(1+a)}[1-e^{-a(s-G)}-e^{-sc}[1-e^{-(a-c)(s-G)}]]},$$

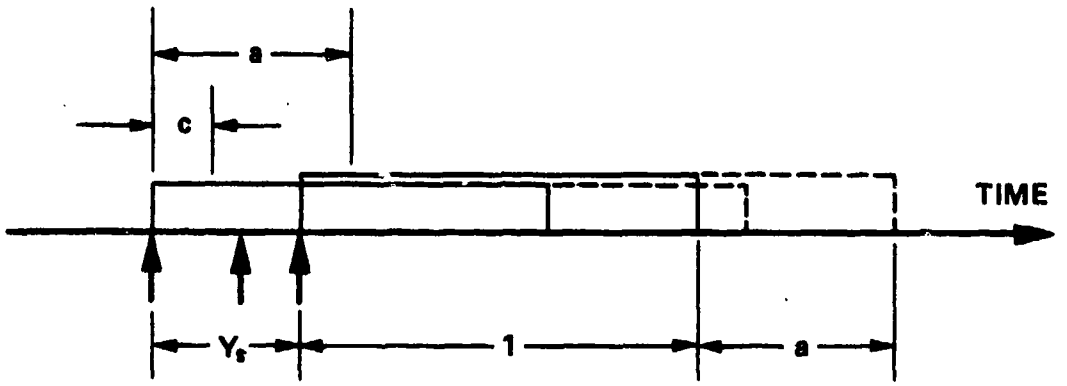
$$Var[X] = \frac{2e^{-cG}-e^{-2aG}}{G^2e^{-2cG}} + \frac{(1+2a)^2}{e^{-2cG}} - \frac{1+2a}{e^{-cG}} - \frac{2(ae^{-cG}-ce^{-aG})}{e^{-2cG}}\left(\frac{e^{-aG}}{G}+1+2a\right) \tag{3.51}$$

3.3.7 Discussions of the Numerical Results

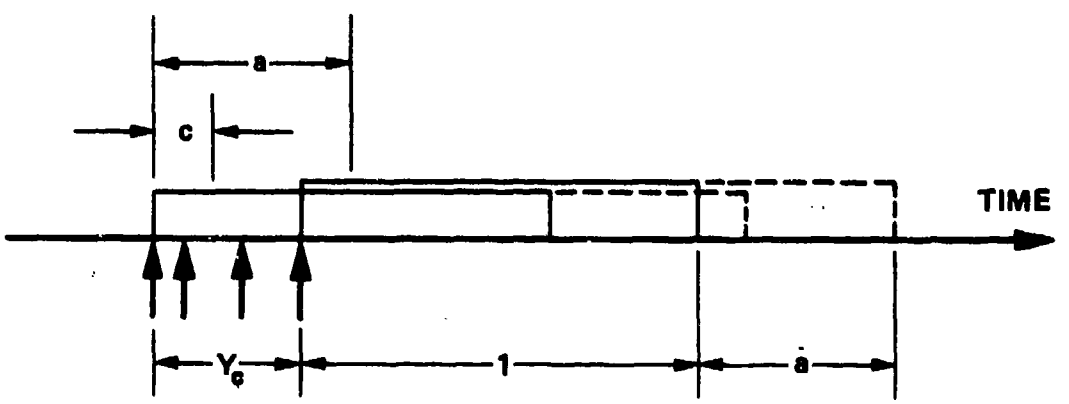
In Figure 3.5(a), we plot the throughput values for the protocols we have studied when they are maximized by optimizing the transmission parameters (e.g., p and g). The results shown are for the limit $M \to \infty$ while holding pM or gM at a fixed finite value. (The curves for CSMA/CD are for the ideal case: $b = a$.) For proper comparison between ALOHA and CSMA in an environment of non-zero propagation delay ($a > 0$), we have uniformly assumed that the duration of a successful transmission period is $1 + a$. Thus, the plots for ALOHA systems are $S/(1+a)$ where S is given by Eq.(3.19) or (3.23). The throughput for perfect scheduling is similarly assumed to be $1/(1+a)$. These maximal throughput curves have been studied in [Toba74,Klei75b,Moll81]. The new results in the present chapter are the corresponding plots in Figure 3.5(b) of the coefficient of variation of the packet interdeparture times. It is remarkable that (at maximal throughputs) they are all below 1. Specifically, in efficient CSMA cases they are almost 0 which implies that the channel service time is nearly constant. (This makes it difficult to numerically invert $X^*(s)$ due to the Gibbs' phenomenon [Otne78].)

In Figures 3.6(a) and 3.6(b), we show S and C^2 for unslotted CSMA with the delay capture effect. Davis and Gronemeyer [Davi80] give an example of a ground radio packet network where the packet duration is 20 msec, the propagation time is 200 μsec and the capture time is 10 μsec. Then we have $a = 0.01$ and $c = 0.0005$. Therefore, in this example, we obtain the throughput of 94.55 % and $C^2 = 0.0215$.

The throughput S and C^2 for slotted CSMA/CD are displayed in Figure 3.7 with several values of collision detection time b ($a \leqslant b \leqslant 1$). Again, for a typical example of a local-area computer network [Toba80d] where a is from 0.01 to 0.1 and b is short (< 0.1, say), we have very small values of C^2 despite the fact that throughput is more or less degraded. This observation suggests an M/D/1 queueing model (with service rate equated to the given channel throughput) approximation to the queue length distribution.



(a) Successful transmission period



(b) Unsuccessful transmission period

Figure 3.4 Transmission periods in unslotted CSMA with delay capture

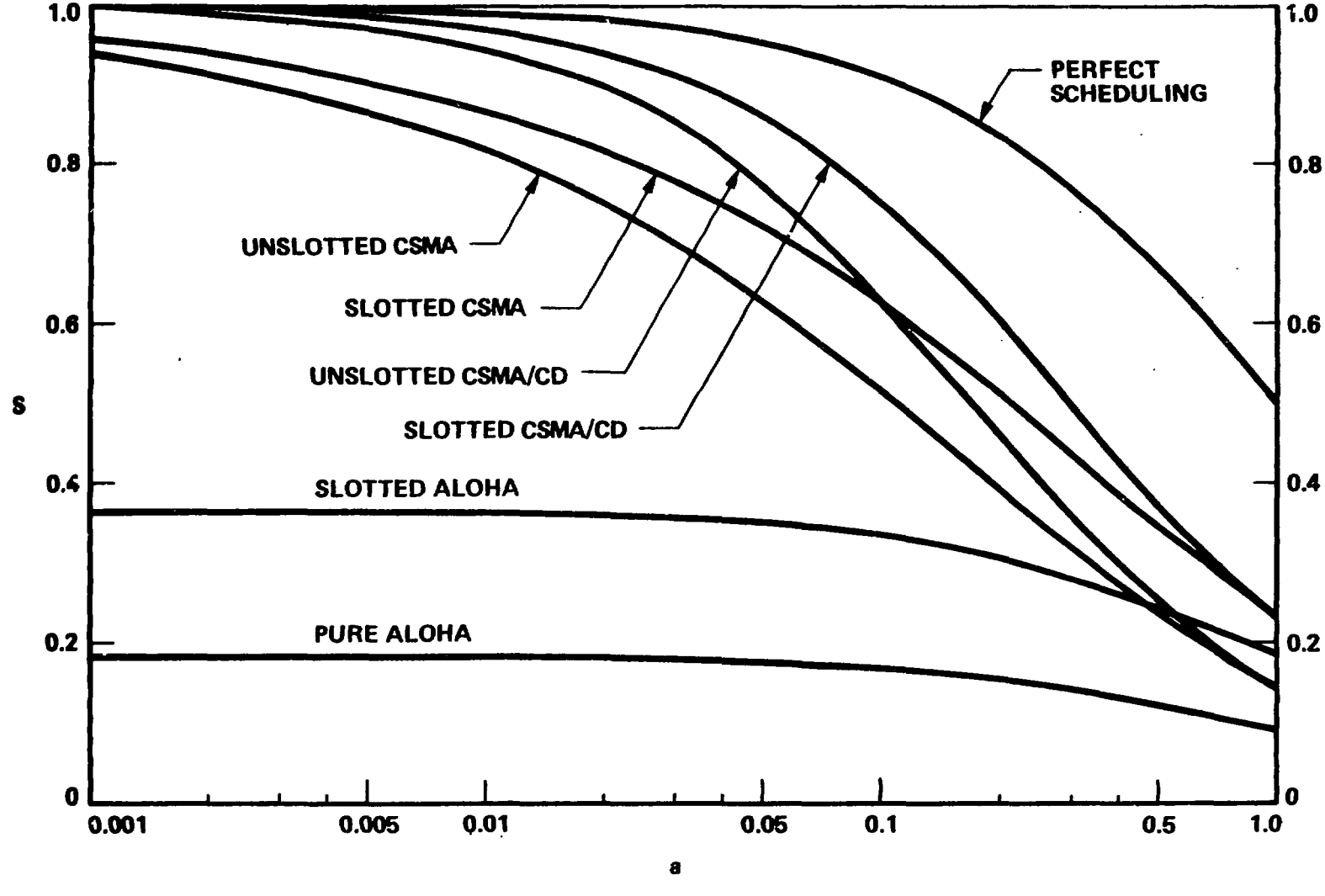


Figure 3.5(a) The (maximized) channel throughput S

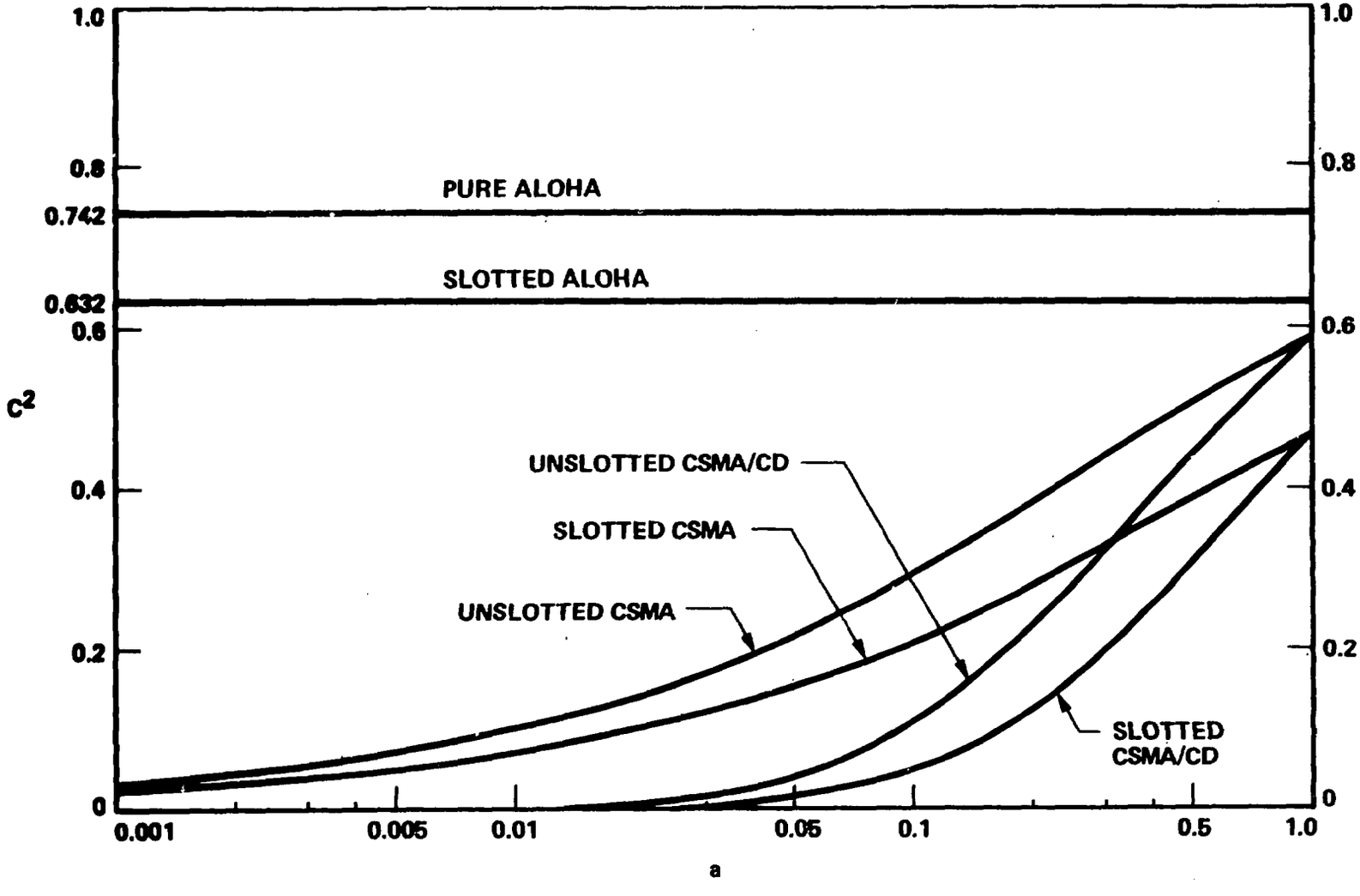


Figure 3.5(b) The coefficient of variation C^2 for packet interdeparture time

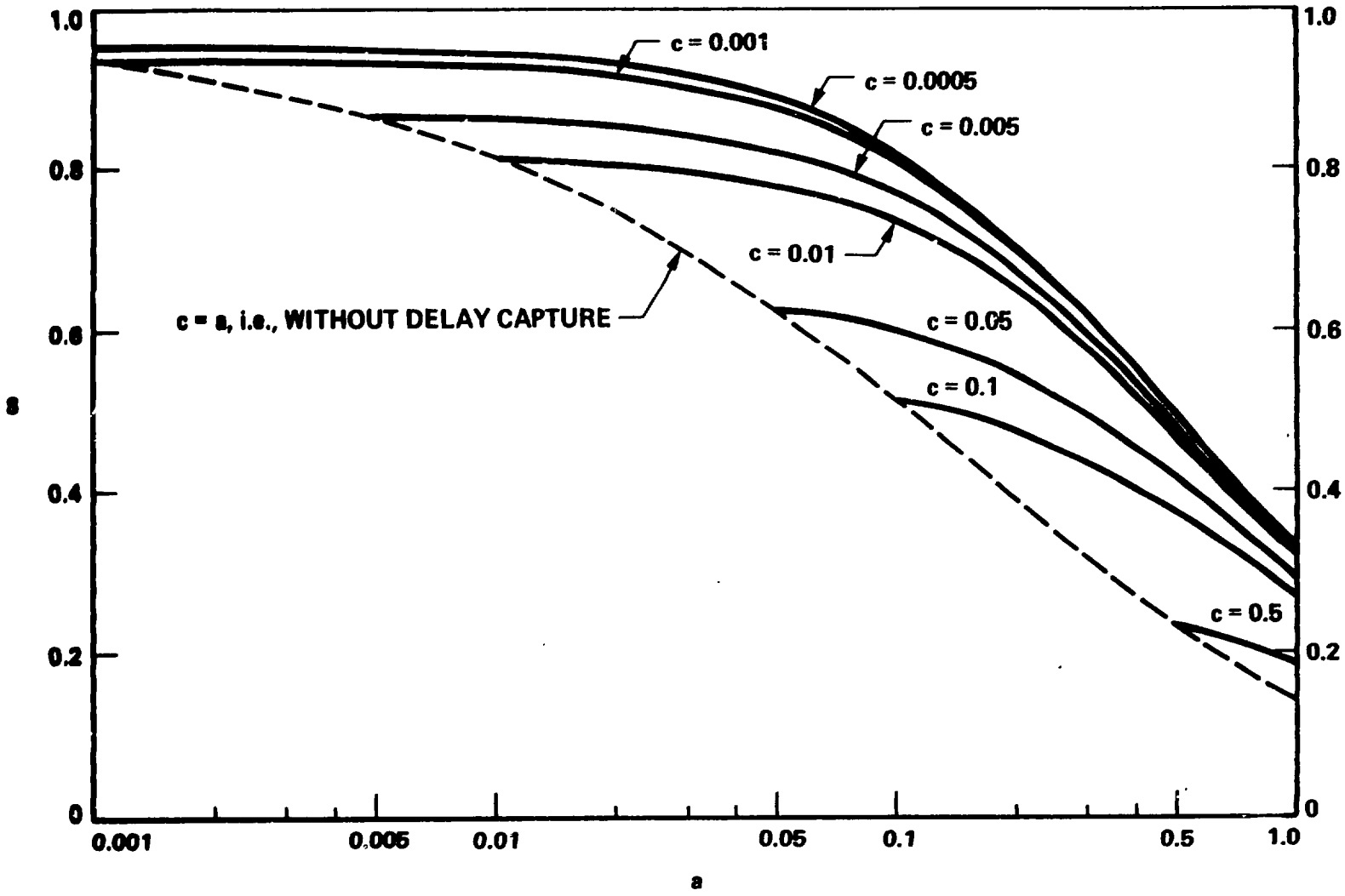


Figure 3.6(a) The channel throughput (S) of unslotted CSMA with delay capture

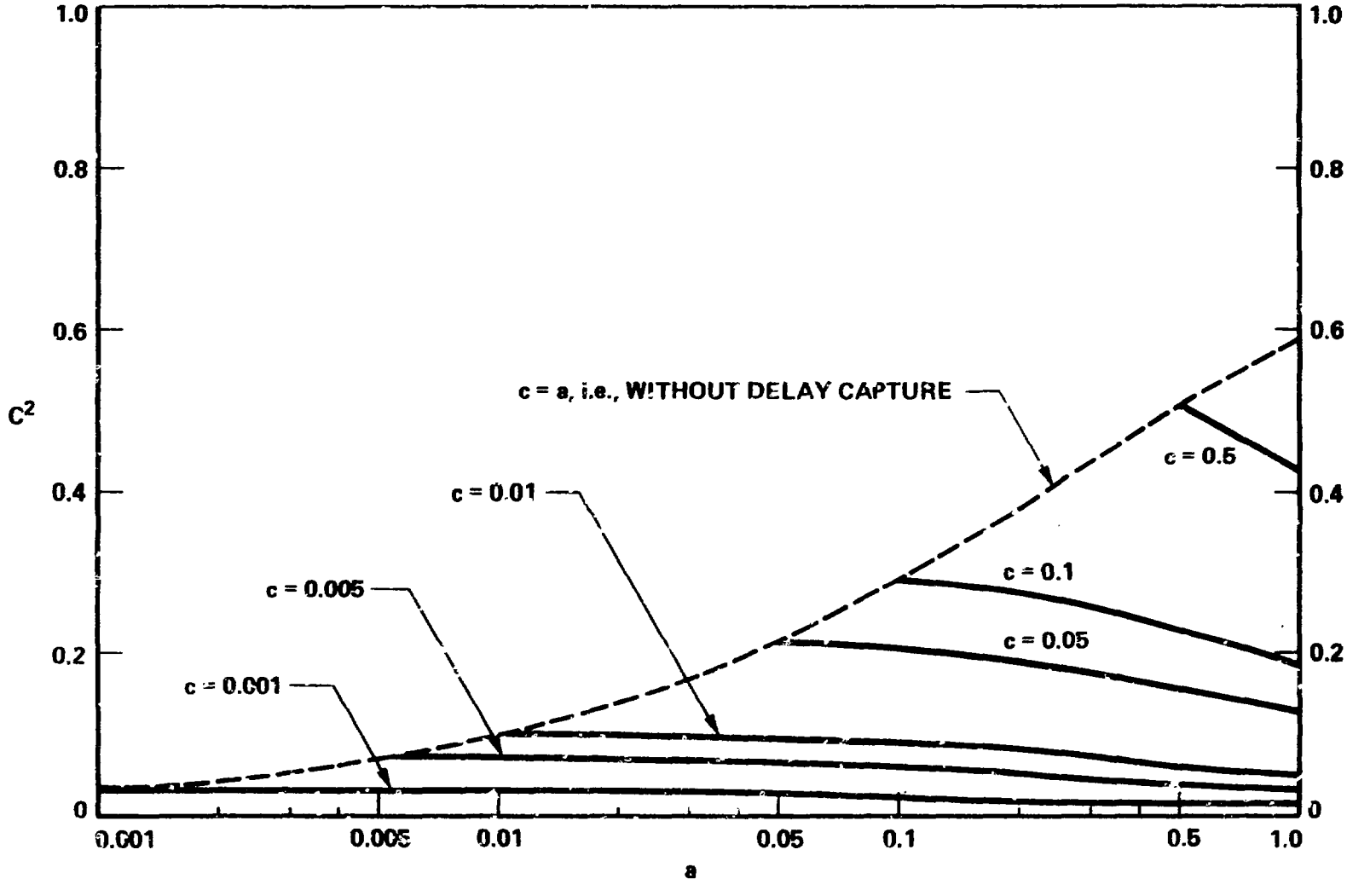


Figure 3.6(b) The coefficient of variation (C^2) for packet interdeparture time in unslotted CSMA with delay capture

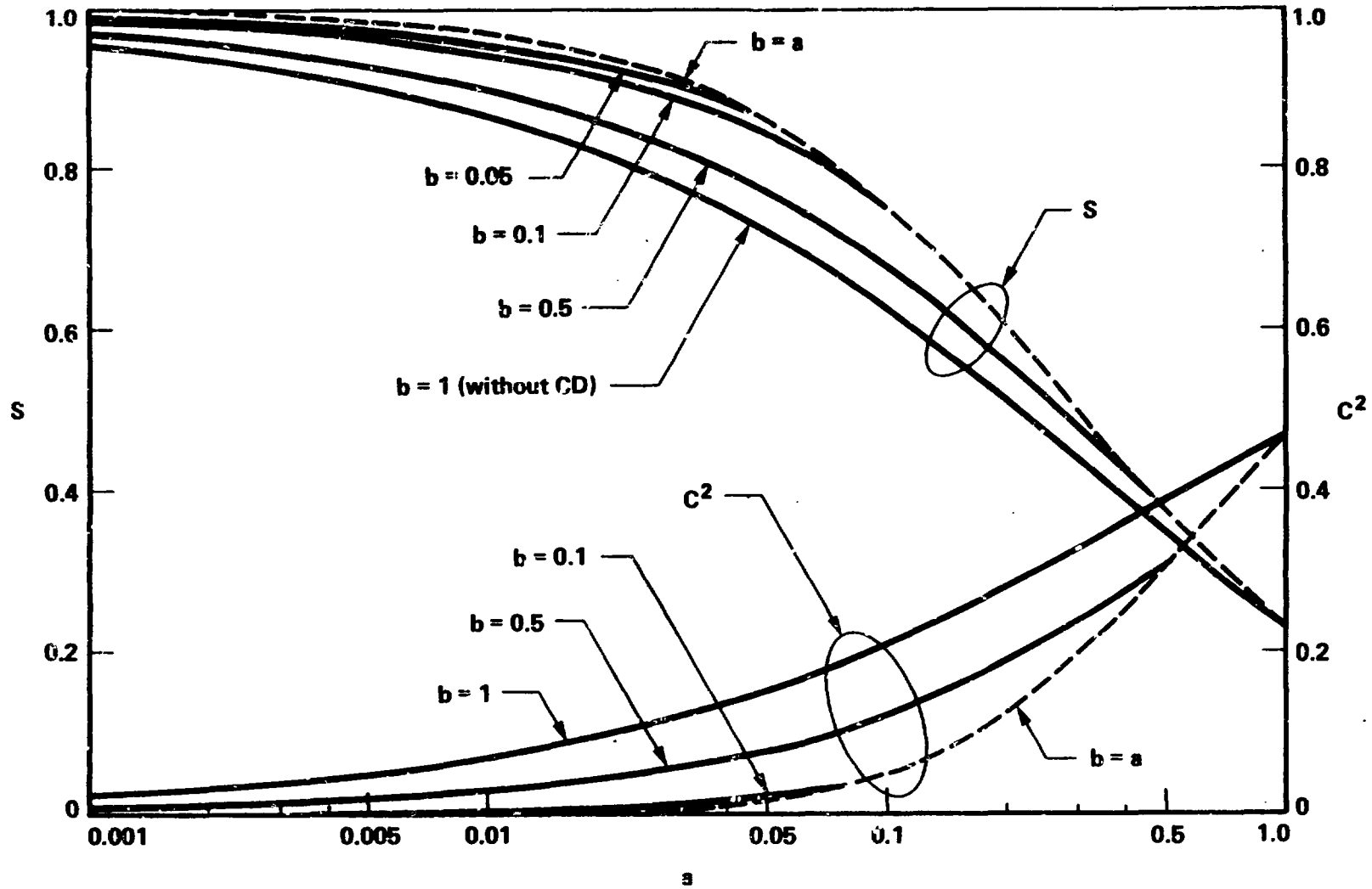


Figure 3.7 The channel throughput (S) and coefficient of variation (C^2) for packet interdeparture time in slotted CSMA/CD

3.4 Output Processes for Nonidentical Users

In this section, we discuss the packet interdeparture time X for a population of users with nonidentical transmission parameter values. First we consider slotted ALOHA and slotted CSMA/CD for which the property of identical distribution of $\{I^{(k)} + F^{(k)};\ k = 1,2,\cdots\}$ still holds because the $F^{(k)}$'s are constant. Then we discuss unslotted CSMA protocols in which the duration of an unsuccessful transmission period depends on the user who initiated the transmission.

3.4.1 Slotted ALOHA

Let p_i be the probability that user i transmits in any slot ($i = 1,2,\cdots,M$). Since $T = F = 1$, we have the Eqs.(3.13), (3.15) and (3.17) as before. However, instead of Eq.(3.14), E and U are now defined by

$$E = \prod_{i=1}^{M} (1-p_i) \quad ; \qquad U = \sum_{i=1}^{M} p_i \prod_{\substack{j=1 \\ (j \neq i)}}^{M} (1-p_j) \tag{3.52}$$

Let us now focus attention on user i. The probability q_i that a successful transmission is achieved by user i is clearly given by

$$q_i = \frac{u_i}{U} \quad ; \qquad U = \sum_{i=1}^{M} u_i \tag{3.53}$$

where

$$u_i = p_i \gamma_i \quad ; \qquad \gamma_i = \prod_{\substack{j=1 \\ (j \neq i)}}^{M} (1-p_j) \tag{3.54}$$

We consider a sequence of interdeparture times $\{X_i^{(n)};\ n = 1,2,\cdots\}$ from user i ($X_i^{(1)}$ begins at the end of an arbitrarily chosen successful transmission by user i). Note that all $X_i^{(n)}$'s are independent and identically distributed (let their generic representation be X_i). Clearly, X_i can be expressed as

$$X_i = \sum_{k=1}^{K_i} [I_i^{(k)} + 1] \tag{3.55}$$

where $I_i^{(k)}$ is the kth idle period duration at user i, and K_i is the number of transmissions by user i until he achieves success. These two are independent and are distributed, respectively, as follows (let I_i be the generic representation of the identically distributed $I_i^{(k)}$'s):

$$Prob[I_i = n] = (1-p_i)^n p_i \qquad n = 0,1,2,\cdots$$

$$\bar{I}_i = \frac{1-p_i}{p_i} \quad ; \qquad Var[I_i] = \frac{1-p_i}{p_i^2} \tag{3.56}$$

$$Prob\,[K_i = k] = (1-\gamma_i)^{k-1}\gamma_i \qquad k = 1,2,\cdots$$

$$\bar{K}_i = \frac{1}{\gamma_i} \;; \quad Var[K_i] = \frac{1-\gamma_i}{\gamma_i^2} \tag{3.57}$$

From Eqs.(3.55)-(3.57), we get

$$S_i \triangleq \frac{1}{\bar{X}_i} = u_i \;; \quad C_i^2 \triangleq \frac{Var[X_i]}{\bar{X}_i^2} = 1-u_i \tag{3.58}$$

Using Eqs.(3.17), (3.53), (3.54) and (3.58), we can confirm Eq.(3.6). The distribution of X_i is given by

$$Prob\,[X_i = n] = (1-u_i)^{n-1}u_i \qquad n = 1,2,\cdots \tag{3.59}$$

3.4.2 Slotted CSMA and CSMA with Collision Detection

The distribution of packet interdeparture time X for slotted CSMA/CD users with parameters $\{p_i\}$ is still given by Eq.(3.27), where E and U are now defined by Eq.(3.52). Accordingly, the channel throughput S and the coefficient of variation of X, C^2, are given by Eq.(3.26). For the individual users, the throughput S_i and the coefficient of variation of interdeparture times C_i^2 can be calculated by Eqs.(3.6), (3.53) and (3.54). The results for CSMA without collision detection may be obtained by letting $b = 1$.

Generally it can be shown that the maximum-allowable throughput contour in the $\{p_i\}$ space is given by

$$a + b(1-E) = (a+b)\sum_{i=1}^{M} p_i \tag{3.60}$$

(This of course reduces to Eq.(3.28) in the case of identical users.)

In Figure 3.8, we plot $\{S_i\}$ and $\{C_i^2\}$ as well as S and C^2 for $M = 5$ users of slotted CSMA (without CD) when $p_i = ip$ $(i = 1,2,\cdots,M)$ with optimal p. We see that the individual throughputs $\{S_i\}$ are distributed nearly in the same proportion to $\{p_i\}$. Also we notice in the region of high throughput (small a) that $C_i^2 \simeq 1 - S_i$. This implies the nearly independent geometric distribution (like in ALOHA case) for the interdeparture time from each user.

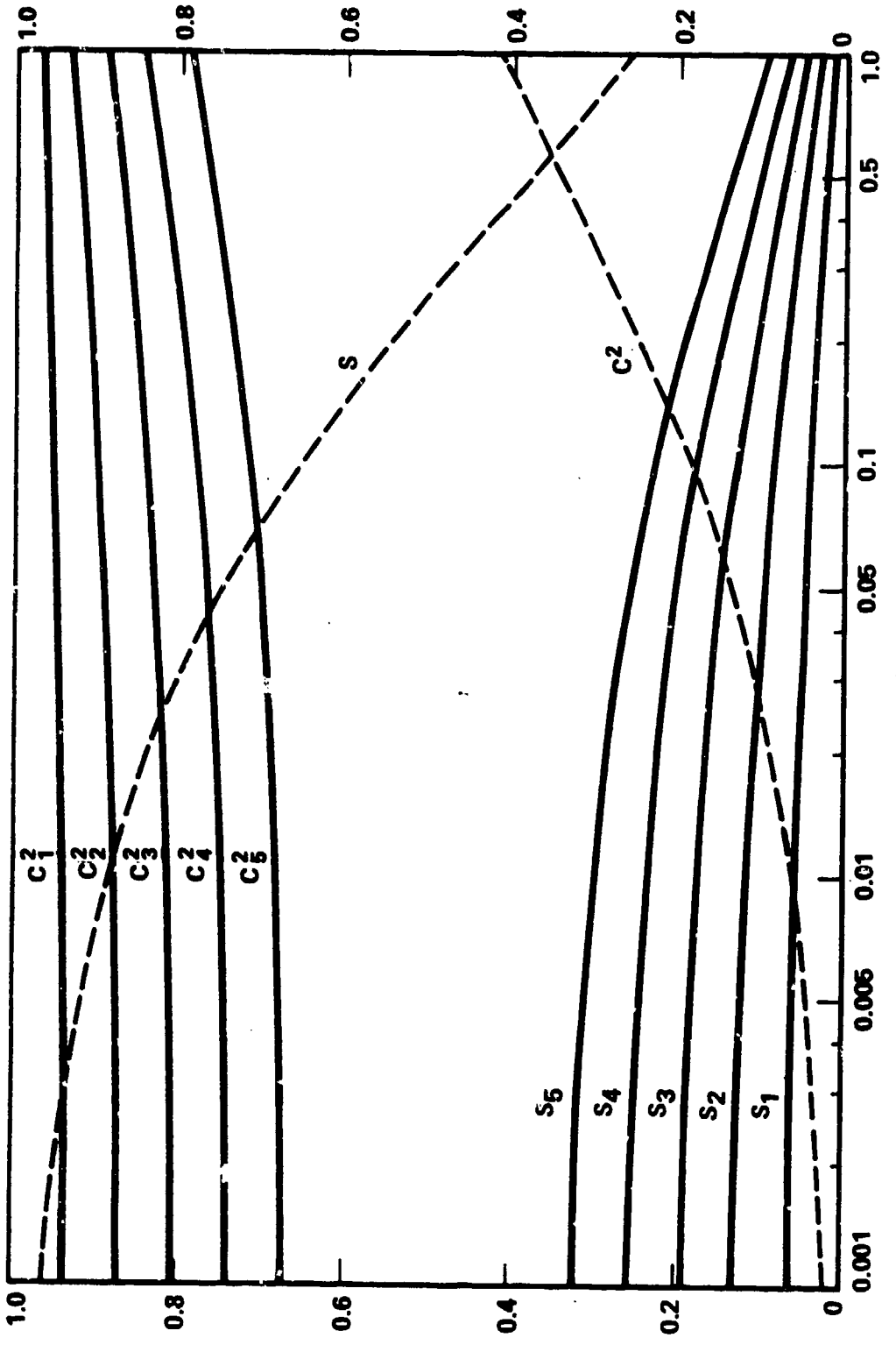


Figure 3.8 $\{S_i\}$ and $\{C_i^2\}$ for 5 nonidentical users of slotted CSMA

3.4.3 Unslotted CSMA

In unslotted systems, the duration of each unsuccessful transmission period depends on which user has begun transmission breaking the channel idle period. We first consider unslotted CSMA without collision detection and without delay capture where the parameter g_i is assigned to user i $(i=1,2,\cdots,M)$. Notice that the channel idle period durations $\{I^{(k)};\ k=1,2,\cdots\}$ are still independent and identically distributed as its generic representation I:

$$Prob[I \leqslant y] = 1-\exp[-y\sum_{i=1}^{M} g_i] \qquad y \geqslant 0$$

$$\bar{I} = \frac{1}{\sum_{i=1}^{M} g_i}\ ; \qquad Var[I] = \frac{1}{(\sum_{i=1}^{M} g_i)^2} \tag{3.61}$$

The terms in a sequence of $\{I^{(k)} + F^{(k)};\ k=1,2,\cdots\}$ in Eq.(3.8) are, however, no longer identically distributed although they are independent. Also, the probability of success for an already started transmission differs from cycle to cycle depending on which user initiates the transmission period. Let us look at these points more closely.

First, notice that the probability that user i among others begins transmission breaking the channel idle period is given by

$$v_i = \frac{g_i}{\sum_{j=1}^{M} g_j}\ ; \qquad \sum_{i=1}^{M} v_i = 1 \tag{3.62}$$

The probability of success in a cycle where user i initiates transmission period is then given by

$$\gamma_i = \exp\left[-a\sum_{\substack{j=1\\(j\neq i)}}^{M} g_j\right] \qquad i=1,2,\cdots,M \tag{3.63}$$

For later use, let us denote by

$$\gamma = \sum_{i=1}^{M} v_i\gamma_i = \frac{\sum_{i=1}^{M} g_i \exp\left[-a\sum_{\substack{j=1\\(j\neq i)}}^{M} g_j\right]}{\sum_{i=1}^{M} g_i} \tag{3.64}$$

the probability that a transmission period is successful.

In an unsuccessful transmission period initiated by user i, let $Y(i)$ be the transmission start time of the last colliding packet. Its distribution is given by

$$Prob[\,Y(i) \leqslant y\,] = \frac{\prod_{\substack{j=1 \\ (j \neq i)}}^{M} (1 - e^{-R_j y} + e^{-R_j a}) - \gamma_i}{1 - \gamma_i} \qquad 0 \leqslant y \leqslant a \tag{3.65}$$

Now, Eq.(3.8) can be written as

$$X = \sum_{n=1}^{\infty} \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_n=1}^{M} \left[\prod_{k=1}^{n-1} \nu_{i_k} (1 - \gamma_{i_k}) \right] \nu_{i_n} \gamma_{i_n} \cdot \{ \sum_{k=1}^{n-1} [\, I^{(k)} + 1 + a + Y(i_k)\,] + I^{(n)} + 1 + a \} \tag{3.66}$$

where $\nu_{i_k}(1 - \gamma_{i_k})$ is the probability that the kth transmission period is initiated by user i_k and that it involves collision. Eq.(3.66) for X can be simplified by using Eqs.(3.62) and (3.64) as follows:

$$X = \sum_{n=1}^{\infty} \{ (1-\gamma)^{n-2} \gamma \sum_{k=1}^{n-1} \sum_{i_k=1}^{M} [\, I^{(k)} + 1 + a + Y(i_k)\,] \nu_{i_k} (1 - \gamma_{i_k}) + (1-\gamma)^{n-1} \gamma [\, I^{(n)} + 1 + a\,] \} \tag{3.67}$$

Let us define a random variable Y by

$$Prob[\,Y \leqslant y\,] = \frac{1}{1-\gamma} \sum_{i=1}^{M} \nu_i (1 - \gamma_i)\, Prob[\,Y(i) \leqslant y\,] \qquad 0 \leqslant y \leqslant a \tag{3.68}$$

Then we have

$$X = \sum_{n=1}^{\infty} (1-\gamma)^{n-1} \gamma \{ \sum_{k=1}^{n-1} [\, I^{(k)} + 1 + a + Y\,] + I^{(n)} + 1 + a \} \tag{3.69}$$

Note that Eq.(3.69) is of the same form as Eq.(3.8) conditionally summed. Therefore, Eq.(3.12) for $X^*(s)$ still holds when T and F are given by Eqs.(3.33) and (3.34), respectively, and Y in Eq.(3.68) is used. Thus we get

$$\bar{X} = \frac{1}{\gamma} (\bar{I} + 1 + a) + (\frac{1}{\gamma} - 1)\, \bar{Y}\,,$$

$$Var[X] = \frac{1}{\gamma} Var[I] + (\frac{1}{\gamma} - 1)\, Var[Y] + (\bar{I} + 1 + a + \bar{Y})^2 \frac{1-\gamma}{\gamma^2} \tag{3.70}$$

where

$$\bar{Y} = \sum_{i=1}^{M} \overline{Y(i)}\, \nu_i (1 - \gamma_i) / (1 - \gamma) = \frac{1}{1-\gamma} \int_0^a [\, 1 - \sum_{i=1}^{M} \nu_i \prod_{\substack{j=1 \\ (j \neq i)}}^{M} (1 - e^{-R_j y} + e^{-R_j a})\,]\, dy\,,$$

$$Var[Y] = \sum_{i=1}^{M} \overline{Y(i)^2}\, \nu_i (1 - \gamma_i) / (1 - \gamma) - \bar{Y}^2 \tag{3.71}$$

Especially we have explicitly

$$S = \frac{\gamma}{\frac{1}{\sum_{i=1}^{M} g_i} + 1 + 2a - \sum_{i=1}^{M} \nu_i \int_0^a \left[\prod_{\substack{j=1 \\ (j \neq i)}}^{M} (1 - e^{-g_j y} + e^{-g_j a}) \right] dy} \tag{3.72}$$

Note that q_i, the probability that a successful transmission is achieved by user i, is now given by

$$q_i = \frac{g_i \gamma_i}{\sum_{j=1}^{M} g_j \gamma_j} = \frac{\nu_i \gamma_i}{\gamma} \qquad i = 1,2,\cdots,M \tag{3.73}$$

Thus the throughput of user i is given by

$$S_i = q_i S = \frac{\nu_i \gamma_i}{\frac{1}{\sum_{j=1}^{M} g_j} + 1 + 2a - \sum_{j=1}^{M} \nu_j \int_0^a \left[\prod_{\substack{k=1 \\ (k \neq j)}}^{M} (1 - e^{-g_k y} + e^{-g_k a}) \right] dy} \qquad i = 1,2,\cdots,M \tag{3.74}$$

The analysis for CSMA with delay capture can be done quite similarly. Specifically, we get the system and individual throughput expressions in Eqs.(3.72) and (3.74), respectively, where the following γ_i is used instead of Eq.(3.63):

$$\gamma_i = \exp\left[-c \sum_{\substack{j=1 \\ (j \neq i)}}^{M} g_j \right] \qquad i = 1,2,\cdots,M \tag{3.75}$$

3.4.4 Unslotted CSMA with Collision Detection

The above analysis can be readily applied to unslotted CSMA with collision detection time b. Instead of Eq.(3.69), we have

$$X = \sum_{n=1}^{\infty} (1-\gamma)^{n-1} \gamma \left[\sum_{k=1}^{n-1} [I^{(k)} + b + a + Y_1] + I^{(n)} + 1 + a \right] \tag{3.76}$$

where Y_1 is a random variable defined by

$$Prob[Y_1 \leq y] = \frac{1}{1-\gamma} \sum_{i=1}^{M} \nu_i (1 - \gamma_i) \, Prob[Y_1(i) \leq y] \qquad 0 \leq y \leq a \tag{3.77}$$

and $Y_1(i)$ is the transmission start time of the first colliding packet in an unsuccessful transmission period initiated by user i. In Eqs.(3.76) and (3.77), $I^{(k)}$ is the kth channel idle period duration, and ν_i, γ and γ_i are given in Eqs.(3.62)-(3.64). The distribution of $Y_1(i)$ is given by

$$Prob[\,Y_1(i) > y\,] = \frac{\exp\left[-y \sum_{\substack{j=1 \\ (j\neq i)}}^{M} g_j\right] - \gamma_i}{1-\gamma_i} \qquad 0 \leqslant y \leqslant a \tag{3.78}$$

from which we have

$$\bar{Y}_1 = \frac{1}{1-\gamma} \sum_{i=1}^{M} \nu_i (1-\gamma_i) \Big(\sum_{\substack{j=1 \\ (j\neq i)}}^{M} g_j \Big)^{-1} - \frac{a\gamma}{1-\gamma} ,$$

$$Var[Y_1] = \frac{1}{1-\gamma} \sum_{i=1}^{M} \nu_i (1-\gamma_i)\, \overline{Y_1(i)^2} - \bar{Y}_1^2 \tag{3.79}$$

where

$$\overline{Y_1(i)^2} = \frac{\hat{z}}{\Big(\sum_{\substack{j=1 \\ (j\neq i)}}^{M} g_j \Big)^2} - \frac{a^2\gamma_i}{1-\gamma_i} - \frac{2a\gamma_i}{(1-\gamma_i)\Big(\sum_{\substack{j=1 \\ (j\neq i)}}^{M} g_j \Big)}$$

Therefore we can compute

$$\bar{X} = \Big(\frac{1}{\gamma} - 1\Big)(\bar{I} + b + a + \bar{Y}_1) + \bar{I} + 1 + a ,$$

$$Var[X] = \frac{1}{\gamma} Var[I] + \Big(\frac{1}{\gamma} - 1\Big) Var[Y_1] + (\bar{I} + b + a + \bar{Y}_1)^2 \frac{1-\gamma}{\gamma^2} \tag{3.80}$$

We have the channel throughput

$$S = \frac{\gamma}{\dfrac{1}{\sum_{i=1}^{M} g_i} + \gamma + (1-\gamma)(b+a) + \sum_{i=1}^{M} \dfrac{\nu_i (1-\gamma_i)}{\sum_{\substack{j=1 \\ (j\neq i)}}^{M} g_j}} \tag{3.81}$$

and the throughput for user i

$$S_i = q_i S = \frac{\nu_i \gamma_i}{\dfrac{1}{\sum_{i=1}^{M} g_i} + \gamma + (1-\gamma)(b+a) + \sum_{i=1}^{M} \dfrac{\nu_i (1-\gamma_i)}{\sum_{\substack{k=1 \\ (k\neq i)}}^{M} g_k}} \qquad i = 1,2,\cdots,M \tag{3.82}$$

3.5 Conclusion

We have studied the packet departure processes in a variety of contention-type packet broadcasting systems. The channel access protocols considered include both slotted and unslotted systems of ALOHA and carrier-sense-multiple-access (CSMA). The effects of collision detection and delay capture on CSMA have also been investigated.

Through the analysis of channel activity cycles alternating between the idle and (successfully or unsuccessfully) transmitting states, we have derived the distribution of the packet interdeparture time X. Then we found the channel throughput ($S = 1/\bar{X}$) and the coefficient of variation of X ($C^2 = Var[X]/\bar{X}^2$) explicitly. All the results for the distribution of X and C^2 are new. Some results for S (specifically Eqs.(3.26), (3.36), (3.41), (3.48), (3.72), and (3.81)) are also newly derived in this chapter. It has been shown that in efficient CSMA systems with collision detection or with delay capture, C^2 is very small while the throughput suffers some degradation. The case where users have different transmission parameter values has also been analyzed.

Using $\bar{X}$ and C^2 together with the elementary renewal theorem, we have also obtained the asymptotic behavior of the number of successful transmissions at interfering individual queues. These results can be used to determine the coefficients in the diffusion process approximation to the queue length distribution at users, which will appear in Chapter 5.

CHAPTER 4
Approximate Output Processes in Hidden-User Packet Radio Systems

The processes consisting of the packet interdeparture times for contention-type packet broadcasting systems in a hidden-user, single-hop environment are studied. The channel access protocols considered include pure ALOHA and slotted and unslotted carrier-sense-multiple-access (CSMA). The theory of superposition of independent renewal processes is applied to approximate the distribution of the duration of each unsuccessful transmission period in channel state. Our analysis results for the channel throughput and the coefficient of variation for the packet interdeparture time in symmetric and 'wall' configurations are shown to be in good agreement with simulation results over a wide range of offered channel traffic.

4.1 Introduction and Assumptions

In Chapter 3, we conducted an exact stochastic analysis of packet interdeparture times (i.e., intervals between two consecutive successful transmissions) for several channel access protocols in packet broadcasting systems. Channel access protocols such as slotted ALOHA [Robe72], pure ALOHA [Abra70], and slotted and unslotted nonpersistent carrier-sense-multiple-access (CSMA) [Toba74,Klei75b] were studied to find explicitly the distributions of packet interdeparture times. (The reciprocal of the mean interdeparture time is the channel throughput, one of our main performance measures.) In our earlier treatment of CSMA, it was assumed that every user is in line-of-sight of all others so that any transmission can be heard (after a finite signal propagation delay) by all parties (i.e., a fully-connected configuration). Regulation of transmission by listening to other transmissions is the essence of CSMA, and it is that which achieves a high throughput (as long as the propagation delay is small compared to the packet transmission time).

However, in applying CSMA to ground-based packet radio communication systems for a population of geographically distributed users, such as PRNET [Kahn78], there are many situations in which some users cannot hear transmissions from certain other users; this is possibly because they are out of transmission range of each other or because they are separated by some physical obstacles (e.g., mountains) blocking the signal. Such a situation, called the *hidden-terminal problem*, was analyzed by Tobagi and Kleinrock [Toba74,Toba75] and a serious throughput degradation was shown to exist. This is because hidden users behave independently ignoring the ongoing transmissions. (The busy-tone multiple-access (BTMA) was then proposed to save the day.)

This chapter focuses on the performance analysis of hidden-user configurations by use of an approach different from [Toba75]. Our method is based on the modeling of packet transmission activity at each user as a two-state (transmitting or not) alternating renewal process. In our models, we assume that each user has packets ready for transmission at all times. Also, the transmission protocol is assumed to be memoryless in the sense that whenever a user experiences an idle (non-transmitting) period he renews his action regardless of the past happenings. Now, let us define the two alternating states in channel. The transmission state in channel is the state where at least one user is transmitting or any transmission is being sensed. Also, the channel idle state is defined as the state where no users are transmitting or no transmissions are being sensed. Thus, the channel state also alternates between the transmission and idle periods. (There can be two consecutive transmission periods with an idle period of duration 0 between them.)

A transmission period is successful if there are no other transmissions heard at the intended receiver during the (protocol-dependent) vulnerable period. Exact analysis is possible for the stochastic property of the durations of the channel idle period and a successful transmission period. However, to analyze the duration of an unsuccessful transmission period, an approximation using the theory of superposition of independent renewal processes is applied. This treatment involves a twofold approximation; (i) we treat each user's transmission process as independent with a properly reduced transmission rate (whereas, in fact, two CSMA users in line-of-sight of each other behave dependently), (ii) we treat consecutive interevent times in a superposed process as if they were independent and identically distributed (whereas in reality they are not). The validation of our approximation will be provided by comparing the results by existing exact analysis (for certain special cases) and simulation.

We assume the existence of a single receiving station which is in line-of-sight of all users. We then are in a position to realize a spectrum of 'hiddenness' ranging from ALOHA (completely hidden) to fully-connected CSMA (completely visible) and the partially hidden configuration of CSMA in between. Our approach makes it possible for the solution to smoothly transfer over all degrees of hiddenness as opposed to the one in [Toba75] where, for example, the expression for the channel throughput in the limit of fully-connected CSMA (in a zero propagation delay, infinite population model) does not agree with the exact result in [Klei75b]. Also, through (approximate) analysis of the durations of alternating channel states (idle and transmitting), we obtain an approximation to the mean packet interdeparture time (whose reciprocal is the channel throughput) as well as its variance. These first two moments of the distribution will be used to determine the coefficients in the diffusion process approximation to users' packet queue length distribution in Chapter 5.

In the following, after an extract from the theory of superposed renewal processes (Section 4.2), we consider in Section 4.3 the packet departure processes of pure ALOHA, unslotted CSMA, unslotted CSMA with delay capture effect (explained shortly), and slotted CSMA, each in a hidden-user environment. Our study of CSMA is restricted to the nonpersistent CSMA protocol [Klei75b] only. Note that slotted ALOHA systems do not need special treatment for hidden users because all users of slotted ALOHA are originally considered to be hidden from each other and (due to the assumed memoryless protocol) the event in any slot is always

independent of the event in any other slot. Comparison of our calculated results with the simulation results in some example systems is discussed in Section 4.4. We also discuss the rationale of our assumptions and give directions to possible refinement of the present formulation.

Throughout this chapter, the packet length is assumed to be constant, and its transmission time is chosen as the unit of time axis. Then, in typical ground-based systems, the signal propagation delay, denoted by a, is small (e.g., $a = 0.01$ normalized time units). In the analysis of pure ALOHA (Section 4.3.1) and unslotted CSMA (Section 4.3.2), we may properly assume that $a = 0$ because performance degradation due to the finiteness of a is considered to be negligible compared to the degradation due to the hidden-terminal effect. We also consider the case $a > 0$ in Section 4.3.4.

Systems which use the spread spectrum modulation technique may exhibit a delay capture phenomenon [Kahn78]. This is the ability of a receiver to successfully receive a leading packet (i.e., a packet which started transmission by breaking the channel idle period) even though most of it arrives at the receiver overlapped in time by other packets. The time needed to capture the leading packet, denoted by c, can be also small (e.g., $c \cong a/20 = 0.0005$ [Davi80]). Thus, in our treatment of CSMA with delay capture (Section 4.3.3), we assume that $c = 0$ ('perfect' capture). Note that even with perfect capture, packets transmitted in the middle of ongoing transmissions are not received correctly and as such cause a degradation in throughput by extending the duration of transmission period.

Analysis of slotted CSMA is conducted by assuming a finite value of the propagation delay a which is chosen as the (mini)slot size. For a successful transmission, clearance of $1 + 1/a$ slots in the channel is required ($1/a$ is assumed to be an integer). Thus, if every user is hidden from one another, the system may be called 'minislotted ALOHA.' We apply the slotted version of the superposition of renewal processes (developed in Section 4.2.2) to deal with the duration of an unsuccessful transmission period in slotted CSMA.

4.2 Superposition of Independent Renewal Processes

In this section, we derive the distribution of interevent times in a superposition of independent renewal processes. We assume that there are a finite number, M, of event sources (indexed 1 through M), at each of which events occur from time to time independently of the others. We first consider the case of continuous-time systems following the existing theory (in Section 4.2.1), and then adapt it into the case of discrete-time systems (in Section 4.2.2).

4.2.1 Continuous-Time Systems

Let the interevent times at source i be independent and identically distributed as represented by Y_i with mean $\bar{Y}_i$ and distribution function $F_i(x)$, $x \geq 0$ $(i = 1,2,\cdots,M)$. Figure 4.1 illustrates a combination of these events into a superposition process. Note that in the superposition process, the interevent times are generally neither independent nor identically distributed. (The correlation among such successive intervals was studied by Lawrance [Lawr73] and Ito [Ito78].) However, what we are seeking here is the steady-state distribution of a single interevent time, denoted by $\tilde{Y}$, following an arbitrarily chosen event.

Conditioned on picking an event from source i, the following interevent time $\tilde{Y}(i)$ can be expressed as

$$\tilde{Y}(i) = \min[\, Y_i, \hat{Y}_1, \cdots, \hat{Y}_{i-1}, \hat{Y}_{i+1}, \cdots, \hat{Y}_M \,] \tag{4.1}$$

where $\hat{Y}_j$ stands for the residual life in source j whose pdf is given by $[1 - F_j(x)]/\bar{Y}_j$ (see, e.g., [Klei75a]). From Eq.(4.1), we have, for $i = 1,2,\cdots,M$,

$$Prob[\, \tilde{Y}(i) > x \,] = Prob[\, Y_i > x, \hat{Y}_j > x, \text{ for all } j \neq i \,]$$

$$= [1 - F_i(x)] \prod_{\substack{j=1 \\ (j \neq i)}}^{M} \frac{1}{\bar{Y}_j} \int_x^\infty [1 - F_j(y)]\, dy \qquad x \geq 0 \tag{4.2}$$

We can uncondition Eq.(4.2) by use of

$$Prob[\text{ picking at random an event from source } i\,] = \frac{E[\tilde{Y}]}{\bar{Y}_i} \qquad i = 1,2,\cdots,M \tag{4.3}$$

where

$$\frac{1}{E[\tilde{Y}]} = \sum_{i=1}^{M} \frac{1}{\bar{Y}_i} \tag{4.4}$$

Eqs.(4.3) and (4.4) come from the observation that the rate of events from source i is given by $1/\bar{Y}_i$. This unconditioning yields

$$Prob[\, \tilde{Y} > x \,] = E[\tilde{Y}] \sum_{i=1}^{M} \frac{1 - F_i(x)}{\bar{Y}_i} \prod_{\substack{j=1 \\ (j \neq i)}}^{M} \frac{1}{\bar{Y}_j} \int_x^\infty [1 - F_j(y)]\, dy \qquad x \geq 0 \tag{4.5}$$

This expression is given in [Lawr73]. When the sources are identical, in which case we drop the subscripts i and j, the result (4.5) reduces to the form given by Cox and Smith [Cox54]:

$$Prob[\, \tilde{Y} > x \,] = [1 - F(x)] \left\{ \frac{1}{\bar{Y}} \int_x^\infty [1 - F(y)]\, dy \right\}^{M-1} \qquad x \geq 0 \tag{4.6}$$

(Note that $\bar{Y} = \bar{Y}_i = E[\tilde{Y}]/M$.)

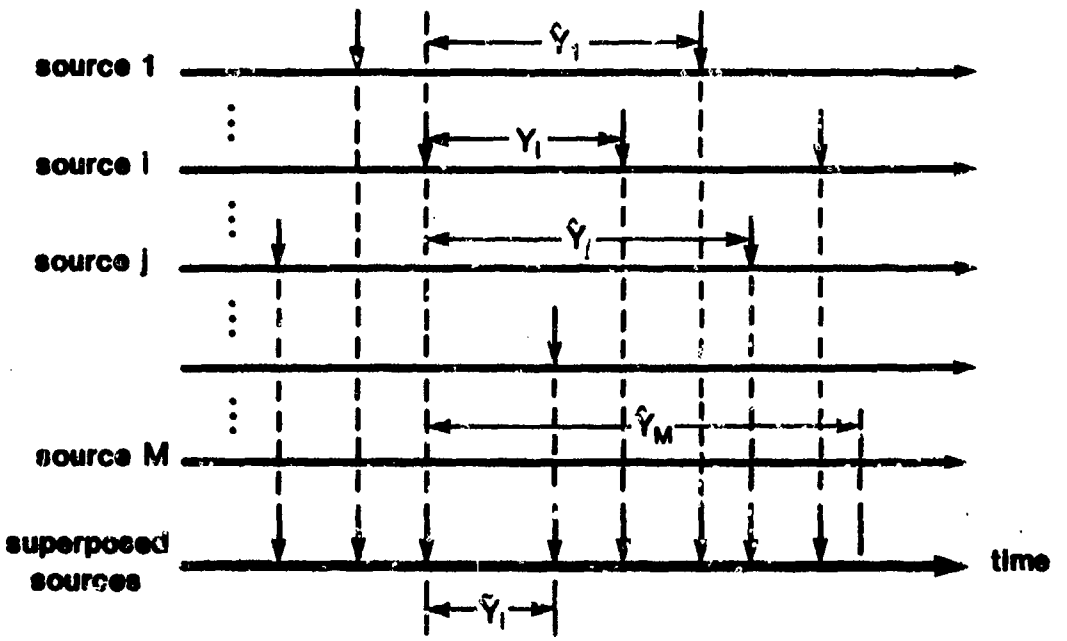


Figure 4.1 Interevent time in the superposition process.

Let us consider two examples. For the first example, we let the interevent times at source i be exponentially distributed: $F_i(x) = 1-\exp(-x/\bar{Y}_i)$, $x \geq 0$ $(i=1,2,\cdots,M)$. Then, it is easily shown that Eq.(4.5) yields

$$Prob\,[\,\tilde{Y} > x\,] = \exp(-x/E[\tilde{Y}]), \quad x \geq 0 \qquad \text{(exponential)}$$

as expected from the fact that merging of independent Poisson streams forms a Poisson stream with aggregate rate given by Eq.(4.4).

As a second example, assume that

$$F_i(x) = \begin{cases} 0 & 0 \leq x < 1 \\ 1 - e^{-g_i(x-1)} & x \geq 1 \end{cases}$$

$$\bar{Y}_i = 1 + \frac{1}{g_i} \qquad i = 1,2,\cdots,M \tag{4.7}$$

which is an exponential distribution shifted by 1. Then, since

$$\int_x^\infty [1-F_i(y)]\,dy = \begin{cases} 1 - x + \dfrac{1}{g_i} & 0 \leq x < 1 \\ \dfrac{1}{g_i}\, e^{-g_i(x-1)} & x \geq 1 \end{cases}$$

Eq.(4.5) gives

$$Prob\,[\,\tilde{Y} > x\,] = \begin{cases} E[\tilde{Y}] \displaystyle\sum_{i=1}^{M} \frac{g_i}{1+g_i} \prod_{\substack{j=1 \\ (j\neq i)}}^{M} \frac{1+g_i(1-x)}{1+g_i} & 0 \leq x < 1 \\ E[\tilde{Y}] \left(\displaystyle\sum_{i=1}^{M} g_i \right) \left(\displaystyle\prod_{i=1}^{M} \frac{e^{-g_i(x-1)}}{1+g_i} \right) & x \geq 1 \end{cases} \tag{4.8}$$

This second example is important since later we will be interested in the distribution of the random variable $\tilde{Y}$ when $\tilde{Y} \leq 1$ (we denote this random variable by f).

$$Prob\,[\,f \leq x\,] \triangleq Prob\,[\,\tilde{Y} \leq x \mid \tilde{Y} \leq 1\,] = \frac{1 - Prob\,[\,\tilde{Y} > x\,]}{1 - Prob\,[\,\tilde{Y} > 1\,]}$$

or

$$Prob\,[\,f > x\,] = \frac{\displaystyle\sum_{i=1}^{M} g_i \left\{ \prod_{\substack{j=1 \\ (j\neq i)}}^{M} [\,1+g_i(1-x)\,] - 1 \right\}}{\displaystyle\sum_{i=1}^{M} g_i \left\{ \prod_{\substack{j=1 \\ (j\neq i)}}^{M} (1+g_i) - 1 \right\}} \qquad 0 \leq x \leq 1 \tag{4.9}$$

In the case of identical sources (again dropping the subscripts i and j), Eq.(4.9) reduces to

$$Prob\,[f > x] = \frac{[1+g(1-x)]^{M-1} - 1}{(1+g)^{M-1} - 1} \qquad 0 \leqslant x \leqslant 1 \tag{4.10}$$

from which we can calculate

$$\bar{f} = \frac{1}{(1+g)^{M-1} - 1}\left[\frac{(1+g)^M - 1}{gM} - 1\right],$$

$$Var[f] = \frac{1}{[(1+g)^{M-1} - 1]^2}$$

$$\cdot\left\{\frac{2[(1+g)^{M+1} - 1][(1+g)^{M-1} - 1]}{g^2M(M+1)} - \frac{[(1+g)^M - 1]^2}{(gM)^2} - (1+g)^{M-1} + \frac{2(1+g)^{M-1}}{M}\right\} \tag{4.11}$$

It is interesting to note that the limiting forms of the expressions in Eqs.(4.10) and (4.11) for $M \to \infty$ with G fixed at $G = gM$ are given by

$$Prob\,[f \leqslant x] = \frac{1 - e^{-Gx}}{1 - e^{-G}} \qquad 0 \leqslant x \leqslant 1$$

$$\bar{f} = \frac{1}{G} - \frac{e^{-G}}{1 - e^{-G}}\ ; \qquad Var[f] = \frac{1}{G^2} - \frac{e^{-G}}{(1 - e^{-G})^2} \tag{4.12}$$

These results are identical to Eq.(B.2) in Appendix B obtained by considering a collective Poisson stream with rate G. We also note from (4.11) that $\bar{f} \simeq M/[2(M-1)]$ as $g \to 0$.

4.2.2 Discrete-Time Systems

With the application to slotted CSMA in mind, let us derive the probability distribution of an arbitrarily chosen interevent time for the superposition of M independent event sources in the discrete-time systems. We denote the slot size by a, and call each slot boundary (when events are allowed to occur) as an *epoch*. Let the interevent times at source i ($i = 1,2,\cdots,M$) be independent and identically distributed as represented by Y_i with mean $\bar{Y}_i$ and the complementary cumulative distribution

$$F_i^c(n) \triangleq Prob\,[Y_i > na] = \sum_{j=n+1}^{\infty} Prob\,[Y_i = ja] \qquad n = 0,1,2,\cdots \tag{4.13}$$

Then the residual life in source i (on the condition of no events in source i at a selected epoch), denoted by $\hat{Y}_i$, is distributed as

$$Prob\,[\hat{Y}_i > na] = \frac{a}{\bar{Y}_i} \sum_{k=n}^{\infty} F_i^c(k) \qquad n = 0,1,2,\cdots \tag{4.14}$$

Note for a slotted system that more than one event can occur at each epoch. The probability that an event occurs in source i at a randomly chosen epoch is given by $a/\bar{Y}_i$. Therefore, by similar arguments as in Section 4.2.1, the distribution of an arbitrary interevent time in the superposition process, denoted by $\tilde{Y}$, is given by

$$Prob[\tilde{Y} > na] = \frac{\prod_{i=1}^{M}[\frac{a}{\bar{Y}_i}F_i^e(n) + (1-\frac{a}{\bar{Y}_i})\frac{a}{\bar{Y}_i}\sum_{k=n}^{\infty}F_i^e(k)] - \prod_{i=1}^{M}[(1-\frac{a}{\bar{Y}_i})\frac{a}{\bar{Y}_i}\sum_{k=n}^{\infty}F_i^e(k)]}{1-\prod_{i=1}^{M}(1-\frac{a}{\bar{Y}_i})}$$

$$n = 0,1,2,\cdots \tag{4.15}$$

where the denominator stands for the probability that we have at least one event occurring at a randomly chosen epoch.

Let us consider again two examples. In the first example, we assume that, in source i an event occurs at any epoch with probability p_i and does not with probability $1-p_i$ $(i=1,2,\cdots,M)$. Thus the interevent times in source i are geometrically distributed as

$$Prob[Y_i = na] = (1-p_i)^{n-1}p_i \qquad n=1,2,\cdots$$

Using this in Eqs.(4.13) and (4.15), we have

$$Prob[\tilde{Y} = na] = E^{n-1}(1-E) \qquad n=1,2,\cdots$$

where $E \triangleq \prod_{i=1}^{M}(1-p_i)$ is the probability of no events at each epoch. This result is as expected since it represents the probability that an event follows $n-1$ epochs of no events.

For the second example, assume that $\tau \triangleq 1/a$ is an integer and that, for source i

$$Prob[Y_i = 1+na] = (1-p_i)^{n-1}p_i \qquad n=1,2,\cdots$$

$$\bar{Y}_i = 1 + \frac{1}{\tau p_i} \qquad i=1,2,\cdots,M \tag{4.16}$$

Then we have

$$F_i^e(n) = \begin{cases} 1 & 0 \leqslant n \leqslant \tau \\ (1-p_i)^{n-\tau} & n \geqslant \tau \end{cases}$$

Substituting these expressions into Eq.(4.15), we obtain the distribution of $\tilde{Y}$ when $\tilde{Y} \leqslant 1$, which is denoted by f again. It is given by

$$Prob[f > na] \triangleq 1 - \frac{1 - Prob[\tilde{Y} > na]}{1 - Prob[\tilde{Y} > \tau a]}$$

$$= \frac{\prod_{i=1}^{M}[p_i + P_i[1+p_i(\tau-n)]] - \prod_{i=1}^{M}[P_i[1+p_i(\tau-n)]] - \prod_{i=1}^{M}(p_i+P_i) + \prod_{i=1}^{M}P_i}{\prod_{i=1}^{M}(1+\tau p_i) - \prod_{i=1}^{M}[1+p_i(\tau-1)] - \prod_{i=1}^{M}(p_i+P_i) + \prod_{i=1}^{M}P_i}$$

$$n = 0,1,2,\cdots,\tau \tag{4.17}$$

where $P_i \triangleq (1-p_i+\tau p_i)/(1+\tau p_i)$. From Eq.(4.17), we can calculate $\bar{f}$ and $Var[f]$ by use of

$$\bar{f} = \sum_{n=0}^{\tau} n\, Prob[f > na] ,$$

$$Var[f] = \sum_{n=0}^{\tau} (2n)\, Prob[f > na] + \bar{f} - \bar{f}^2 \tag{4.18}$$

Finally, let us show the limit form of Eq.(4.17) as $M \to \infty$ in the case of identical sources (dropping the subscript i) with G kept at a fixed value such that $aG = pM$:

$$Prob[f \leq na] = \frac{1-e^{-naG}}{1-e^{-G}} \qquad n = 0,1,2,\cdots,\tau \tag{4.19}$$

This expression should be as expected since such an infinite population of sources constitutes a group of arrivals at each epoch with Poisson distributed group size with mean aG. We may also note the correspondence between Eqs.(4.12) and (4.19).

4.3 Analysis of Output Processes

As in Chapter 3, we can express the packet interdeparture time X as consisting of $K-1$ cycles of alternating channel idle periods $\{I^{(k)}\}$ and unsuccessful transmission periods $\{F^{(k)}\}$ $(k = 1,2,\cdots,K-1)$ terminated by the last cycle of $I^{(K)}$ and a successful transmission period T:

$$X = \sum_{k=1}^{K-1} [I^{(k)} + F^{(k)}] + I^{(K)} + T \tag{4.20}$$

If our channel access protocol assumes that $\{I^{(k)} + F^{(k)};\ k = 1,2,\cdots,K-1\}$ and $\{I^{(K)} + T\}$ are mutually independent and also each of them is independent of K, then we can express the mean and variance of X as

$$\bar{X} = (\bar{K}-1)(\bar{I}+\bar{F}) + \bar{I} + \bar{T} ,$$

$$Var[X] = \bar{K}\, Var[I] + (\bar{K}-1)\, Var[F] + Var[T] + (\bar{I}+\bar{F})^2 Var[K] \tag{4.21}$$

where I and F represent each of $\{I^{(k)}\}$ and $\{F^{(k)}\}$ identically distributed, respectively, and we have assumed the independence of $I^{(k)}$, $F^{(k)}$ and T for each k. For all the protocols we consider below, K is geometrically distributed as

$$Prob[K=k] = (1-\gamma)^{k-1}\gamma \qquad k=1,2,\cdots$$

$$\overline{K} = \frac{1}{\gamma} \; ; \quad Var[K] = \frac{1-\gamma}{\gamma^2} \tag{4.22}$$

where γ is the probability that a transmission is successful once it has been started by breaking the channel idle period.

Through Eq.(4.21), the throughput S and the coefficient of variation C^2 of packet interdeparture times for the whole system are given by

$$S = \frac{1}{\overline{X}} \; ; \quad C^2 = \frac{Var[X]}{\overline{X}^2} \tag{4.23}$$

The throughput S_i and the coefficient of variation C_i^2 of packet interdeparture times for user i is given by

$$S_i = q_i S \; ; \quad C_i^2 = 1 - q_i(1-C^2) \qquad i=1,2,\cdots,M \tag{4.24}$$

where q_i is the probability that a successful transmission is achieved by user i. (See Eq.(3.6).)

For fully-connected CSMA systems and an infinite population of pure ALOHA users, we have found the exact expressions for the distribution of X as well as for the mean and variance of X in Chapter 3. For systems involving a finite number of hidden users, however, it seems very difficult to find the distribution of F for the reasons mentioned in Section 4.1. Therefore, we introduce several approximations and validate them in some cases where exact analysis or simulation results are available (Section 4.4).

4.3.1 Pure ALOHA

Let us begin our (approximate) analysis of output processes with pure (or unslotted) ALOHA for a finite number M of users. We assume that the propagation delay $a = 0$ in this section. (The case of nonzero a is given as a special case of the system analyzed in Section 4.3.4.) Whether hidden or not, each user of pure ALOHA behaves independently of all others. So, let user i alternate between the transmitting state of duration 1 and the idle state of duration exponentially distributed with mean $1/g_i$ ($i=1,2,\cdots,M$). Thus, if we focus attention on the instants of starting transmission at each user, the intervals between those instants are independent and identically distributed as given by the distribution function in Eq.(4.7). Therefore, the interval between two arbitrarily chosen successive starts of transmissions in the whole system is distributed as given by Eq.(4.8).

Now, let us find γ, the distributions of I, F and T, and $\{q_i\}$ for pure ALOHA. First, obviously the channel idle period is exponentially distributed with aggregate parameter $\sum_{i=1}^{M} g_i$:

$$Prob[I \leq y] = 1 - \exp(-y \sum_{i=1}^{M} g_i) \qquad y \geq 0$$

$$\bar{I} = (\sum_{i=1}^{M} g_i)^{-1} \; ; \quad Var[I] = (\sum_{i=1}^{M} g_i)^{-2} \tag{4.25}$$

A successful transmission is obtained when a packet which breaks the channel idle period is not overlapped by any other transmission during its entire transmission period of length 1. The probability that user i among others begins transmission by breaking the channel idle period is given by

$$\nu_i = \frac{g_i}{\sum_{j=1}^{M} g_j} \; ; \quad \sum_{i=1}^{M} \nu_i = 1 \tag{4.26}$$

The probability of this user's success is then given by

$$\gamma_i = \exp\left[- \sum_{\substack{j=1 \\ (j \neq i)}}^{M} g_j \right] \qquad i = 1,2,\cdots,M \tag{4.27}$$

Thus we have the (conditioned) probability, q_i, that a successful transmission is achieved by user i as

$$q_i = \frac{g_i \gamma_i}{\sum_{j=1}^{M} g_j \gamma_j} = \frac{\nu_i \gamma_i}{\sum_{j=1}^{M} \nu_j \gamma_j} \qquad i = 1,2,\cdots,M \tag{4.28}$$

The probability of success for any user leading a transmission period is given by

$$\gamma = \sum_{i=1}^{M} \nu_i \gamma_i = \frac{\sum_{i=1}^{M} g_i \exp\left[- \sum_{\substack{j=1 \\ (j \neq i)}}^{M} g_j \right]}{\sum_{i=1}^{M} g_i} \tag{4.29}$$

The duration of a successful transmission period is constant:

$$T = 1 \tag{4.30}$$

The results in Eqs.(4.25)-(4.30) are exact. It remains for us to find the distribution of F. Note that F consists of an indefinite number of successive transmissions such that the intervals between their successive start times are all less than 1. Such an interval when arbitrarily chosen is distributed as in Eq.(4.9). A difficulty arises in finding the distribution of F; the successive intervals between transmission start times are neither independent nor identically distributed. However, let us introduce an approximation that they are independent and identically distributed as is given by Eq.(4.9). Thus, defining $f^{(n)}$ as the nth such interval in an interval F, we have

$$F = \sum_{n=1}^{L} f^{(n)} + 1 \tag{4.31}$$

The number of transmissions contained in an unsuccessful transmission period, denoted by L, is geometrically distributed (by approximation) as

$$Prob[L = n] \cong (1-\delta)^{n-1}\delta \qquad n = 1,2,\cdots$$

$$\bar{L} = \frac{1}{\delta} \;\; ; \qquad Var[L] = \frac{1-\delta}{\delta^2} \tag{4.32}$$

where δ is the probability that an arbitrary interval between two successive transmission start times is no shorter than the packet transmission time 1. In the context of a superposition process, it is defined by

$$\delta \triangleq Prob[\tilde{Y} \geq 1]$$

or, from Eq.(4.8),

$$\delta = \frac{\sum_{i=1}^{M} g_i}{\sum_{i=1}^{M} g_i \prod_{\substack{j=1 \\ (j \neq i)}}^{M} (1+g_j)} \tag{4.33}$$

Note that, in the above approximation, δ and each distribution of $f^{(n)}$ do not depend on the users who are transmitting during $f^{(n)}$. This fact implies that we have neglected the dependence of F on the sequence of the users whose transmissions constitute the interval F.

We are now in a position to calculate $\bar{F}$ and $Var[F]$ by use of the formula for the sum of independent random variables. From (4.31), they are expressed as

$$\bar{F} = \bar{L}\,\bar{f} + 1 \;\; ; \quad Var[F] = \bar{L}\,Var[f] + \bar{f}^2\,Var[L] \tag{4.34}$$

where f is the generic representation of the identically distributed $f^{(n)}$'s, and its distribution is given by Eq.(4.9). Also, $\bar{L}$ and $Var[L]$ are given in Eqs.(4.32) and (4.33). In the case of identical users (dropping the subscripts), using Eq.(4.11) we get

$$\bar{F} = \frac{(1+g)^M - 1 - gM(1+g)^{-(M-1)}}{gM[1-(1+g)^{-(M-1)}]} , \tag{4.35}$$

$$Var[F] = \frac{2(1+g)^{M-1}[(1+g)^{M+1}-1]}{g^2M(M+1)[(1+g)^{M-1}-1]} + \frac{(1+g)^{M-1}[(1+g)^M-1]^2[(1+g)^{M-1}-2]}{g^2M^2[(1+g)^{M-1}-1]^2}$$

$$+ \frac{(1+g)^{M-1}}{[(1+g)^{M-1}-1]^2}\left\{\frac{2(1+g)^{M-1}}{M} - \frac{2[(1+g)^M-1][(1+g)^{M-1}-1]}{gM} - 1\right\} \tag{4.36}$$

Let us examine the validity of our approximation just introduced by comparing $\bar{F}$ given in Eq.(4.35) and the exact expression for $\bar{F}$ (obtained through the mean-value argument; see Eq.(2.18)):

$$\bar{F}_{exact} = \frac{(1+g)^M - 1 - gM\,e^{-g(M-1)}}{gM\,[1-e^{-g(M-1)}]} \tag{4.37}$$

The difference between Eqs.(4.35) and (4.37) only comes from the difference between the terms $(1+g)^{-(M-1)}$ and $e^{-g(M-1)}$ which are close when $M \to \infty$ with G fixed at $G = gM$ and are identical in the limit. Figure 4.2 plots the relative error $(\bar{F} - \bar{F}_{exact})/\bar{F}_{exact}$. As expected, the error decreases as M increases. Even for $M = 3$, the relative error is less than 10%. For $M = 20$ users, the relative error is about 1%. Thus we adopt the independence assumption about the consecutive intervals between successive transmission start times. We may also note that

$$\bar{L} = (1+\frac{G}{M})^{(M-1)} \simeq e^G \quad \text{as} \quad M \to \infty$$

gives the average number of transmissions (except the leading one) involved in an unsuccessful transmission period.

Thus we have expressed all variables needed to evaluate Eqs.(4.23) and (4.24) in terms of $\{g_i\}$ and M. The numerical results in some example configurations are provided later along with those for unslotted CSMA (see Figures 4.4 and 4.5).

4.3.2 Unslotted CSMA with Zero Propagation Delay

We now proceed to study the packet interdeparture times for a population of unslotted nonpersistent CSMA users in a hidden-user environment. We assume the propagation delay to be negligible. Let g_i be the rate of transmission starts at user i when he hears an idle channel ($i = 1,2,\cdots,M$). Then, it is clear that I, $\{\nu_i\}$, and T are given by Eqs.(4.25), (4.26) and (4.30), respectively. In the following, we first determine $\{\gamma_i\}$ (from which $\{q_i\}$ can be calculated by Eq.(4.28)), and then assess the distribution of F approximately. To represent the hearing configuration, let us denote by $H(i)$ a set of user indices whose transmission can be heard by user i (including i himself) and let $H_i(i) \triangleq H(i) - \{i\}$. We assume symmetry in the hearing configuration; i.e., $H(i)$ is also the set of users who can hear transmission from user i.

When we consider the condition that a transmission by user i started by breaking the channel idle period be successful, we must concern ourselves with the behavior of the users who do not hear this transmission. (Due to the assumption of zero propagation delay, transmissions from users in $H_i(i)$ are immediately suppressed.) Since each of the users outside $H(i)$ may start transmission independently, the probability of success for user i is given by

$$\gamma_i = \exp\left[-\sum_{j \notin H(i)} g_j\right] \qquad i = 1,2,\cdots,M \tag{4.38}$$

As in Eq.(4.29), we have that the probability of success for any transmission period is

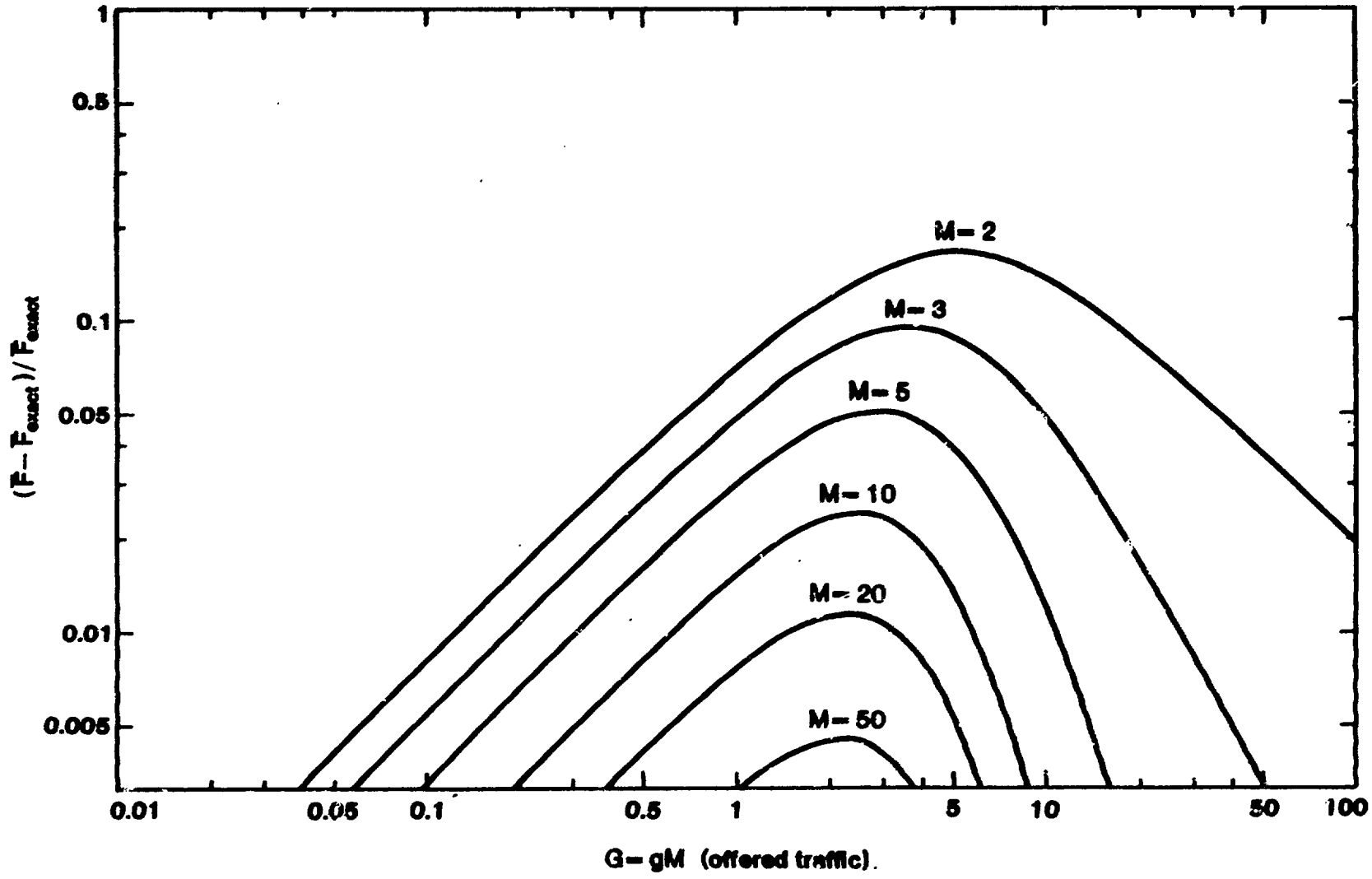


Figure 4.2 Relative error of the mean duration of an unsuccessful transmission period ($\bar{F}$) in pure ALOHA

$$\gamma = \sum_{i=1}^{M} \nu_i \gamma_i = \frac{\sum_{i=1}^{M} g_i \exp\left[-\sum_{j \in H(i)} g_j\right]}{\sum_{i=1}^{M} g_i} \tag{4.39}$$

We note that γ_i in Eq.(4.38) reduces to the expression in Eq.(4.27) in the case of pure ALOHA ($H(i) = \{i\}$), and that $\gamma_i = 1$ in the case of fully-connected CSMA ($H(i) = \{1,2,\cdots,M\}$).

A transmission started by breaking the channel idle period is unsuccessful with probability $1 - \gamma$. Then, how long is the duration of this unsuccessful transmission period F? Here again, F consists of a random number of consecutive transmissions such that the duration of each interval between two successive starts of transmission is less than 1. Therefore, we have the same intractability as in the analysis of pure ALOHA which has forced us to the approximation in Eq.(4.31). In addition, since each CSMA user does not behave independently once any transmission has started (he stops transmission initiations when he hears other transmissions), the independence of source processes in a superposition on which our approximation in pure ALOHA was based is no longer applicable. Nevertheless, we here introduce another assumption for approximation saying that the intervals between two successive transmission start times at each user are independent and identically distributed as given by Eq.(4.7) but with properly reduced transmission rates $\{g'_i\}$. We propose that the reduced rates $\{g'_i\}$ be determined as follows:

$$g'_i = g_i \cdot Prob\,[\text{user } i \text{ does not hear the current transmission(s)} \mid \text{at least one other user is transmitting}]$$

$$= g_i \cdot \frac{\prod_{j \in H_i(i)} \frac{g_j^{-1}}{1+g_j^{-1}} - \prod_{\substack{j=1 \\ (j \neq i)}}^{M} \frac{g_j^{-1}}{1+g_j^{-1}}}{1 - \prod_{\substack{j=1 \\ (j \neq i)}}^{M} \frac{g_j^{-1}}{1+g_j^{-1}}} \qquad i = 1,2,\cdots,M \tag{4.40}$$

where the factor $g_j^{-1}/(1+g_j^{-1})$ is the probability that user j is not transmitting under the assumption that he (independently of others) alternates between the transmission of length 1 and the exponentially distributed idle time (with mean $1/g_j$). Note that if user i is completely hidden from all other users, then Eq.(4.40) duly gives $g'_i = g_i$. On the other hand, if user i can be heard by all others, then Eq.(4.40) yields $g_i = 0$ which is fine again. In between these extreme cases, Eq.(4.40) gives g'_i between 0 and g_i depending on the connectivity of user i. The more user i is heard, the closer g'_i is to 0.

We thus assume that during any unsuccessful transmission period each user alternates between the transmission and idle states (independently of others as in pure ALOHA) with the parameter $\{g'_i\}$. So, for the distribution and the moments of F, we use the results in Eqs.(4.9), (4.31)-(4.36) where $\{g_i\}$ are replaced by $\{g'_i\}$. Now that we have expressed all

variables in terms of $\{g_i\}$ and $\{H(i)\}$ needed to calculate the distribution of X, let us proceed to apply our formulation to the two example systems whose hearing configurations are taken from [Toba75].

The first example, called a symmetric hidden-user configuration, consists of M identical users ($g_1 = g_2 = \cdots = g_M = g$) each of whom can hear transmissions from m users (including himself). The hearing graph for the case $M = 8$ and $m = 7$ in [Toba75] is reproduced in Figure 4.3(a). So, the case $m = 1$ corresponds to a pure ALOHA system while the case $m = M$ is equivalent to a population of fully-connected CSMA users. For this configuration, we have exactly

$$Prob[I \leq y] = 1 - e^{-Mgy} \qquad y \geq 0$$

$$\gamma_i = \gamma = e^{-(M-m)g} \qquad i = 1,2,\cdots,M \tag{4.41}$$

The reduced parameter $\{g'_i\}$ to calculate F approximately is given by

$$g'_i = g \cdot \frac{(1+g)^{-(m-1)} - (1+g)^{-(M-1)}}{1-(1+g)^{-(M-1)}} \triangleq g' \qquad i = 1,2,\cdots,M \tag{4.42}$$

Thus $\bar{F}$ and $Var[F]$ are evaluated via Eqs.(4.35) and (4.36), respectively, with g replaced by g'. Substituting these expressions into Eq.(4.21), we obtain the mean and variance of packet interdeparture time X as

$$\bar{X} = \left(\frac{1}{\gamma} - 1\right)\left(\frac{1}{G} + \bar{F}\right) + \frac{1}{G} + 1 \ ,$$

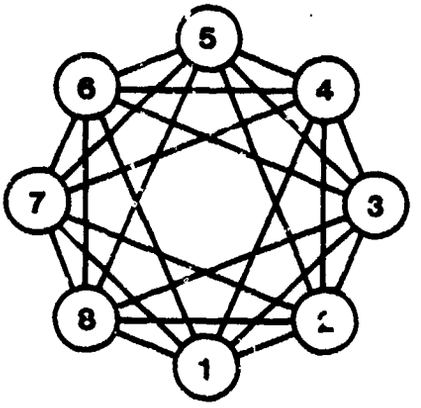
$$Var[X] = \frac{1}{\gamma G^2} + \left(\frac{1}{\gamma} - 1\right) Var[F] + \left(\frac{1}{G} + \bar{F}\right)^2 \frac{1-\gamma}{\gamma^2} \tag{4.43}$$

where we have used $G = gM$ as the aggregate rate of starting a transmission when the channel is idle. Note that $\bar{X}$ and $Var[X]$ depend on m through γ given by Eq.(4.41). In the case of fully-connected users ($m = M$), we have $g' = 0$ so that $\bar{F} = Var[F] = 0$ and $\gamma = 1$ (every transmission is successful). Then we have

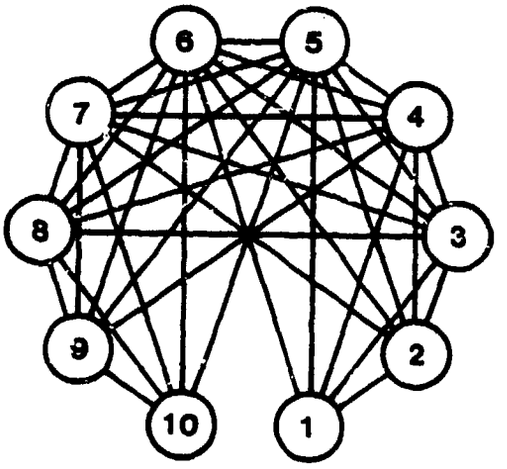
$$S = \frac{1}{\bar{X}} = \frac{G}{1+G} \quad ; \quad C^2 = \frac{Var[X]}{\bar{X}^2} = \frac{1}{(1+G)^2} \tag{4.44}$$

which is an exact result for the nonpersistent CSMA with no hidden users in the limit of zero propagation delay [Klei75b]. Note that the formulation in [Toba75] (which applies the technique of 'reduced rate' to both successful and unsuccessful transmissions indistinctly) fails to reach Eq.(4.44) when $M \to \infty$ with $m = M - 1$. In the case of pure ALOHA ($m = 1$), the results in Eq.(4.43) conform to those given in Section 4.3.1 because $g' = g$.

For the symmetric hidden-user configuration, [Toba75] does not give any results of numerical calculation or simulation for nonpersistent CSMA. Instead they show the simulation results for 1-persistent CSMA with propagation delay $a = 0.01$ and configuration $m = M - 1$. Figures 4.4(a) and (b) display and compare the numerical results for S and C^2 based on Eq.(4.43) for $M = 20$ users with various degree of connectivity. (The simulation results are discussed in Section 4.4.) The curves for $m = 1$ and $m = M$ are exact ones and other curves for



(a) Symmetric hidden-user configuration (M= 8, m= 7).



(b) Wall configuration (M= 10).

Figure 4.3 The hearing graphs of example configurations.

partially-hidden configurations are smoothly placed between the two extremities. This again may justify our approximation. In these figures the effects of hidden users on the system performance are clearly portrayed.

The second example, called a wall configuration in [Toba75], also consists of M users with identical parameters ($g_1 = g_2 = \cdots = g_M = g$) but with different connectivity. The graph representation of a hearing configuration for our example ($M = 10$) is shown in Figure 4.3(b) (reproduction of Figure 8(b) in [Toba75]). For this case, we have

$$v_i = \frac{1}{M} \quad ; \qquad \gamma_i = \exp\left[-g\left(M-1-|H_i(i)|\right)\right],$$

$$g'_i = g \cdot \frac{(1+g)^{-|H_i(i)|} - (1+g)^{-(M-1)}}{1-(1+g)^{-(M-1)}} \qquad i = 1,2,\cdots,M \tag{4.45}$$

where $|H_i(i)|$ stands for the number of user indices in set $H_i(i)$, i.e., the number of edges from node i in the hearing graph. We note in this case that

$$g'_1 > g'_2 > g'_3 > g'_4 > g'_5 = 0$$

In Figures 4.5(a) and (b), we show the numerical results for S_i and C_i^2 for individual users (solid curves) as well as system-wide S and C^2 (broken curves). They are compared to our simulation results in Section 4.4. The performance differentiation among users according to their connectivity is outstanding.

4.3.3 Unslotted CSMA with Delay Capture

When perfect delay capture is available in unslotted CSMA with zero propagation delay, every transmission period contains exactly one successful transmission (achieved by the user who starts first). However, due to possible overlapped transmissions by hidden users, the duration of each (successful) transmission period is now variable. The activity in the channel heard by the receiver consists of alternating cycles between transmission and idle periods. The duration of an idle period is exponentially distributed as given in Eq.(4.25), and $\{v_i\}$ are given by Eq.(4.26). With perfect capture, we have

$$\gamma_i = \gamma = 1 \quad ; \qquad q_i = v_i \qquad i = 1,2,\cdots,M \tag{4.46}$$

Note however that the sequence of channel activity is not different from the one of unslotted CSMA without capture; only the probability of attaining success is different. Therefore, we let

$$T = \begin{cases} 1 & \text{with probability } \gamma' \\ F' & \text{with probability } 1-\gamma' \end{cases} \tag{4.47}$$

where γ' and F' are γ and F, respectively, for the corresponding CSMA system without delay

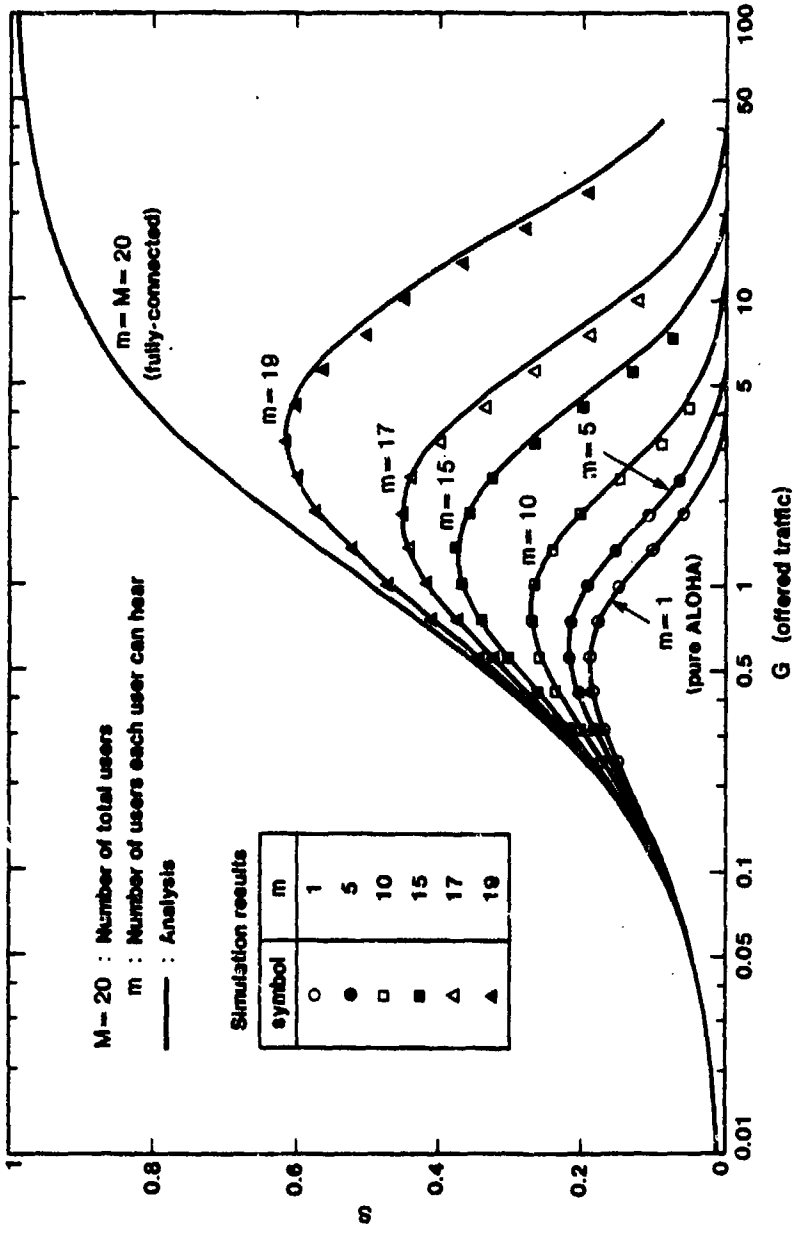


Figure 4.4(a) Throughput (S) in unslotted CSMA (zero propagation delay) for symmetric hidden-user configurations

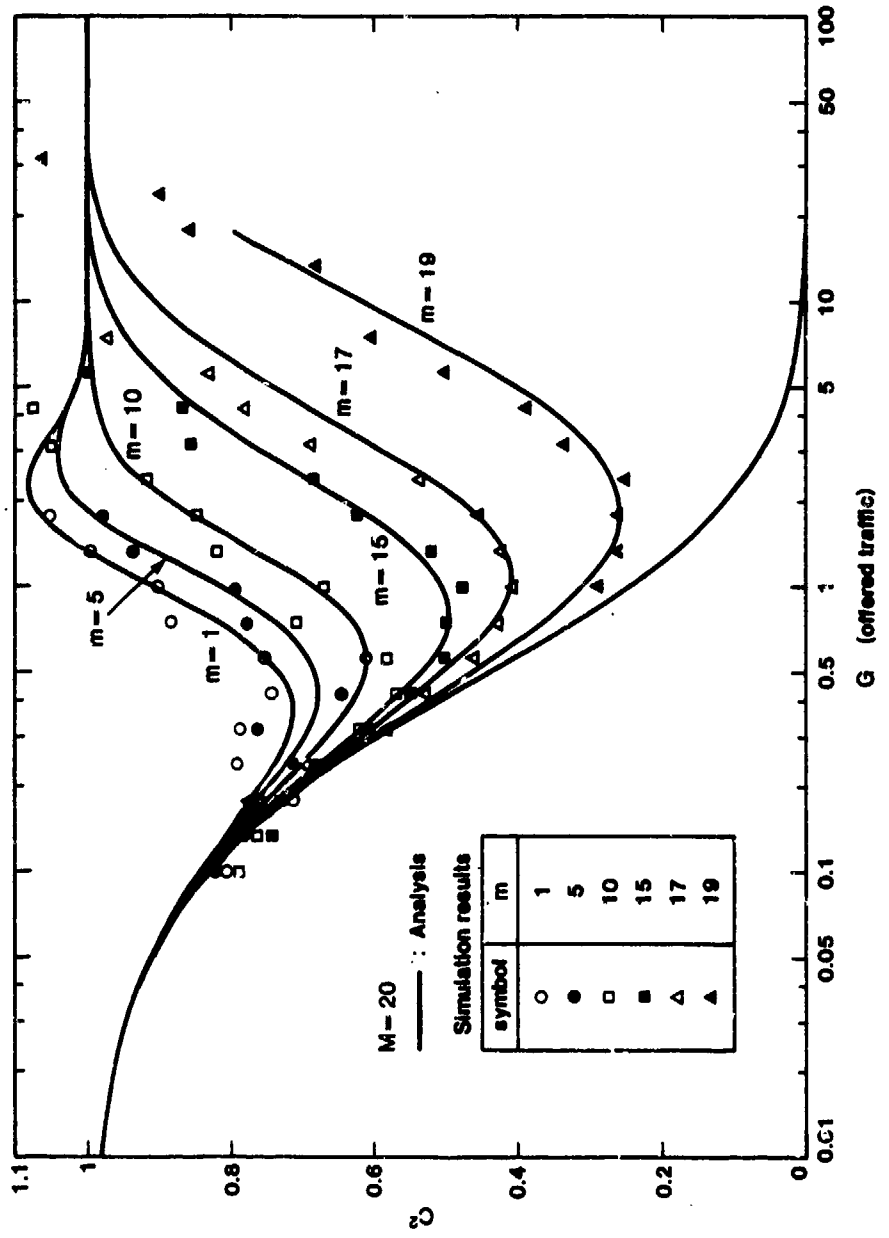


Figure 4.4(b) Coefficient of variation of the packet interdeparture time (C^2) in unslotted CSMA (zero propagation delay) for symmetric hidden-user configurations

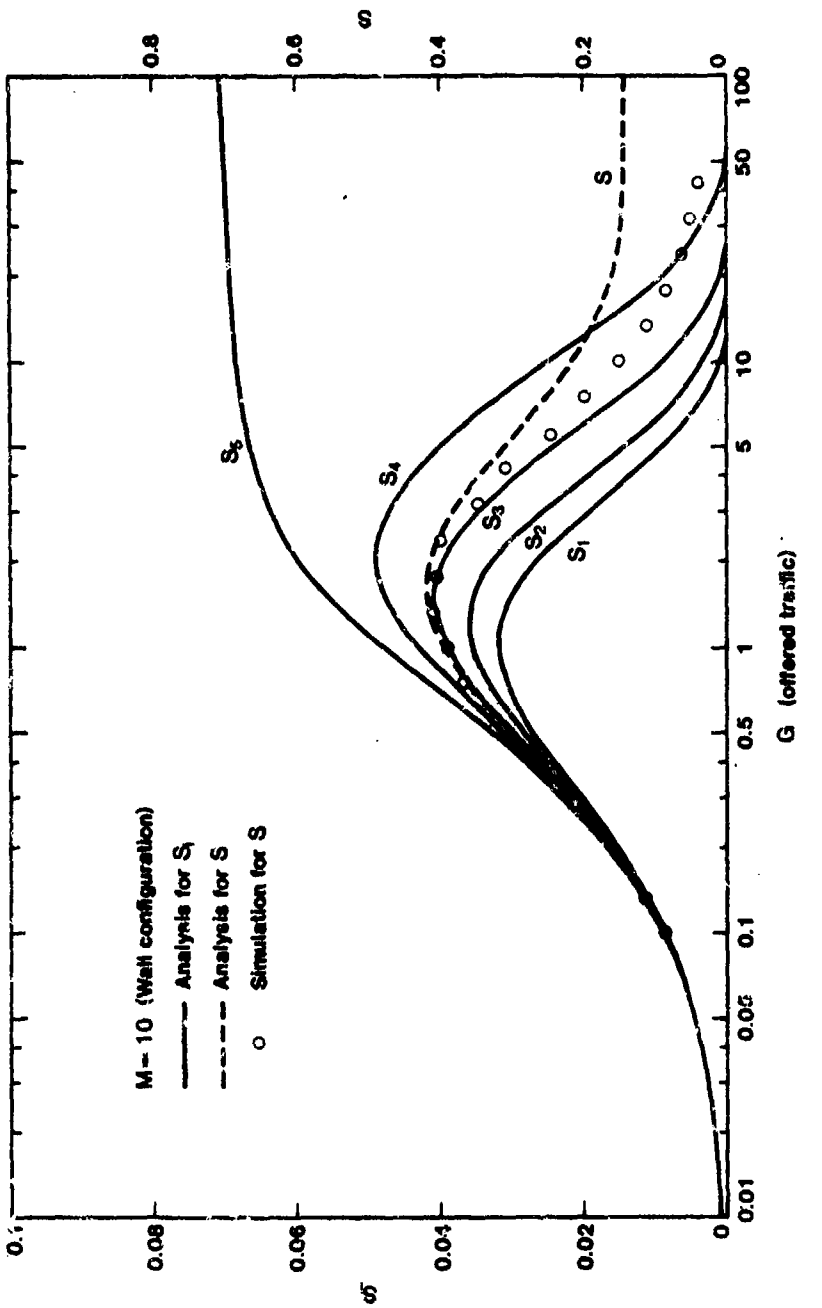


Figure 4.5(a) Individual user throughput (S_i) and channel throughput (S) in unslotted CSMA (zero propagation delay) for a wall configuration in Figure 4.3(b)

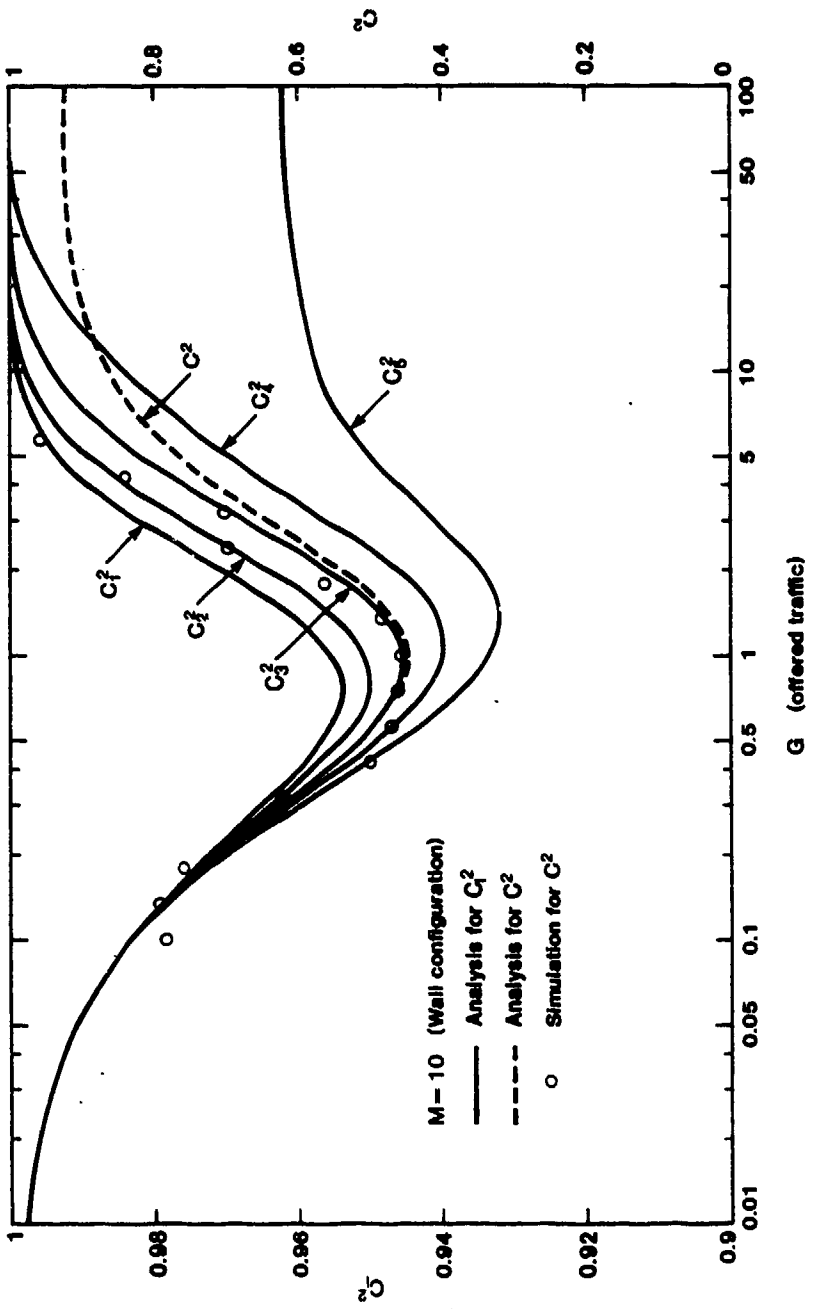


Figure 4.5(b) Coefficient of variation of the packet interdeparture time (C^2 for the system and C_i^2 for user i) in unslotted CSMA (zero propagation delay) for a wall configuration in Figure 4.3(b)

capture. Then we have

$$\bar{X} = \gamma'(\bar{I}+1) + (1-\gamma')(\bar{I}+\bar{F}'),$$

$$Var[X] = \gamma' Var[I] + (1-\gamma')(Var[I] + Var[F']) + \gamma'(1-\gamma')(\bar{F}'-1)^2 \qquad (4.48)$$

In Figures 4.6(a) and (b), we show the results of numerical calculation for S and C^2 in the same configuration as the first example of Section 4.3.2 ($M=20$ identical users with connectivity m). Note that $G=gM$. Again, the case $m=1$ corresponds to a pure ALOHA system for which we have an exact expression for the throughput (see Eq.(2.22) with $c=0$):

$$S_{exact} = \frac{gM}{(1+g)^M} \qquad (4.49)$$

(No exact expressions for $Var[X]$ are available.) For a system of as many as $M=20$ users, the difference in the throughput between the above exact values and our approximate values is indiscernible in the plot. In the other extreme case, $m=M$, corresponding to a fully-connected CSMA system, we recover the results in Eq.(4.44). In these figures, we recognize the effects of hidden users on X which change smoothly depending on the connectivity. If we compare these figures with Figures 4.4(a) and (b), the throughput enhancement due to delay capture is clear.

4.3.4 Unslotted CSMA with Nonzero Propagation Delay

The treatment of unslotted CSMA with nonzero propagation delay a is a straightforward extension of the formulation in Section 4.3.2, and the notation there is carried over here. The results below in a special case where $H(i)=\{i\}$ correspond to those for pure ALOHA with nonzero propagation delay, while the results in another special case where $H(i)=\{1,2,\cdots,M\}$ conform to those for fully-connected CSMA considered in Chapter 3. Also, in the limit $a \to 0$, the results reduce to those in Section 4.3.2.

Given the rates of starting transmission $\{g_i\}$, the distribution of a channel idle period I is given as in Eq.(4.25), and the probability ν_i that user i initiates a transmission period is given by Eq.(4.26). The probability of this user's success is then given by

$$\gamma_i = \gamma_i^{(1)}\gamma_i^{(2)} \qquad i=1,2,\cdots,M \qquad (4.50)$$

where

$$\gamma_i^{(1)} = \exp\left[-(1+a)\sum_{j\in H(i)} g_j\right]; \quad \gamma_i^{(2)} = \exp\left[-a\sum_{j\in H_i(i)} g_j\right] \qquad (4.51)$$

It should be clear that $\gamma_i^{(1)}$ accounts for the probability that no users hidden from user i (initiating the transmission period) do not start transmission during time $1+a$, and that $\gamma_i^{(2)}$ is the probability that those users who can hear the leading transmission by user i do not start transmission during time a. Then, from $\{\gamma_i\}$, the probabilities $\{q_i\}$ with which each successful

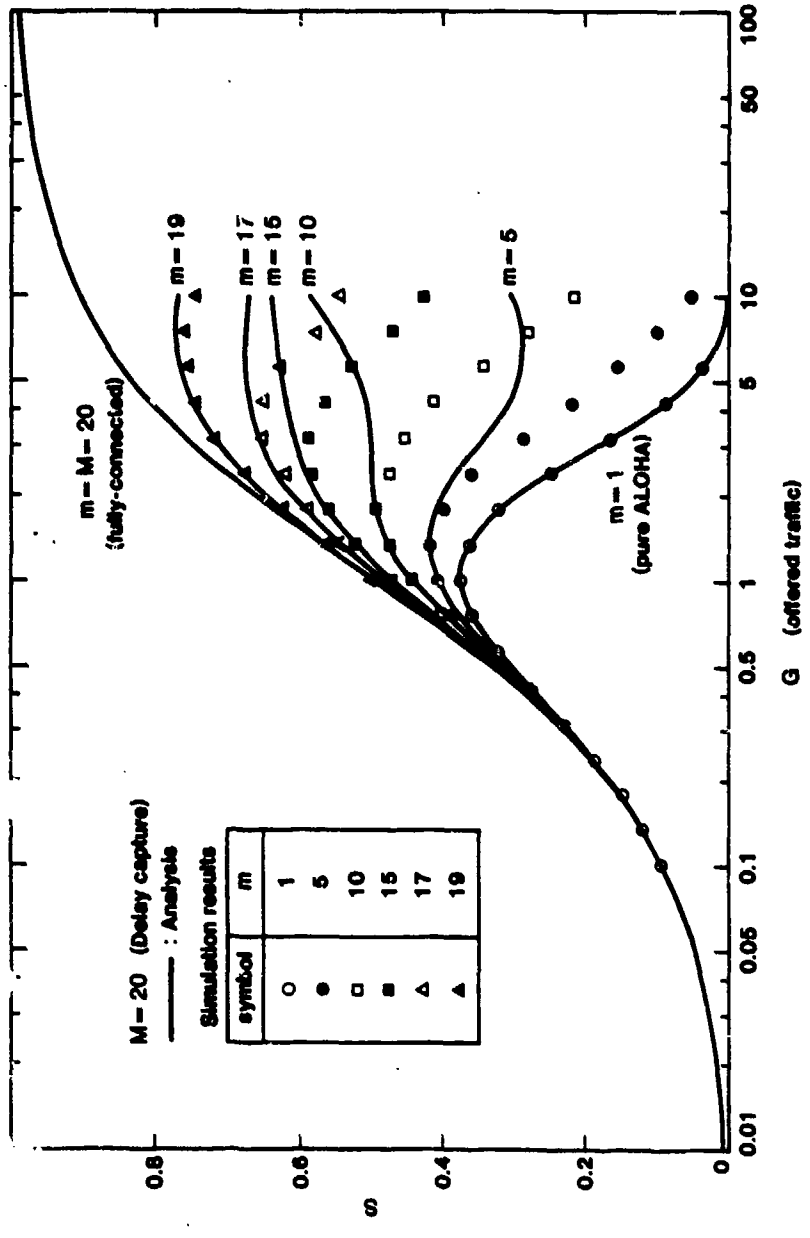


Figure 4.6(a) Throughput (S) in unslotted CSMA (zero propagation delay) with perfect capture for symmetric hidden-user configurations

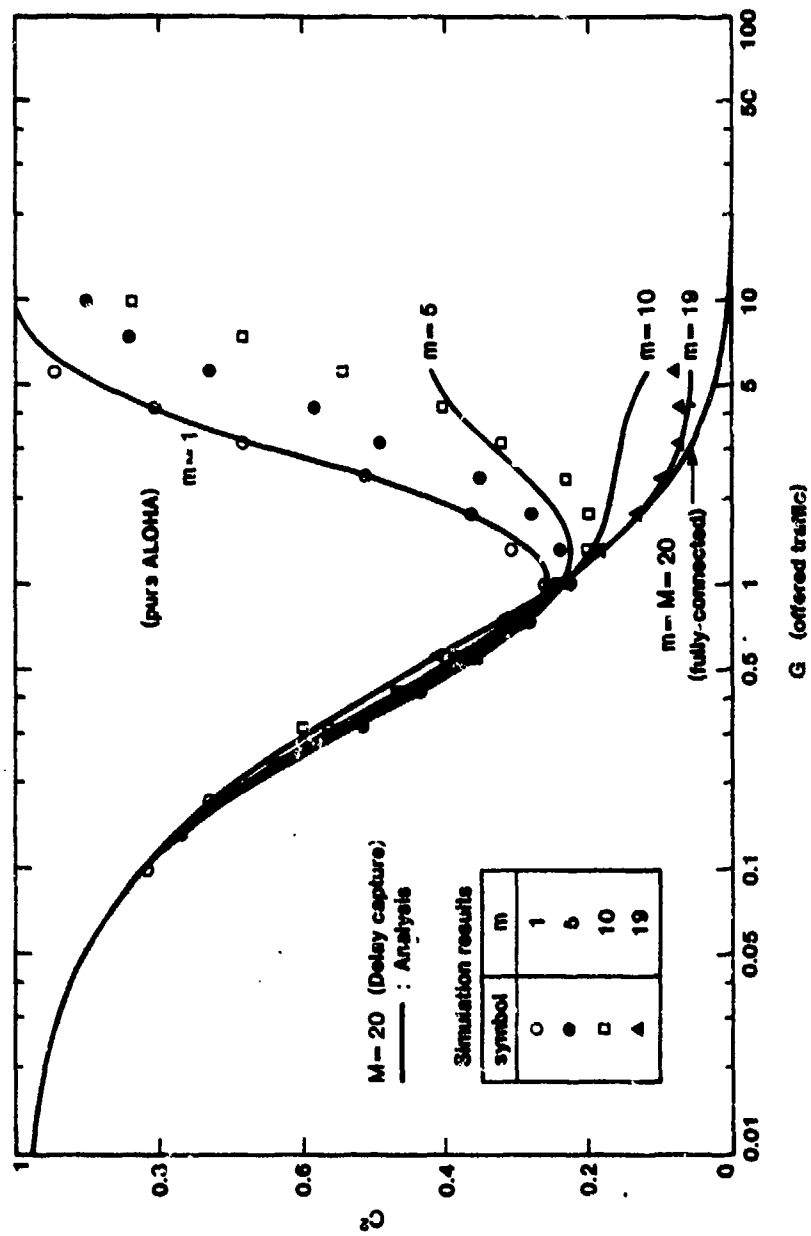


Figure 4.6(b) Coefficient of variation of the packet interdeparture time (C^2) in unslotted CSMA (zero propagation delay) with perfect capture for symmetric hidden-user configurations

transmission is achieved by each user can be calculated via Eq.(4.28). The duration of a successful transmission period is constant:

$$T = 1 + a \tag{4.52}$$

In order to deal with the duration of an unsuccessful transmission period F, we distinguish two kinds of unsuccessful transmission periods; the first kind (whose duration is denoted by $F^{(1)}$) is one such that no transmissions by the users hidden from the leading user are involved in the transmission period, and the second kind (whose duration is denoted by $F^{(2)}$) is one containing transmissions from hidden users. Thus, if we denote by $F(i)$, $F^{(1)}(i)$ and $F^{(2)}(i)$ the durations of an unsuccessful transmission period and its two kinds which is initiated by user i, respectively, then we have

$$F(i) = \begin{cases} F^{(1)}(i) & \text{with probability } \gamma_i^{(1)}[1-\gamma_i^{(2)}]/(1-\gamma_i) \\ F^{(2)}(i) & \text{with probability } (1-\gamma_i^{(1)})/(1-\gamma_i) \end{cases} \tag{4.53}$$

Now the distribution of $F^{(1)}(i)$ is given in Chapter 3 since this corresponds to the duration of an unsuccessful transmission period initiated by user i in a fully-connected environment. That is, if $Y(i)$ is the transmission start time of the last colliding packet in an unsuccessful transmission period initiated by user i, we have (see Eqs.(3.34) and (3.35))

$$F^{(1)}(i) = 1 + a + Y(i) ,$$

$$Prob\,[\,Y(i) \leqslant y\,] = \frac{\prod_{j \in H_i(i)} (1 - e^{-g_j y} + e^{-g_j a}) - \gamma_i^{(2)}}{1 - \gamma_i^{(2)}} \qquad 0 \leqslant y \leqslant a \tag{4.54}$$

It remains for us to find the distribution of $F^{(2)}(i)$ by approximation. For convenience, let us define

$$\gamma_1 = \sum_{i=1}^{M} \nu_i \gamma_i^{(1)} \tag{4.55}$$

and express F in the two cases:

$$F = \begin{cases} F^{(1)} & \text{with probability} \quad \dfrac{1}{1-\gamma} \displaystyle\sum_{i=1}^{M} \nu_i \gamma_i^{(1)}[1-\gamma_i^{(2)}] \\ F^{(2)} & \text{with probability} \quad \dfrac{1}{1-\gamma} \displaystyle\sum_{i=1}^{M} \nu_i [1-\gamma_i^{(1)}] \end{cases} \tag{4.56}$$

where

$$Prob\,[\,F^{(1)} \leqslant x\,] = \frac{1}{\gamma_1 - \gamma} \sum_{i=1}^{M} \nu_i \gamma_i^{(1)}[1-\gamma_i^{(2)}]\,Prob\,[\,F^{(1)}(i) \leqslant x\,] ,$$

$$Prob\,[F^{(2)} \leqslant x] = \frac{1}{1-\gamma_1} \sum_{i=1}^{M} \nu_i [1-\gamma_i^{(1)}]\, Prob\,[F^{(2)}(i) \leqslant x] \tag{4.57}$$

Recall that $F^{(1)}$ is the duration of an unsuccessful transmission period (unconditioned on the initial user) in a fully-connected environment, and that $F^{(2)}$ corresponds to the case involving hidden-users' transmissions. Since Eq.(4.56) implies

$$Prob\,[F \leqslant x] = \frac{\gamma_1-\gamma}{1-\gamma}\, Prob\,[F^{(1)} \leqslant x] + \frac{1-\gamma_1}{1-\gamma}\, Prob\,[F^{(2)} \leqslant x] \tag{4.58}$$

it follows that

$$\bar{F} = \frac{\gamma_1-\gamma}{1-\gamma} E[F^{(1)}] + \frac{1-\gamma_1}{1-\gamma} E[F^{(2)}]\,,$$

$$Var[F] = \frac{\gamma_1-\gamma}{1-\gamma} E[\{F^{(1)}\}^2] + \frac{1-\gamma_1}{1-\gamma} E[\{F^{(2)}\}^2] - \bar{F}^2 \tag{4.59}$$

By Eqs.(4.54) and (4.57) we know the distribution of $F^{(1)}$. Specifically, the first two moments of $F^{(1)}$ are given by

$$E[F^{(1)}] = \frac{1}{\gamma_1-\gamma} \sum_{i=1}^{M} \nu_i \gamma_i^{(1)} [1-\gamma_i^{(2)}]\,(1+a+\overline{Y(i)})\,,$$

$$E[\{F^{(1)}\}^2] = \frac{1}{\gamma_1-\gamma} \sum_{i=1}^{M} \nu_i \gamma_i^{(1)} [1-\gamma_i^{(2)}]\,\overline{(1+a+Y(i))^2} \tag{4.60}$$

where

$$E[\{Y(i)\}^k] = \frac{1}{1-\gamma_i^{(2)}} \int_0^a (k y^{k-1}) [1 - \prod_{j \in H_i(i)} (1 - e^{-\beta_j y} + e^{-\beta_j a})]\, dy \qquad k = 1,2 \tag{4.61}$$

It is the distribution of $F^{(2)}$ (which has the same twofold intractability as for F in Section 4.3.2) that we have recourse to approximation. We here simply write down the expressions to calculate the distribution of $F^{(2)}$ similar to the treatment in Section 4.3.2 but with a modification due to the finite value of a.

$$F^{(2)} = \sum_{n=1}^{L} f^{(n)} + 1 + a \tag{4.62}$$

where

$$Prob\,[L = n] \cong (1-\delta)^{n-1}\delta \qquad\qquad n = 1,2,\cdots$$

$$\delta = \frac{\sum_{i=1}^{M} g'_i}{\sum_{i=1}^{M} g'_i \prod_{\substack{j=1 \\ (j\neq i)}}^{M} [(1+a)g'_j+1]} \tag{4.63}$$

and

$$Prob\,[f^{(a)} > x] = \frac{\sum_{i=1}^{M} g'_i \left\{ \prod_{\substack{j=1 \\ (j\neq i)}}^{M} [(1+a-x)g'_j+1]-1 \right\}}{\sum_{i=1}^{M} g'_i \left\{ \prod_{\substack{j=1 \\ (j\neq i)}}^{M} [(1+a)g'_j+1]-1 \right\}} \qquad 0 \leqslant x \leqslant 1+a \tag{4.64}$$

The reduced rates of starting transmission $\{g'_i\}$ are given by

$$g'_i = g_i \cdot \frac{\prod_{j\in H_i(i)} \frac{g_j^{-1}}{1+a+g_j^{-1}} - \prod_{\substack{j=1 \\ (j\neq i)}}^{M} \frac{g_j^{-1}}{1+a+g_j^{-1}}}{1-\prod_{\substack{j=1 \\ (j\neq i)}}^{M} \frac{g_j^{-1}}{1+a+g_j^{-1}}} \qquad i=1,2,\cdots,M \tag{4.65}$$

In the case of identical users (dropping the subscripts), we have the following simple expressions for the mean and variance of $F^{(2)}$:

$$E[F^{(2)}] = \frac{g_1^M - 1 - (1+a)\,g'Mg_1^{-(M-1)}}{g'M\,[1-g_1^{-(M-1)}]},$$

$$Var[F^{(2)}] = \frac{2g_1^{M-1}(g_1^{M+1}-1)}{(g')^2M(M+1)(g_1^{M-1}-1)} + \frac{g_1^{M-1}(g_1^M-1)^2(g_1^{M-1}-2)}{(g'M)^2(g_1^{M-1}-1)^2}$$

$$+ \frac{(1+a)^2 g_1^{M-1}}{(g_1^{M-1}-1)^2}\left(\frac{2g_1^{M-1}}{M}-1\right) - \frac{2(1+a)g_1^{M-1}(g_1^M-1)}{g'M(g_1^{M-1}-1)} \tag{4.66}$$

where

$$g_1 \triangleq (1+a)g'+1$$

Thus we can calculate $\bar{F}$ and $Var[F]$ by Eq.(4.59). All other quantities needed to evaluate Eqs.(4.23) and (4.24) have also been given. Figures 4.7(a) and (b) plot S and C^2 for the symmetric hidden-user configurations when $M = 20$ and $m = M-1 = 19$. The curves for $a=0$ are copied from Figures 4.4(a) and (b).

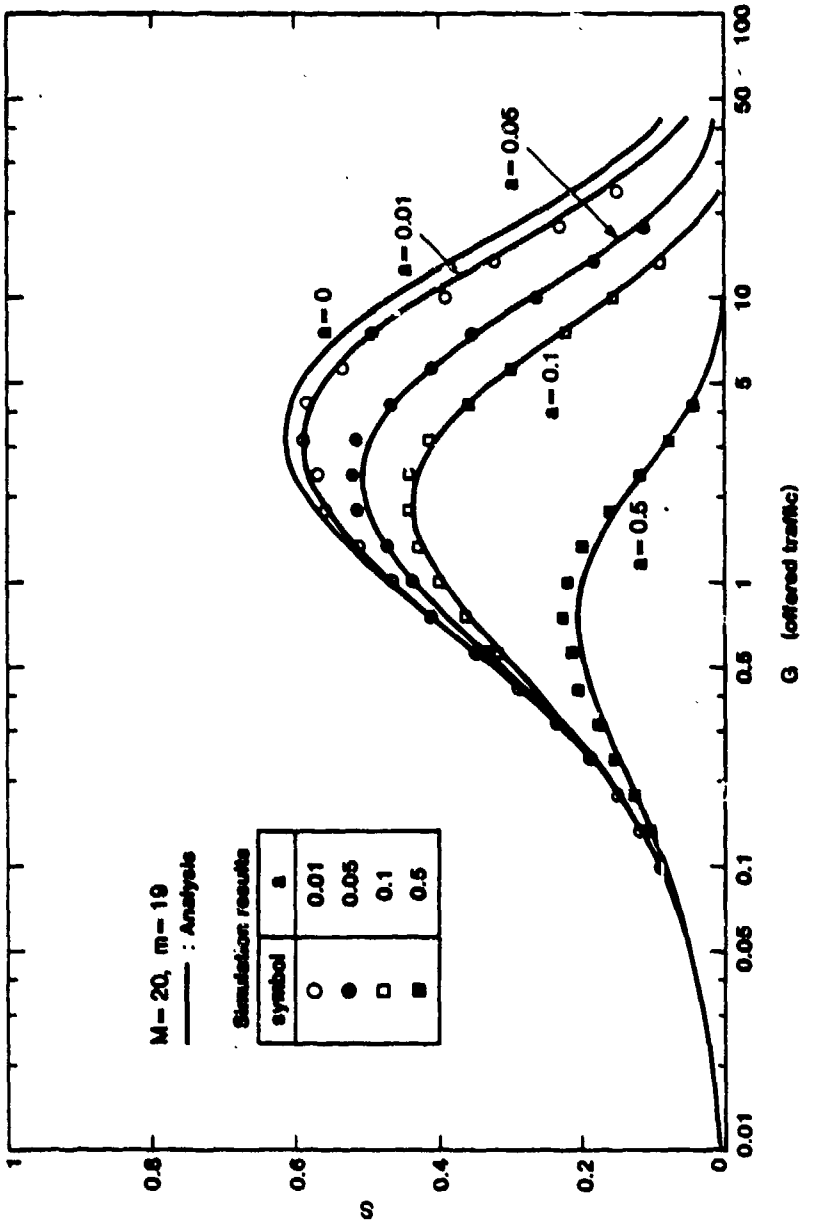


Figure 4.7(a) Throughput (S) in unslotted CSMA (nonzero propagation delay) for a symmetric hidden-user configuration (M = 20, m = 19)

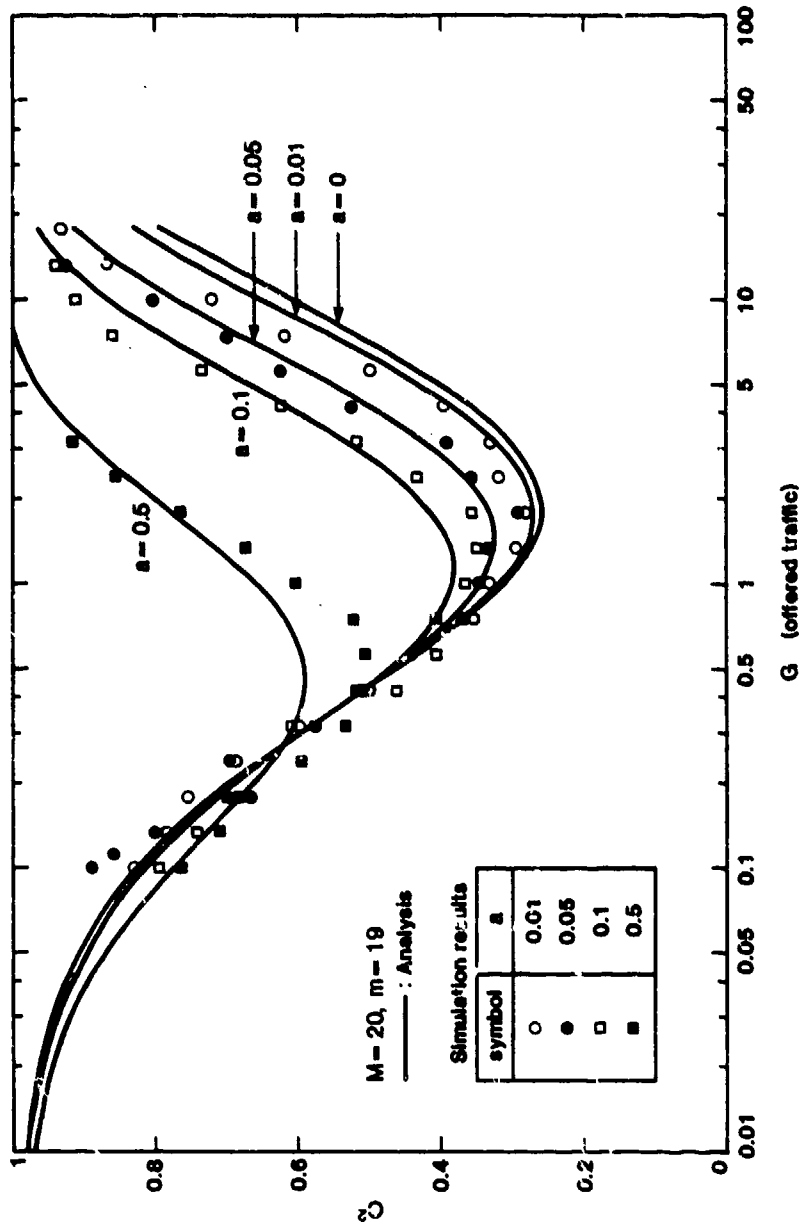


Figure 4.7(b) Coefficient of variation of the packet interdeparture time (C^2) in unslotted CSMA (nonzero propagation delay) for a symmetric hidden-user configuration (M = 20, m = 19)

4.3.5 Slotted CSMA

Lastly, we consider slotted CSMA where the (mini)slot size is equal to the propagation delay a (>0). As in Section 4.2.2, let $\tau \triangleq 1/a$ be an integer. Note then that a transmission period at each user consists of $(1+a)/a = \tau+1$ slots. Let p_i be the probability of starting transmission by user i when the channel is sensed idle ($i=1,2,\cdots,M$). Then the probability that a transmission started by user i is successful is given by

$$\gamma_i = \prod_{\substack{j=1 \\ (j\neq i)}}^{M} (1-p_j) \left[\prod_{j\notin H(i)} (1-p_j) \right]^{\tau} \qquad i=1,2,\cdots,M \tag{4.67}$$

i.e., the probability that no other users start transmission in the same slot as user i does and that no users hidden from user i start transmission during the rest of transmission period (of duration τ slots). The probability that user i is one of the users who start transmission breaking the channel idle period is given by

$$\nu_i = p_i/(1-E) \qquad i=1,2,\cdots,M \tag{4.68}$$

where

$$E = \prod_{i=1}^{M} (1-p_i) \tag{4.69}$$

is the probability that an idle slot is followed by another idle slot. Note that $\sum_{i=1}^{M} \nu_i \neq 1$ because there can be more than one simultaneous transmission starts. The probability that any user starts a successful transmission is given by

$$\gamma = \sum_{i=1}^{M} \nu_i \gamma_i = \frac{\sum_{i=1}^{M} p_i \gamma_i}{1-E} \tag{4.70}$$

and the (conditional) probability that a successful transmission is achieved by user i is

$$q_i = \frac{\nu_i \gamma_i}{\gamma} = \frac{p_i \gamma_i}{\sum_{j=1}^{M} p_j \gamma_j} \qquad i=1,2,\cdots,M \tag{4.71}$$

The duration of channel idle period is geometrically distributed as

$$Prob[I = na] = E^n(1-E) \qquad n=0,1,2,\cdots$$

$$\bar{I} = \frac{Ea}{1-E} \quad ; \quad Var[I] = \frac{Ea^2}{(1-E)^2} \tag{4.72}$$

The duration of a successful transmission period is constant:

$$T = 1+a \tag{4.73}$$

Note that all the results up to now (in this subsection) are exact. However, we can evaluate the duration of an unsuccessful transmission period F only approximately. To do this, we first determine a proper reduction of transmission parameter value for each user according to his connectivity. If user i behaved independently of others alternating the idle period (denoted by I_i, distributed as $Prob[I_i = na] = (1-p_i)^n p_i$, $n = 0,1,2,\cdots$, with mean $\bar{I}_i = (1-p_i)a/p_i$) and the transmission period (of constant duration $1+a$), then the probability that user i is in the idle state at an arbitrary epoch is given by

$$\frac{\bar{I}_i}{1+a+\bar{I}_i} = \frac{(1-p_i)\,a}{p_i+a}$$

Now, similar to Eq.(4.40) let us define

$$p'_i = p_i \cdot \frac{\prod\limits_{j\in H_i(i)} \frac{(1-p_j)\,a}{p_j+a} - \prod\limits_{\substack{j=1\\(j\neq i)}}^{M} \frac{(1-p_j)\,a}{p_j+a}}{1 - \prod\limits_{\substack{j=1\\(j\neq i)}}^{M} \frac{(1-p_j)\,a}{p_j+a}} \qquad i = 1,2,\cdots,M \tag{4.74}$$

We use $\{p'_i\}$ given by Eq.(4.74) in Eqs.(4.17) and (4.18) to calculate the mean and variance of $f^{(n)}$, the nth interval between two successive transmission starts which is no longer than τ slots. The duration of an unsuccessful transmission period F is expressed as

$$F = \sum_{n=1}^{\tau} f^{(n)} + 1 + a \tag{4.75}$$

The distribution of L is given by

$$Prob[L = n] \cong (1-\delta)^{n-1}\delta \qquad n = 1,2,\cdots \tag{4.76}$$

where δ is equivalent to $Prob[\tilde{Y} > 1]$ in Section 4.2.2; i.e.,

$$\delta = \frac{\prod\limits_{i=1}^{M}(p'_i + P'_i) - \prod\limits_{i=1}^{M} P'_i}{\prod\limits_{i=1}^{M}(1+\tau p'_i) - \prod\limits_{i=1}^{M}[1+p'_i(\tau-1)]} \tag{4.77}$$

with $P'_i \triangleq (1-p'_i+\tau p'_i)/(1+\tau p'_i)$.

We are now able to evaluate S and C^2. As an example, we again consider a symmetric hidden-user configuration with $M = 20$ and $m = M-1 = 19$ ($p_1 = p_2 = \cdots p_M = p$). The total offered channel traffic is defined as $G = pM/a$. In Figures 4.8(a) and (b), we plot S and C^2 against G for some values of a. When $a \to 0$, the results conform to those for unslotted CSMA with $a = 0$ shown in Figures 4.4(a) and (b). Also comparison of the corresponding curves between Figures 4.7 (unslotted CSMA) and 4.8 shows the higher S and smaller C in slotted CSMA.

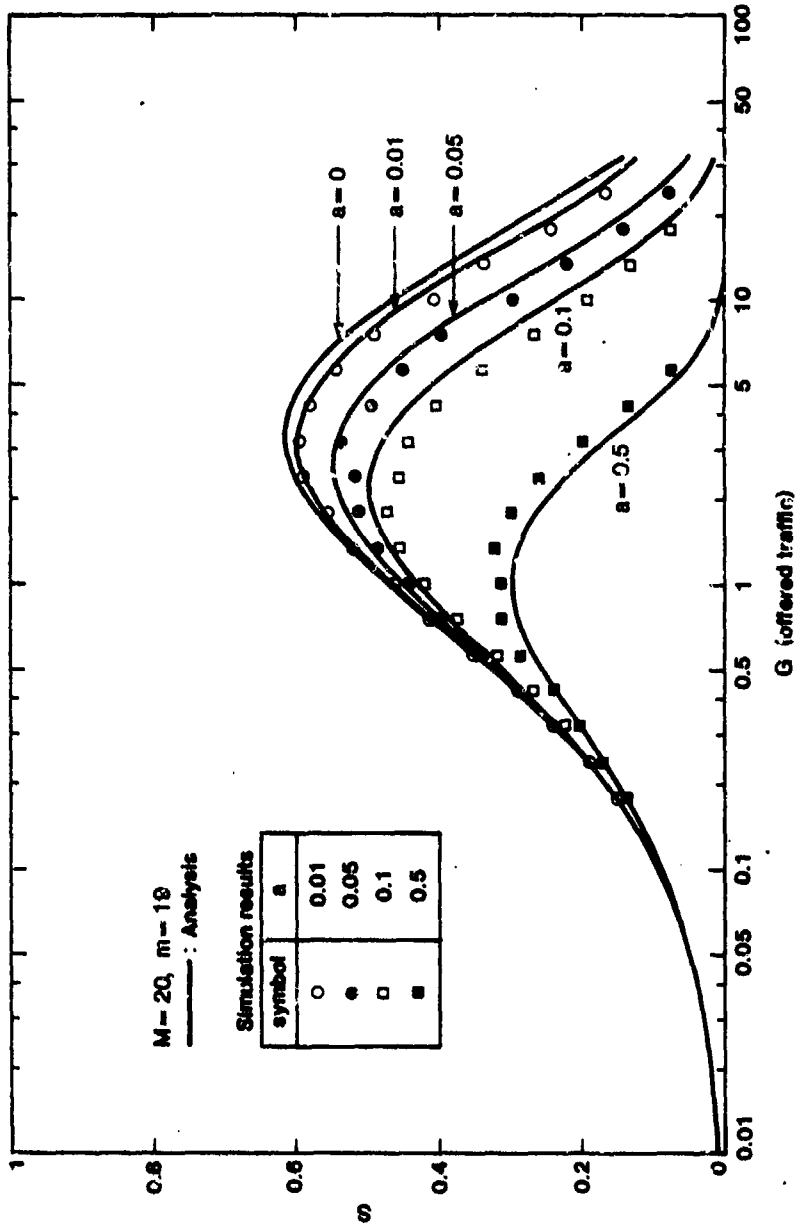


Figure 4.8(a) Throughput (S) in slotted CSMA for a symmetric hidden-user configuration (M = 20, m = 19)

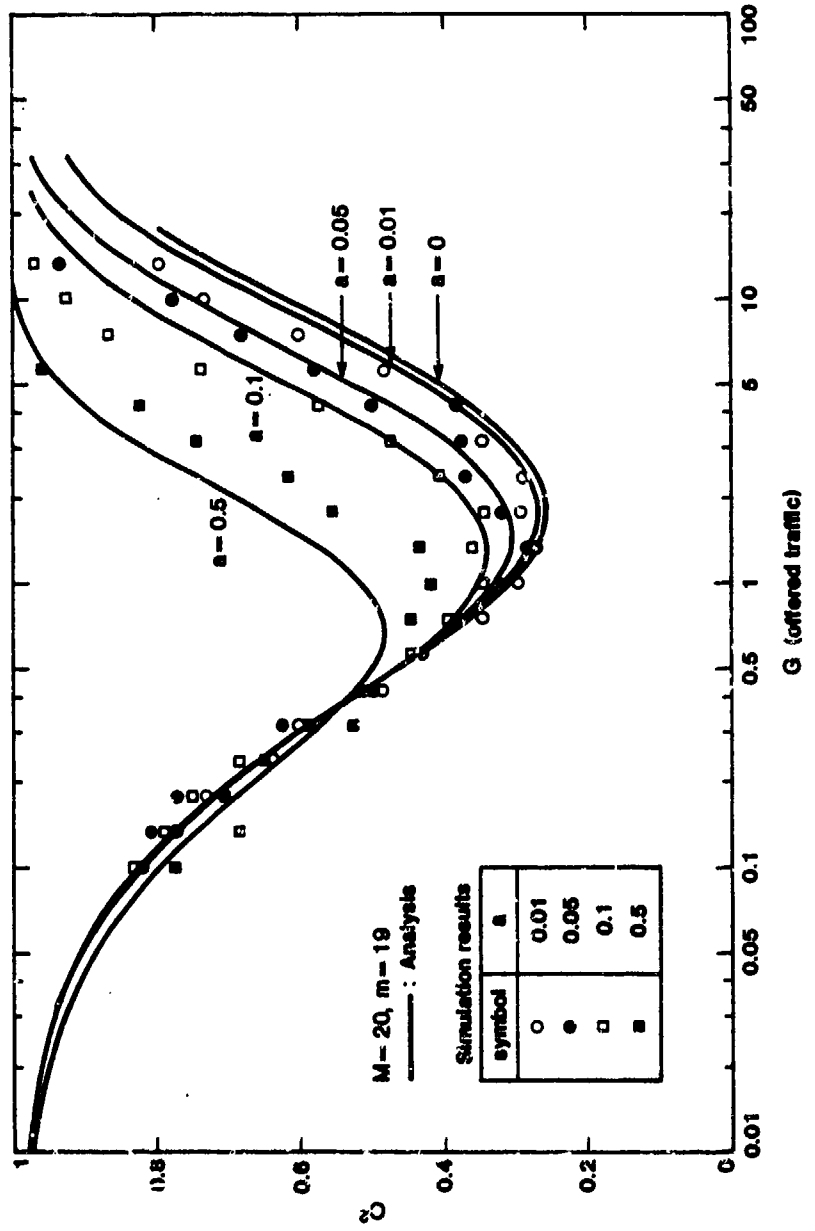


Figure 4.8(b) Coefficient of variation of the packet interdeparture time (C^2) in slotted CSMA for a symmetric hidden-user configuration (M = 20, m = 19)

4.4 Discussion

In this section, after describing the simulation model we compare the simulation results with those calculated by using our approximation. Then some of our basic assumptions for the approximation are discussed with regard to their rationale. At the same time, suggestions on the refinement of the present model are made.

In the simulation program, the nonpersistent CSMA protocol is implemented as follows: when a user finds any ongoing transmission at the time of his scheduled transmission, he defers (reschedules) his transmission by an exponentially distributed time after the end of the current transmission. Given a set of parameter values (a, M, m, $\{g_i\}$ or $\{p_i\}$, and $\{H(i)\}$), we collected 2,000 interdeparture times and computed their sample mean and coefficient of variation. These results are shown in the figures which plot the corresponding analytical results (Figures 4.4-4.8).

Figure 4.4 for the symmetric hidden-user configurations in unslotted CSMA with zero propagation delay manifests excellent agreement between our approximation and simulation results over almost the whole range of the offered channel traffic value G in each case of hiddenness. Although agreement for small G is not surprising because we do not have many collisions there anyway, the agreement for large G is noteworthy. (The treatment in [Toba75] claims its applicability only for relatively small G.) In Figure 4.5, we have shown the results for the wall configuration depicted in Figure 4.3(b). The simulation results are only shown for the system-wide quantities S and C^2 because their distribution among the users according to Eq.(4.24) are exact. In this case, agreement is not very good for large G, specifically after G exceeds its optimal value which makes S maximal. The discrepancy at large G may be explained as follows. Consider the case $G \to \infty$. In the simulation (and in reality) there is a probability 1/10 for each of fully-connected users 5 and 6 with which he can initiate a transmission period and (due to full-connectivity) be successful (so $\gamma = 1/5$). However, since the duration of other transmission periods (initiated by partially-connected users) can be infinitely long, the contribution to the throughput from users 5 and 6 becomes infinitesimal. Our modeling, however, assumes no user identities in determining the distribution of F as noted after Eq.(4.33). As a result of transmission rate reduction by Eq.(4.45), we have $\{g'_i\} \simeq \{0\}$ as $G \to \infty$, so that $\delta \to 1$, $f^{(n)} \to 1/2$, and so $F \to 3/2$. It follows that $S \to \gamma/[(1-\gamma)F + \gamma \cdot 1] \simeq 1.42$ as shown in Figure 4.5(a). Therefore it is clear that the discrepancy comes from the inaccurate evaluation of F. However, our approximation may still be useful because we usually operate the system around the optimal value of G. A possible refinement for a population of dissimilar users as in this example is suggested below. Figure 4.6 is for the same settings as in Figure 4.5 but with perfect delay capture. Here again (except the case of pure ALOHA for which agreement is uniformly excellent) the agreement is good only until G attains its optimum. Figures 4.7 and 4.8 are for the symmetric hidden-user configurations in CSMA with a finite propagation delay a. The agreement (over the whole range of G) is again excellent for any reasonable value of a.

Now, let us examine some of our assumptions which have been introduced to make the analysis tractable. In Eqs.(4.31), (4.62) and (4.75), we have assumed that each $f^{(n)}$, the interval between two successive transmission start times such that it is shorter than the packet transmission time, is independent and identically distributed, while, in reality, they are not. We have, however, a theoretical rationale for this assumption for a large user population: i.e., the fact that a large number of merged point processes tends to be a Poisson stream with aggregate rate (Palm-Khintchine theorem; see, e.g., [Heym82a]). The apparent good agreement of our results with simulation for as many as 20 users endorses this assertion. In our second assumption, to account for the carrier-sensing effects, we have reduced the transmission rates and used them as parameters in the exponential distributions for transmission rescheduling. In reality, however, since we reschedule transmission at indefinite times until we sense the channel idle, the interval between the actual transmissions is likely to be distributed as a random number of exponentially distributed times rather than as a single exponential distribution with reduced rate. An example of the resultant inaccurate evaluation for the duration of an unsuccessful transmission period is the above-mentioned discrepancy for the wall configuration. Thus, an improvement based on this observation is desired for refinement.

Also in expressing F in Eqs.(4.31), (4.62) and (4.75), we have ignored the dependence of F on the detailed sequence of user indices involved therein for the sake of mathematical tractability. However, in the case of unslotted CSMA (including pure ALOHA) with zero propagation delay, for example, we know exactly the distribution of $f^{(1)}$ depending on the user who initiates the transmission period. Specifically, if we denote by $f^{(1)}(i)$ the $f^{(1)}$ when the transmission period is initiated by user i, then its distribution is given by

$$Prob\,[f^{(1)}(i) \leqslant x] = \frac{1 - \exp[-x(\sum_{j \notin H(i)} g_j)]}{1 - \exp[-(\sum_{j \notin H(i)} g_j)]} \qquad 0 \leqslant x \leqslant 1 \tag{4.78}$$

We may also use the (exact) joint distribution of a few successive interevent times given in [Lawr73,Ito78]. Thus for those values of offered traffic such that an unsuccessful transmission involves only a few transmissions, this refinement is expected to improve the present formulation results. It is also noted in this connection that Ito [Ito77] has derived the interevent time distribution conditioned on the initial and terminal event sources in the superposed renewal processes.

4.5 Conclusion

We have given an approximate analysis for the packet departure processes in a hidden-user environment of single-hop packet broadcasting systems. The channel access protocols considered include pure ALOHA, and unslotted and slotted carrier-sense-multiple-access (CSMA). The effect of (perfect) delay capture on unslotted CSMA has also been evaluated.

Exact stochastic analysis has been given for the durations of a channel idle period, a successful transmission period, and an unsuccessful transmission period consisting only of those packets from the users who can hear the initiating transmission. An approximate analysis has been developed for the duration of an unsuccessful transmission period involving hidden users' packets. Our approximation is based on the theory of superposition of independent renewal processes, together with a proper reduction of transmission start rates to take care of carrier-sense effects.

The channel throughput and the coefficient of variation of the packet interdeparture time calculated by use of our approximation have been compared with the simulation results in the symmetric and wall configurations for a variety of degrees of hiddenness. The agreement between them is excellent in the symmetric hidden-user configurations (without delay capture) for almost the whole range of offered channel traffic and all reasonable values of propagation delay. For wall configurations and symmetric hidden-user configurations with perfect delay capture, the agreement is good until the offered traffic value exceeds its optimum which gives maximal channel throughput.

Lastly we have discussed some rationale for our assumptions and suggested possible refinements of the present model for the dissimilar user case.

The first two moments of packet interdeparture time will be used in Chapter 5 to determine the coefficients in the diffusion process approximation to the queue length distribution at the users.

CHAPTER 5
Queueing Delays in Contention Packet Broadcasting Systems

The average packet delay (including queueing and randomized retransmission delays) for a finite number of random access users of a channel with infinite buffers is studied in this chapter in several ways. First, exact expressions for the average packet delay are given for two identical users of slotted ALOHA and of slotted CSMA-CD (carrier-sense multiple-access with collision detection). Then, for the cases of more than two users of slotted ALOHA, some upper and lower bounds on the mean delay are discussed. Finally, for a general class of contention-type memoryless protocols, a diffusion process approximation for the joint queue length distribution is formulated, and on the basis of its stationary solution, two approximate mean delay formulas are proposed and examined against simulation.

5.1 Introduction

An important performance measure in packet broadcasting communication systems such as ground packet radio networks and local-area computer networks is the average packet delay at a given throughput value. When the channel access protocol falls in the class of random access schemes, this delay versus throughput performance for a finite user population has been studied mainly by use of *linear feedback models*; for example, ALOHA-type schemes are studied in [Lam75], [Carl75] and [Davi80], and carrier-sense-multiple-access (CSMA)-type protocols are studied in [Toba77], [Toba80d], [Hans79] and [Heym82b]. In a linear feedback model, a Markov or semi-Markov process is formulated for a finite population of statistically identical users each being capable of storing at most one backlogged packet. (The system state is usually the number of backlogged packets.) This model may be realistic for a system of users who can actually have at most one outstanding request (like interactive terminals), or a system of so many users that traffic per user must be held at a sufficiently low level in order for the system to be stable.

To analyze the throughput-delay relationship for a group of users with capability of storing more than one packet, some extension of the linear feedback model has been attempted in [Toba80a] and in Chapter 7 of this dissertation. One of the conclusions obtained in these studies is expressed by the phrase in [Toba80a] that the (optimally controlled) system is mostly *channel bound* as opposed to *storage bound*. This statement is drawn from an observation that the improvement (in the throughput-delay performance) brought about by increasing the number of packet buffers from 2 to 3 is not as significant as that by increasing it from 1 to 2. Studying the performance in the case of users each having an infinite buffer motivates us not

only on its own value but also with interest in comparing the difference between the finite-buffer and infinite-buffer cases.

In this chapter, we study the mean packet delay (which includes the queueing and randomized retransmission delays) in a finite population of users each of whom has an independent packet arrival process and an infinite capacity of storing outstanding packets. When the channel access protocol is slotted ALOHA, this problem has been addressed in several papers. For example, Tobagi and Kleinrock [Toba76] showed simulation results. Kleinrock and Yemini [Klei80,Yemi80] developed a Wiener-Hopf technique in the case of two users. Saadawi and Ephremides [Saad81] proposed an iterative approximation method using the notion of user and system Markov chains. Finally, Sidi and Segall [Sidi83] found an explicit expression for the mean delay in the case of two identical users. The present chapter continues these efforts by extending the technique in [Sidi83] and by introducing another approximation method.

The organization of the following sections is as follows. In Section 5.2, we first reproduce the analysis of [Sidi83] for the case of two identical users of slotted ALOHA (since our study is based on this approach), and then carry out a similar analysis for slotted CSMA with collision detection. In Section 5.3, applying the same method to the case of more than 2 users of slotted ALOHA, we derive a formal expression for the mean queue length (containing undetermined constants) from which an explicit upper bound on the mean delay is obtained. Section 5.4 is devoted to the development of a diffusion process approximation to the joint queue length distribution for a finite population of users of one of the contention-type memoryless protocols (defined and analyzed in Chapter 3); this class of protocols (ALOHA and nonpersistent CSMA are included) have independent and identically distributed packet interdeparture times (i.e., the intervals between two consecutive successful transmissions) whose first two moments are used to determine the coefficients in the diffusion equation. Concluding remarks are given in Section 5.5.

Throughout the chapter, we assume a constant packet length whose transmission time is chosen as the unit of time. The notation convention $\bar{x} = 1 - x$ is used for $0 \leqslant x \leqslant 1$, and

$$G_1(z_1,z_2,\cdots,z_M) \triangleq \frac{d}{dz_1} G(z_1,z_2,\cdots,z_M) \tag{5.1}$$

for a function G of M variables $z_1,z_2,\cdots,z_M$.

5.2 Exact Analysis for Two Identical Users

In this section, we give an exact analysis leading to an explicit expression for the mean packet delay in the case of two identical users. Section 5.2.1 is concerned with slotted ALOHA (this portion is essentially a reproduction of the analysis in [Sidi83], but we have simplified the derivation and included the general arrival process), and Section 5.2.2 deals with slotted CSMA with collision detection by the same technique. Each section begins with a definition of protocols and parameters, followed by the stationary equation for the joint generating function of the

queue length distribution. By manipulating this equation, we get the mean queue length from which we obtain the mean packet delay with the aid of Little's result [Litt61].

5.2.1 Slotted ALOHA

In slotted ALOHA, time is slotted with slot size equal to 1 (i.e., the packet transmission time). The start of a packet transmission is synchronized with one of the slot boundaries. Consider two identical users with independent arrival processes and infinite buffers. Let λ and $f(z)$ be the mean and the generating function, respectively, for the number of arrivals at each user in any slot. For example, $f(z) = \exp[\lambda(z-1)]$ for Poisson arrivals, and $f(z) = \lambda z + \bar{\lambda}$ for Bernoulli arrivals. In any slot, each user behaves as follows: if he has at least one packet at the beginning, he transmits one of them with probability p (and does not with probability $1-p$), where $0 < p \leq 1$. The transmission is successful if and only if only one of the users transmits; in such a case, a packet is dequeued at the end of the slot. This protocol is identical with DFT (delay first transmission) defined in [Toba80a]; note that the first transmission of a newly arrived packet (as well as the subsequent retransmissions if any) is delayed by a geometrically distributed number of slots with mean $1/p$.

Let us denote by Q_i $(i=1,2)$ the number of packets stored by user i at any slot boundary (we consider the steady state only); this includes the arrival(s) and excludes the departure if any in the preceding slot. We define the steady-state joint generating function for queue length distribution by

$$G(z_1,z_2) \triangleq \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} Prob[Q_1 = k_1, Q_2 = k_2] z_1^{k_1} z_2^{k_2} \tag{5.2}$$

By using a heuristic shown in [Klei80,Yemi80] (rather than the complicated derivation in [Sidi83]), we can readily write down the equation for $G(z_1,z_2)$:

$$\begin{aligned} G(z_1,z_2) = F(z_1,z_2) \Big\{ & G(0,0) \\ & + (\bar{p} + \frac{p}{z_1}) [G(z_1,0) - G(0,0)] + (\bar{p} + \frac{p}{z_2}) [G(0,z_2) - G(0,0)] \\ & + [1 - 2p\bar{p} + p\bar{p}(\frac{1}{z_1} + \frac{1}{z_2})] [G(z_1,z_2) - G(z_1,0) - G(0,z_2) + G(0,0)] \Big\} \end{aligned} \tag{5.3}$$

where

$$F(z_1,z_2) \triangleq f(z_1) f(z_2) \tag{5.4}$$

(The implication of each term in Eq.(5.3) should be clear; e.g., $G(z_1,0) - G(0,0)$ represents the case where user 1 is nonempty and user 2 is empty, and in such a case the queue size at user 1 decreases by 1 only when he transmits (i.e., with probability p) as shown in the term

$\bar{p} + p/z_1$. Other terms in the braces can be interpreted similarly. The factor $F(z_1,z_2)$ stands for the change in queue size due to the independent arrival process.)

Although we have been unable to solve Eq.(5.3) for $G(z_1,z_2)$, we can obtain the mean queue length $\bar{Q}_1 = \bar{Q}_2 = G_1(1,1)$ as follows (following the approach in [Sidi83]). First, use the condition $G(1,1) = 1$ and symmetry $G(1,0) = G(0,1)$ to get

$$p(1-2p)G(1,0) + p^2G(0,0) = p\bar{p} - \lambda \tag{5.5}$$

Then, from Eq.(5.3), we can express $G_1(1,1)$ and $dG(z,z)/dz\,|_{z=1}$ in terms of $G_1(1,0)$. By observation that $dG(z,z)/dz\,|_{z=1} = 2G_1(1,1)$ due to symmetry, we find

$$G_1(1,0) = \frac{2\lambda - \lambda^2 + f''(1)}{2p} \tag{5.6}$$

$$G_1(1,1) = \frac{f''(1)\bar{p} + 2\lambda\bar{p} + \lambda^2 p - 2\lambda^2}{2(p\bar{p} - \lambda)} \tag{5.7}$$

Hence, by Little's result [Litt61], we get the mean packet delay denoted by D:

$$D = \frac{G_1(1,1)}{\lambda} = 1 + \frac{\bar{p}^2 + \lambda p/2 + f''(1)\bar{p}/(2\lambda)}{p\bar{p} - \lambda} \tag{5.8}$$

Note that $f''(1) = \lambda^2$ for Poisson arrivals and $f''(1) = 0$ for Bernoulli arrivals. Since there are no lost packets, the throughput of this system is equal to the total mean arrival rate 2λ.

5.2.2 Slotted CSMA with Collision Detection

In CSMA we take into account the nonzero propagation delay denoted by a, so that a successful transmission takes up $1+a$ in channel time. We assume the collision detection to be such that an unsuccessful transmission lasts $b+a$, where $a \leq b \leq 1$. Now the time is slotted with slot size equal to $b+a$, and the start of any transmission is synchronized with one of these slot boundaries. At the end of every slot, each user can recognize what has happened in the slot. (The special case of this slotting when $b = a$ appears in [Lam80].) Clearly, an unsuccessful transmission takes up 1 slot and the duration of a channel idle period is also counted by slots. Let us define

$$\tau \triangleq \left\lceil \frac{1-b}{b+a} \right\rceil \tag{5.9}$$

where $\lceil x \rceil$ is the ceiling of x (let $\lceil i \rceil = i$ for an integer i). Then, a successful transmission takes $\tau + 1$ slots. See Figure 5.1 for an illustration of successful and unsuccessful transmission periods. We note that the case without collision detection ($b = 1$) is equivalent to slotted ALOHA with slot size $1+a$.

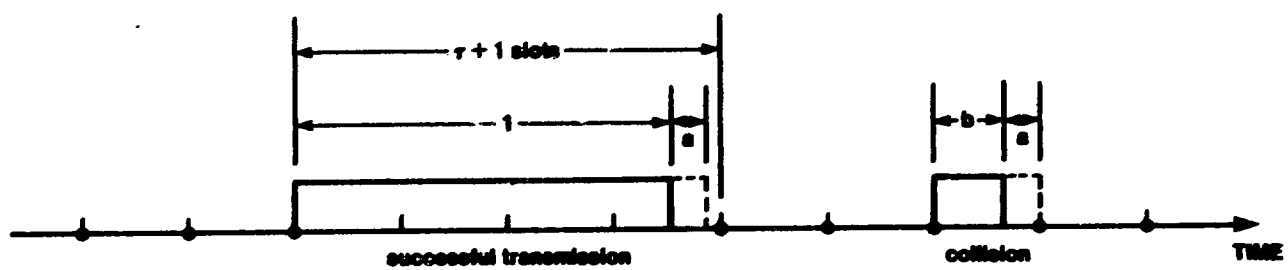


Figure 5.1 An illustration of the channel state in slotted CSMA with collision detection (The embedded Markov epochs are shown by •.)

Consider two identical users with independent arrival processes and infinite buffers. Let λ and $f(z)$ be the mean and the generating function, respectively, for the number of arrivals at each user in any slot (of size $b+a$; note the difference in time units from Section 5.2.1). Suppose that a user has at least one packet at the beginning of a given slot when he is not transmitting. If the preceding slot was sensed busy (due to the other user's transmission), he does not start transmission with probability 1. If the preceding slot was sensed idle (including the case where the preceding slot was an unsuccessful transmission or the last slot of a successful transmission), he starts transmission with probability p (and does not with probability $1-p$), where $0 < p \leq 1$. The simultaneous starts of transmission by both users result in an unsuccessful transmission. Otherwise the transmission will be successful since its start is perceived in the first slot by the other user who then suppresses his transmission.

Numbering the slot boundaries as $t = 1,2,\cdots$, let $Q_i(t)$ $(i=1,2)$ be the number of packets stored at user i at time t; this includes arrival(s) in the slot $[t-1,t]$ and excludes any packet that has successfully completed transmission in the slot $[t-1,t]$. From the above-mentioned arrival process and transmission protocol, it is clear that the process $[Q_1(t),Q_2(t)]$ is a (discrete-time) *semi-Markov process*. We can then construct an *embedded Markov chain* $[Q'_1(t'),Q'_2(t')]$, where $Q'_i(t')$ $(i=1,2)$ is defined to be $Q_i(t')$ when t' is one of those slot boundaries which are not (properly) included in the transmission period. Obviously, these slot boundaries are the embedded Markov epochs in the sense that the process after t' depends only on the state at t'. In Figure 5.1, the Markov epochs are shown by $\bullet$.

Now, we define the stationary joint generating function for the queue length distribution at the Markov epochs by

$$G(z_1,z_2) \triangleq \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} Prob\,[Q'_1=k_1, Q'_2=k_2]\, z_1^{k_1} z_2^{k_2} \tag{5.10}$$

The equation for $G(z_1,z_2)$ is then given by

$$\begin{aligned} G(z_1,z_2) = F(z_1,z_2)\Big\{ & G(0,0) \\ & + [\bar{p} + F(z_1,z_2)^{\tau}\frac{p}{z_1}]\,[G(z_1,0)-G(0,0)] + [\bar{p} + F(z_1,z_2)^{\tau}\frac{p}{z_2}]\,[G(0,z_2)-G(0,0)] \\ & + [1-2p\bar{p} + F(z_1,z_2)^{\tau}p\bar{p}(\frac{1}{z_1}+\frac{1}{z_2})]\,[G(z_1,z_2)-G(z_1,0)-G(0,z_2)+G(0,0)]\Big\} \end{aligned} \tag{5.11}$$

where $F(z_1,z_2)$ is defined in the same way as Eq.(5.4).

We can now find $G_1(1,0)$ and $G_1(1,1)$ as before. First, from $G(1,1)=1$ we have

$$p(1-2p)G(1,0) + p^2 G(0,0) = p\bar{p} - \frac{\lambda}{1-2\lambda\tau} \tag{5.12}$$

Then we get

$$G_1(1,0) = \frac{2\lambda - \lambda^2 + f''(1)}{2p\,(1-2\lambda\tau)} \tag{5.13}$$

$$G_1(1,1) = \frac{(1-2\lambda\tau)[f''(1)\bar{p}+2\lambda\bar{p}+\lambda^2 p] - \lambda[2\lambda - \tau(\lambda^2+f''(1)) - 2\tau(\tau+1)\lambda^2]}{2(1-2\lambda\tau)[p\bar{p}(1-2\lambda\tau)-\lambda]} \tag{5.14}$$

Recall that $G_1(1,1)$ is the mean queue length observed only at the embedded Markov epochs defined above. To find the mean queue length at an arbitrary slot boundary, note that the number of stored packets can be viewed as a 'reward' in the context of the *semi-Markov process with reward* (see, e.g., [Heym82a]). Thus, if we denote by $l(k_1,k_2)$ the expected duration of the state (k_1,k_2), and by $b(k_1,k_2)$ the expected contribution to the backlog accumulation at state (k_1,k_2), then we have the total mean queue length at an arbitrary epoch (or the average reward) as

$$2\bar{Q} = \frac{B}{L} \tag{5.15}$$

where

$$L \triangleq \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} l(k_1,k_2)\, Prob\,[\,Q'_1=k_1, Q'_2=k_2] \tag{5.16}$$

and

$$B \triangleq \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} b(k_1,k_2)\, Prob\,[\,Q'_1=k_1, Q'_2=k_2] \tag{5.17}$$

Since the length between two successive Markov epochs is $\tau+1$ when this interval involves a successful transmission, and it is 1 otherwise (either idle or unsuccessful transmission), we have clearly

$$l(k_1,k_2) = \begin{cases} 1 & k_1=k_2=0 \\ p(\tau+1)+\bar{p} & k_1\geq 1, k_2=0 \text{ or } k_1=0, k_2\geq 1 \\ 2p\bar{p}(\tau+1)+1-2p\bar{p} & k_1\geq 1, k_2\geq 1 \end{cases} \tag{5.18}$$

Similarly, the backlog accumulation is expressed as

$$b(k_1,k_2) = \begin{cases} 2\lambda & k_1=k_2=0 \\ pB(k_1,\tau)+\bar{p}(k_1+2\lambda) & k_1\geq 1, k_2=0 \\ pB(k_2,\tau)+\bar{p}(k_2+2\lambda) & k_1=0, k_2\geq 1 \\ (1-2p\bar{p})\,(k_1+k_2+2\lambda)+2p\bar{p}\,B(k_1+k_2,\tau) & k_1\geq 1, k_2\geq 1 \end{cases} \tag{5.19}$$

where $B(k,\tau)$ is the average accumulated backlog (in packets·slots) during a successful transmission period (of length $\tau+1$ slots) which begins with k packets, and is given by

$$B(k,\tau) = (k-1)(\tau+1)+1\cdot\left(\left\lceil\frac{1+a}{b+a}\right\rceil - \frac{1+a}{b+a} + \tau\right) + \sum_{i=1}^{\tau+1}(2\lambda)i$$

$$= (k-1)(\tau+1) + \lambda(\tau+1)(\tau+2) + \frac{1-b}{b+a} \tag{5.20}$$

Substituting Eqs.(5.18)-(5.20) into Eqs.(5.16) and (5.17), and making use of Eq.(5.12), we have the expressions for L and B reduced to

$$L = \frac{1}{1-2\lambda\tau} \tag{5.21}$$

and

$$B = 2(1+2p\bar{p}\tau)\, G_1(1,1) - 2p(1-2p)\tau\, G_1(1,0) + \frac{2\lambda}{1-2\lambda\tau}\left[\lambda\tau(\tau+1) - \tau + \frac{1-b}{b+a}\right] \tag{5.22}$$

The mean queue length at each user at an arbitrary slot boundary is given by

$$\bar{Q} = (1-2\lambda\tau)\left[(1+2p\bar{p}\tau)\, G_1(1,1) - p(1-2p)\tau\, G_1(1,0)\right] + \lambda^2\tau(\tau+1) - \lambda\left(\tau - \frac{1-b}{b+a}\right) \tag{5.23}$$

Since λ is the mean number of arrivals per user in a slot of length $b+a$, the total throughput of this system is given by $2\lambda/(b+a)$. By Little's result [Litt61], the mean response time D is given by

$$D = \frac{\bar{Q}(b+a)}{\lambda} \tag{5.24}$$

Substitutions of the expression for $G_1(1,0)$ and $G_1(1,1)$ in Eqs.(5.13) and (5.14) into Eq.(5.23) and some manipulation finally yield

$$D = 1 + a + (b+a)\cdot \frac{\bar{p}^2 + \frac{\lambda p}{2} + \bar{p}\,\frac{f''(1)}{2\lambda} + p\bar{p}\tau\left[\frac{f''(1)}{2\lambda} + \lambda\left(\tau + \frac{3}{2}\right)\right]}{p\bar{p}\,(1-2\lambda\tau) - \lambda} \tag{5.25}$$

In Figure 5.2, we show the mean packet delay for given values of throughput, each being optimized with respect to p, in the case of Poisson arrivals. For comparison, we also show the mean response time in a perfect scheduling system (i.e., an M/D/1 queue with arrival rate 2λ and service time $1+a$). This is a plot of

$$D = (1+a)\cdot\frac{1-\lambda(1+a)}{1-2\lambda(1+a)} \tag{5.26}$$

against the total throughput of 2λ. An intersting observation in Figure 5.2 is that D for $b=0.75$ is mostly greater than D for $b=1$. This is because when $a=0.1$ and $b=0.75$, 71% of the second slot in every successful transmission is wasted (note that $(1+a)/(b+a) = 1.294$ is 0.71 short of the next integer 2). The closeness of the curves for $b=0.25$ and $b=0.5$ can be explained similarly.

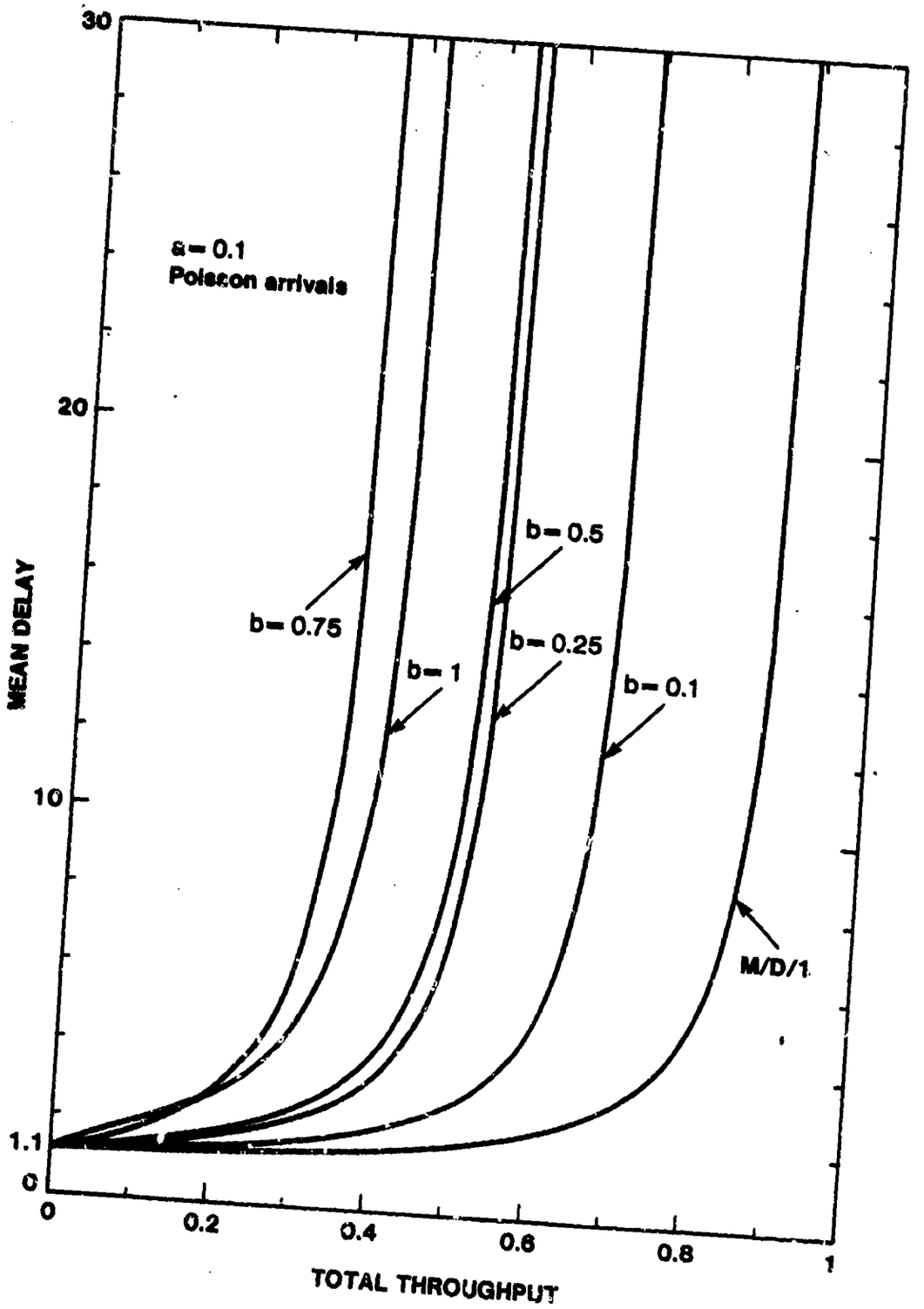


Figure 5.2 Mean delay for two identical users of CSMA with collision detection.

5.3 Bounds on the Mean Delay in Slotted ALOHA

If we apply the technique of the preceding section to the case of more than 2 (identical) users, it turns out that we cannot determine some of the unknown constants which remain in the expression for the mean queue length. From it, however, we can derive an (explicit) upper bound which is tighter than the upper bound obtained by the heavy-traffic assumption. In this section, we show this analysis for slotted ALOHA with Bernoulli arrivals, and also discuss lower bounds which are obtained numerically.

Consider a population of M identical slotted ALOHA users with independent Bernoulli arrival processes and infinite buffers. The assumptions about the arrival process and protocol are the same as in Section 5.2.1, so we use the notations λ and p defined there. Let Q_i denote the number of packets in user i ($i=1,2,\cdots,M$) and define the steady-state joint generating function of the queue length distribution by

$$G(z_1,z_2,\cdots,z_M) = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_M=0}^{\infty} Prob\,[Q_1=k_1,Q_2=k_2,\cdots,Q_M=k_M]\, z_1^{k_1}z_2^{k_2}\cdots z_M^{k_M} \tag{5.27}$$

By considering all the possible events to occur depending on the states of the queues, we can write down the algebraic equation for $G(z_1,z_2,\cdots,z_M)$ in terms of z_i's and those G's for which at least one of their arguments is 0 (this equation is similar to Eq.(5.3)). Then we may eliminate a constant $G(0,0,\cdots,0)$ by using the condition $G(1,1,\cdots,1)=1$. In addition, we have the conditions on the mean queue length $\bar{Q}$ in each user:

$$\bar{Q} \triangleq \frac{d}{dz}G(z,1,1,\cdots,1)\Big|_{z=1} = \frac{1}{2}\frac{d}{dz}G(z,z,1,\cdots,1)\Big|_{z=1} = \cdots = \frac{1}{M}\frac{d}{dz}G(z,z,z,\cdots,z)\Big|_{z=1} \tag{5.28}$$

where we have taken into account the symmetry among the M users. These conditions manifest the observation that the marginal mean length of a single queue is equal to $1/m$ of the sum of the marginal mean lengths of m queues for $m=2,3,\cdots,M$.

In carrying out the calculation, however, it turns out that all the $M-1$ conditions in Eq.(5.28) lead to an identical equation given M. For example, they are

$$\begin{aligned}
M=2: &\quad p\,G_1(1,0) = \lambda-\lambda^2/2\\
M=3: &\quad p\bar{p}\,G_1(1,1,0)+p^2\,G_1(1,0,0) = \lambda-\lambda^2/2\\
M=4: &\quad p\bar{p}^2\,G_1(1,1,1,0)+2p^2\bar{p}\,G_1(1,1,0,0)+p^3\,G_1(1,0,0,0) = \lambda-\lambda^2/2
\end{aligned} \tag{5.29}$$

where G_1 is defined in Eq.(5.1), and symmetry conditions (e.g., $G_1(1,1,0)=G_1(1,0,1)$) have been used to reduce the number of undetermined constants. (The results above and hereafter for $M=2$ are due to Sidi and Segall [Sidi83].) Now, it may be conjectured that, for a general M,

$$\sum_{k=0}^{M-2}\binom{M-2}{k}p^{M-k-1}\bar{p}^k\,G_1(1_{k+1}^{M-1}) = \lambda-\lambda^2/2 \tag{5.30}$$

where 1_k^M is an M-dimensional vector whose first k elements are 1 and the remaining $M-k$ elements are 0. Since we have only one equation (5.29) for $M-1$ unknowns $G_1(1_k^M)$, $k=1,2,\cdots,M-1$, we cannot solve for them except the case $M=2$.

The mean queue length $\bar{Q}=G_1(1_M^M)$, defined by Eq.(5.28), can be expressed in terms of $G_1(1_k^M)$'s as follows:

$$M=2: \quad \bar{Q}=\lambda+\frac{-p^2G_1(1,0)+\lambda(1-p\bar{p})}{p\bar{p}-\lambda}$$

$$M=3: \quad \bar{Q}=\lambda+\frac{-2p^2\bar{p}\,G_1(1,1,0)-p^3G_1(1,0,0)+\lambda(1-p\bar{p}^2)}{p\bar{p}^2-\lambda}$$

$$M=4: \quad \bar{Q}=\lambda+\frac{-3p^2\bar{p}^2G_1(1,1,1,0)-3p^3\bar{p}\,G_1(1,1,0,0)-p^4G_1(1,0,0,0)+\lambda(1-p\bar{p}^3)}{p\bar{p}^3-\lambda} \tag{5.31}$$

Thus, in general, we may induce that

$$\bar{Q}=\lambda+\frac{-\sum_{k=0}^{M-2}\binom{M-1}{k}p^{M-k}\bar{p}^k\,G_1(1_{k+1}^M)+\lambda(1-p\bar{p}^{M-1})}{p\bar{p}^{M-1}-\lambda} \tag{5.32}$$

The mean packet delay D is then given through Little's result [Litt61]:

$$D=\bar{Q}/\lambda \tag{5.33}$$

(Note that for $M=2$ we recover Eq.(5.8) with $f''(1)=0$.) Eqs.(5.30) and (5.32) are the basis for the upper bounds obtained below.

First, an obvious upper bound on $\bar{Q}$ is found by noting that

$$G_1(1_{k+1}^M)\geq 0 \qquad k=0,1,\cdots,M-2 \tag{5.34}$$

Thus, from Eq.(5.32), we have

$$\bar{Q}\leq\lambda+\frac{\lambda(1-p\bar{p}^{M-1})}{p\bar{p}^{M-1}-\lambda} \tag{5.35}$$

which leads to

$$D\leq\frac{1-\lambda}{p\bar{p}^{M-1}-\lambda}\triangleq D_{HT} \tag{5.36}$$

It can be easily seen that this upper bound corresponds to the heavy-traffic approximation, i.e., the solution which is obtained by assuming that all other users are always nonempty. For, in this case, in any slot, we have a successful transmission with probability $p\bar{p}^{M-1}$. Then, the queue length distribution in each user can be found independently (as a birth-and-death process) to give D_{HT} in Eq.(5.36).

Now, a tighter bound can be derived by taking Eq.(5.30) into consideration. Specifically, multiplying the both sides of Eq.(5.30) by p and adding the left-hand side and subtracting the right-hand side to and from the numerator of Eq.(5.32), respectively, we get

$$\bar{Q} = \lambda + \frac{-\sum_{k=1}^{M-2}\binom{M-2}{k-1} p^{M-k}\bar{p}^{k}\, G_1(1_{k+1}^{M}) + \lambda(1-p\bar{p}^{M-1}) - p(\lambda-\lambda^2/2)}{p\bar{p}^{M-1}-\lambda} \tag{5.37}$$

Then, again by use of Eq.(5.34), we obtain

$$D \leqslant 1 + \frac{\bar{p}(1-p\bar{p}^{M-2})+\lambda p/2}{p\bar{p}^{M-1}-\lambda} \triangleq D_U \tag{5.38}$$

Note that $D_U = D$ for $M = 2$. This upper bound may not be very tight as $\lambda \to 0$ because we then expect that $T \simeq 1/p$ due to negligible chances of collision, whereas $D_U \simeq 1/p\bar{p}^{M-2}$. The fact that $\bar{Q}$ and D are finite as long as $\lambda < p\bar{p}^{M-1}$ agrees with the classical result by Abramson [Abra73b] for the throughput of a finite-population slotted-ALOHA model where all the users are assumed to be always nonempty.

As for lower bounds on D, we have not yet obtained any satisfactory expression, although D for $M = 2$ is obviously one of them. Clearly, however, a solution to any finite-buffer system (with the same M) gives a lower bound for our model where an infinite buffer is assumed for each user. In a finite-buffer system, packets which arrive to find the buffer full are lost and not accounted for in $\bar{Q}$ or D. Such a solution is now available only numerically (i.e., we solve a stationary Markov-chain problem). As an extreme, if we assume a single-packet buffer for each user, we are essentially having the linear feedback model for the DFT protocol considered by Tobagi [Toba80a]. Numerical solutions to the cases of more buffers certainly yield tighter lower bounds for each M.

In Figure 5.3, we display the upper and lower bounds described above for the case of $M = 3$ with $p = 0.4$. We have shown the results of numerical calculation for the single-buffer case as well as the 9-buffer case. The latter case should be close to the infinite-buffer system so long as λ is not in the vicinity of its maximum. (In fact, for all the parameter values shown in Figure 5.3, the customer blocking probability in the 9-buffer system is less than 10^{-3}.) The curves labeled with D^* and D^{**} are due to our proposed mean delay formulas derived in the next section. Similar curves for the case of $M = 10$ and $p = 0.1$ are shown in Figure 5.4.

5.4 Diffusion Process Approximation for a Contention System

In this section, we present a diffusion process approximation to the joint queue length distribution in an open contention system. We assume that a population of M users, each having an independent arrival stream of packets, contend for a communication channel which can administer service to one user at a time. If more than one user demands service simultaneously, none of them get a successful service by the channel (the case called collision). In the case of a successful service, one packet is removed from the originating user's queue. The

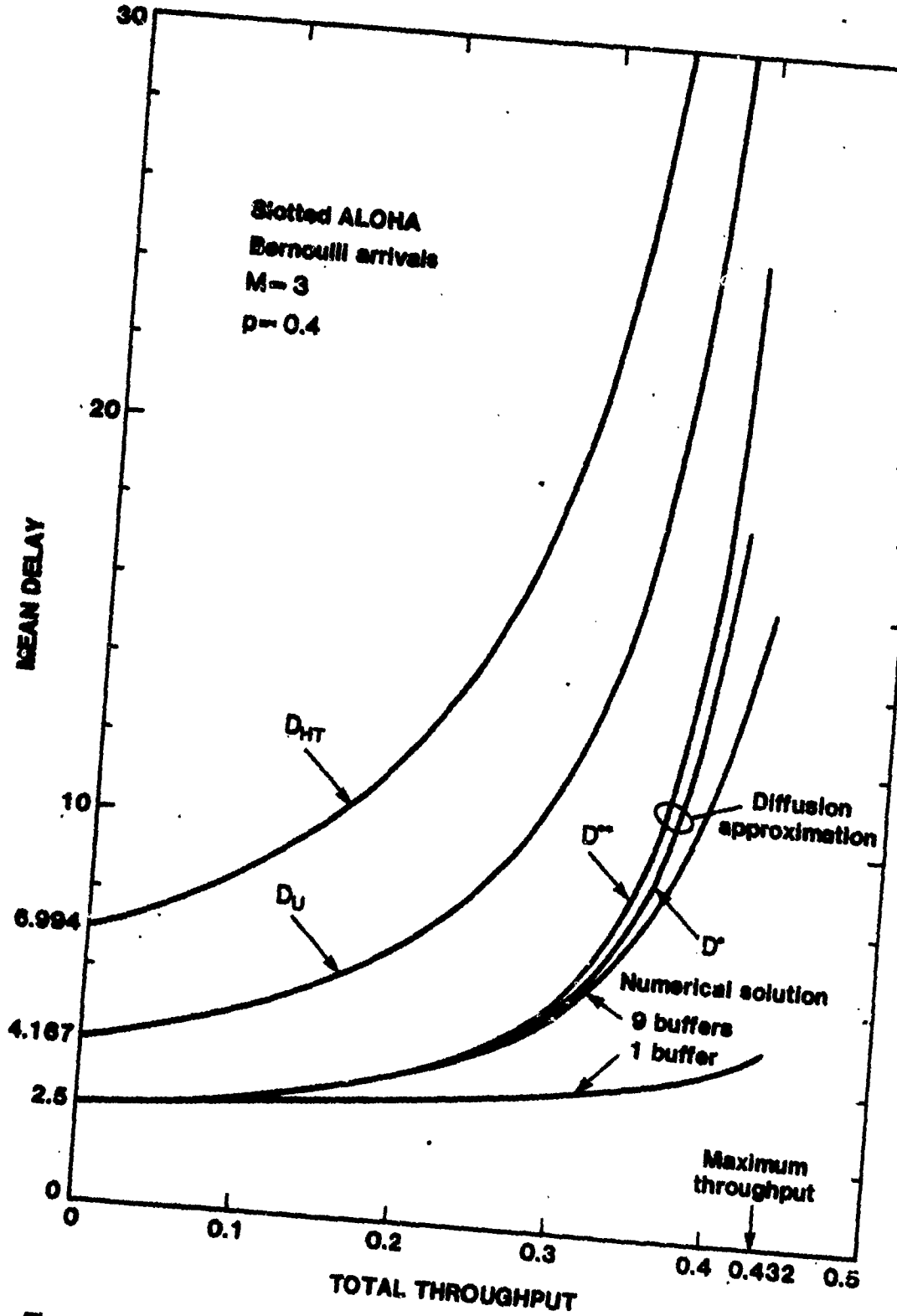


Figure 5.3 Mean delay for a system of 3 identical users of slotted ALOHA.

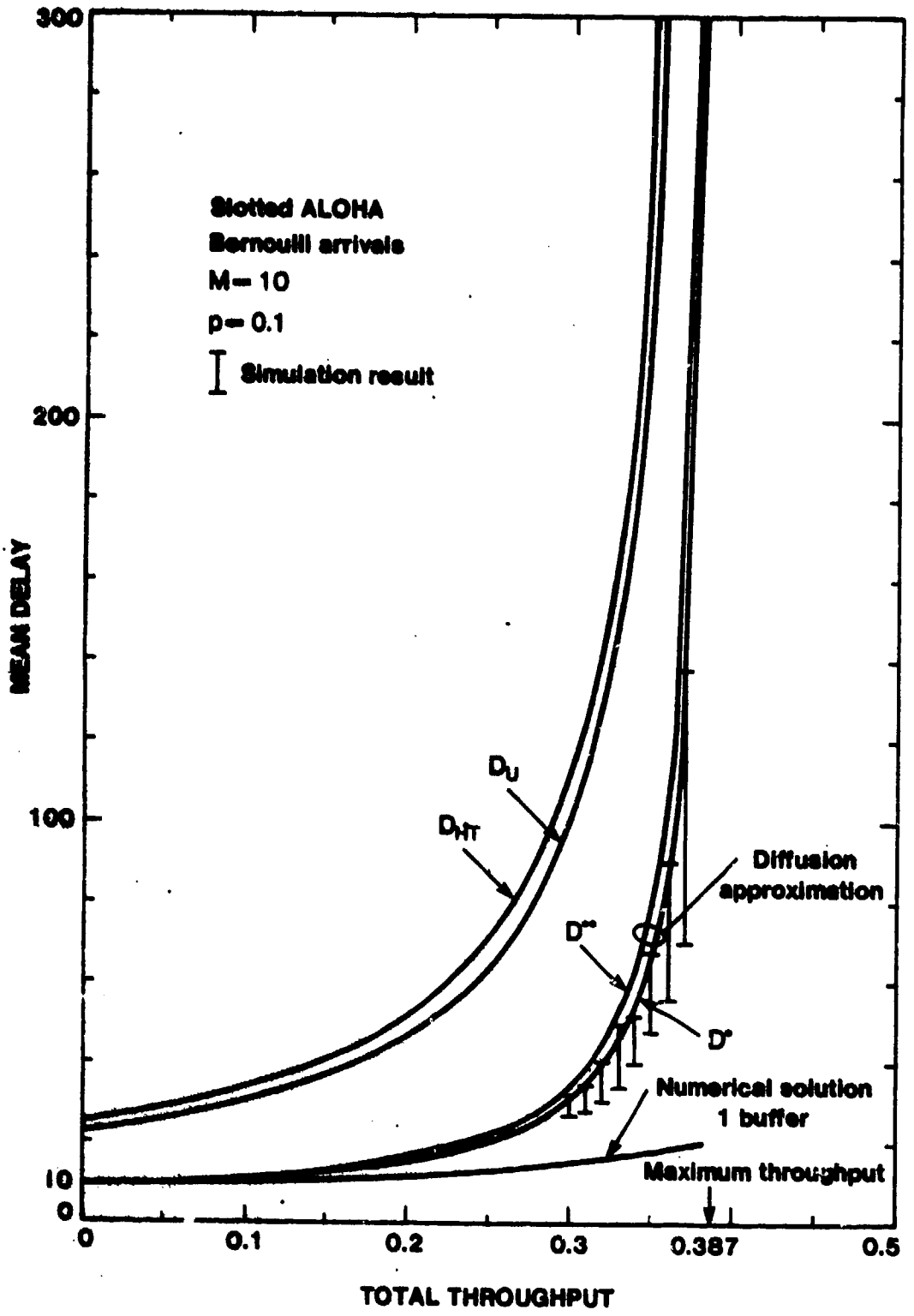


Figure 5.4 Mean delay for a system of 10 identical users of slotted ALOHA (with simulation results).

time of next service request is scheduled independently by each user according to a given randomization procedure. It should be noted that because of contention the queue lengths in all users are dependent on each other. A diffusion process approximation is applied to the joint process of queue lengths in all users.

The time axis of the system may be slotted or unslotted (continuous); its unit is chosen to be the constant packet transmission time. (In a slotted-time system, the slot size may equal this unit time (for ALOHA) or its fraction (for CSMA).) Let the M users be indexed as $1,2,\cdots,M$. Let $1/\lambda_i$ and $C_{a,i}^2$ be the mean and the coefficient of variation, respectively, of the packet interarrival time at user i ($i=1,2,\cdots,M$). Likewise, since there can be no more than one successful service (called departure) in a unit time, let $1/S$ and C^2 be the mean and the coefficient of variation, respectively, of the system interdeparture time. Note that S is equivalent to the channel throughput. Furthermore, we assume that a successful transmission is achieved by user i with probability q_i ($i=1,2,\cdots,M$), where $\sum_{i=1}^{M} q_i = 1$. If we define $1/S_i$ and C_i^2 as the mean and the coefficient of variation, respectively, of the packet interdeparture time from user i, then it can be shown (see Section 3.2 for derivation) that

$$S_i = q_i S \;; \qquad 1 - C_i^2 = q_i (1 - C^2) \qquad i = 1,2,\cdots,M \tag{5.39}$$

We note that S and C^2 have been calculated in Chapters 3 and 4 for a number of contention-type memoryless protocols.

5.4.1 Diffusion Equation for a Contention System

Let us choose the time origin $t=0$ arbitrarily and let $A_i(t)$ be the number of packet arrivals at user i during interval $[0,t]$ ($i=1,2,\cdots,M$). Similarly, let $D_i(t)$ denote the number of departures from user i during the same interval $[0,t]$. Our approximation is based on the assumption that all users are nonempty at all times. It follows that $Q_i(t)$, the number of packets existing in user i at time t, is given by

$$Q_i(t) = Q_i(0) + A_i(t) - D_i(t) \qquad i = 1,2,\cdots,M \tag{5.40}$$

Therefore, the change in $Q_i(t)$ during an interval $[t,t+\Delta]$ is expressed as

$$Q_i(t+\Delta) - Q_i(t) = [A_i(t+\Delta) - A_i(t)] - [D_i(t+\Delta) - D_i(t)]$$

which we write as

$$\Delta Q_i(t) = \Delta A_i(t) - \Delta D_i(t) \qquad i = 1,2,\cdots,M \tag{5.41}$$

We consider an M-dimensional process

$$\Delta \mathbf{Q}(t) = [\Delta Q_1(t), \Delta Q_2(t), \cdots, \Delta Q_M(t)]$$

Note that $A_i(t)$ and $A_j(t)$ are independent for $i \neq j$. We assume that $A_i(t)$ and $D_j(t)$ (possibly $i=j$) are also independent due to the assumption that all users are always nonempty (so that the arrival process does not affect the departure process). Then, it follows from Eq.(5.41)

that

$$\overline{\Delta Q_i(t)} = \overline{\Delta A_i(t)} - \overline{\Delta D_i(t)} ,$$

$$Var[\Delta Q_i(t)] = Var[\Delta A_i(t)] + Var[\Delta D_i(t)] ,$$

$$Cov[\Delta Q_i(t),\Delta Q_j(t)] = Cov[\Delta D_i(t),\Delta D_j(t)] , \qquad i \neq j$$

$$i,j = 1,2,\cdots,M \tag{5.42}$$

We are now to find the quantities on the right-hand sides of Eq.(5.42) by approximation. Our approximation replaces the integer-valued variables $\Delta A_i(t)$ and $\Delta D_i(t)$ by the corresponding continuous-valued Gaussian variables.

If Δ is sufficiently large that we observe many arrivals and departures during $[t,t+\Delta]$, then on the basis of the central limit theorem, we can approximate each $\Delta A_i(t)$ by a Gaussian variable such that

$$\overline{\Delta A_i(t)} = \lambda_i \Delta ; \quad Var[\Delta A_i(t)] = \lambda_i C_{a,i}^2 \Delta \qquad i = 1,2,\cdots,M \tag{5.43}$$

Similarly, the number of departures from all users in $[t,t+\Delta]$

$$\Delta D(t) \triangleq \sum_{i=1}^{M} \Delta D_i(t) \tag{5.44}$$

can be approximated by a Gaussian variable such that

$$\overline{\Delta D(t)} = S\Delta ; \quad Var[\Delta D(t)] = SC^2\Delta \tag{5.45}$$

The M-dimensional process $[\Delta D_1(t),\Delta D_2(t),\cdots,\Delta D_M(t)]$ is also approximated by a multivariate Gaussian process because $\Delta D_i(t)$ and $\Delta D_j(t)$, $i \neq j$, are dependent. It can be readily shown (see Appendix A) that

$$\overline{\Delta D_i(t)} = q_i \overline{\Delta D(t)} ,$$

$$Cov[\Delta D_i(t),\Delta D_j(t)] = q_i q_j \{ Var[\Delta D(t)] - \overline{\Delta D(t)} \} + \delta_{ij} q_i \overline{\Delta D(t)}$$

$$i,j = 1,2,\cdots,M \tag{5.46}$$

where

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Substituting Eqs.(5.39), (5.43), (5.45) and (5.46) into Eq.(5.42), we determine the coefficients for the diffusion equation given shortly:

$$m_i \Delta \triangleq \overline{\Delta Q_i(t)} = (\lambda_i - S_i)\Delta ,$$

$$\sigma_{ij}^2 \Delta \triangleq Cov[\Delta Q_i(t), \Delta Q_j(t)] = [\delta_{ij}(\lambda_i C_{a,i}^2 + S_i) - S_i(1 - C_j^2)]\Delta$$

$$i,j = 1,2,\cdots,M \tag{5.47}$$

Since $\Delta\mathbf{Q}(t)$ has been defined as a linear combination of the two independent multivariate Gaussian processes by Eq.(5.41), it can also be approximated by a multivariate Gaussian process whose means and covariances are given by Eq.(5.47).

Let $p(\mathbf{x};t)$ be the joint probability density function of $\Delta\mathbf{Q}(t)$, where $\mathbf{x} = [x_1, x_2, \cdots, x_M]$. It satisfies the M-dimensional forward diffusion equation

$$\frac{\partial p(\mathbf{x};t)}{\partial t} = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\sigma_{ij}^2 \frac{\partial^2 p(\mathbf{x};t)}{\partial x_i \partial x_j} - \sum_{i=1}^{M} m_i \frac{\partial p(\mathbf{x};t)}{\partial x_i} \tag{5.48}$$

If there is no boundary condition imposed on $\Delta\mathbf{Q}(t)$, then it is an M-dimensional Brownian motion with drift. However, since we have assumed that all users are nonempty at all times (i.e., $\Delta Q_i(t) > 0$), each of M boundaries $x_i = 0$ $(i = 1,2,\cdots,M)$ should act as a reflecting barrier such that no probability mass can collect at $x_i = 0$. The reflecting boundary condition is given by

$$\frac{1}{2}\sum_{j=1}^{M}\sigma_{ij}^2 \frac{\partial p(\mathbf{x};t)}{\partial x_j} - m_i\, p(\mathbf{x};t)\Big|_{x_i=0} = 0 \qquad i = 1,2,\cdots,M \tag{5.49}$$

5.4.2 Stationary Solution to the Diffusion Equation

The stationary solution to Eq.(5.48) where the time derivative is set to zero which satisfies the reflecting boundary condition in Eq.(5.49) is given by

$$p(\mathbf{x}) = \prod_{i=1}^{M}(-\omega_i)\exp(\omega_i x_i) \tag{5.50}$$

Here the column vector $\boldsymbol{\omega} = [\omega_i]^T$ is computed from the column vector $\mathbf{m} = [m_i]^T$ and matrix $\sigma = [\sigma_{ij}^2]$ by

$$\boldsymbol{\omega} = 2\sigma^{-1}\mathbf{m} \tag{5.51}$$

where σ^{-1} is an inverse matrix of σ.

From Eq.(5.47), it can be shown that

$$\det(\sigma) = \left[1 - \sum_{i=1}^{M}\frac{S_i(1 - C_i^2)}{\lambda_i C_{a,i}^2 + S_i}\right]\cdot\prod_{i=1}^{M}(\lambda_i C_{a,i}^2 + S_i) ,$$

$$(\sigma^{-1})_{ij} = \frac{1}{\det(\sigma)} \prod_{\substack{k=1 \\ (k \neq i)}}^{M} (\lambda_k C_{a,k}^2 + S_k) \cdot \left\{ \delta_{ij} \left[1 - \sum_{k=1}^{M} \frac{S_k (1 - C_k^2)}{\lambda_k C_{a,k}^2 + S_k} \right] + \frac{S_i (1 - C_j^2)}{\lambda_j C_{a,j}^2 + S_j} \right\}$$

$$i,j = 1,2,\cdots,M \tag{5.52}$$

Therefore, ω can be easily computed through Eq.(5.51).

Let us now consider the special case of statistically identical users such that

$$\lambda_i = \lambda \; ; \quad C_{a,i}^2 = c_a^2 \; ; \quad q_i = \frac{1}{M}$$

$$S_i = \frac{S}{M} \triangleq s \; ; \quad C_i^2 = 1 - \frac{1}{M}(1 - C^2) \triangleq c^2 \qquad i = 1,2,\cdots,M \tag{5.53}$$

In such a case, from Eq.(5.47) we have

$$m_i = \lambda - s \; ,$$

$$\sigma_{ij}^2 = \delta_{ij} (\lambda c_a^2 + s) - s (1 - c^2) \qquad i,j = 1,2,\cdots,M \tag{5.54}$$

It follows from Eq.(5.54) that

$$\det(\sigma) = (\lambda c_a^2 + s)^{M-1} (\lambda c_a^2 + s C^2) \; ,$$

$$(\sigma^{-1})_{ij} = \frac{(\lambda c_a^2 + s)^{M-2}}{\det(\sigma)} [\delta_{ij} (\lambda c_a^2 + s C^2) + s (1 - c^2)] \qquad i,j = 1,2,\cdots,M \tag{5.55}$$

Therefore, from Eq.(5.51), we obtain

$$\omega_i = \frac{2(\lambda - s)}{\lambda c_a^2 + s C^2} \qquad i = 1,2,\cdots,M \tag{5.56}$$

We note that this is of a similar form to what we meet in the diffusion process approximation to the queue size of a G/G/1 queueing system. The only difference here is the second term of the denominator of Eq.(5.56). The coefficient of variation C^2 for the whole system (instead of c^2) accounts for the interaction among users.

5.4.3 Proposed Mean Delay Formulas

Eq.(5.50) has a product form for the marginal probability density functions. Therefore, the mean queue length in each user can be calculated separately. In the context of the diffusion process approximation, there seem to be several ways to properly evaluate the mean queue length as shown in [Gele80]. A straightforward way is to simply calculate the mean of an exponential distribution in each term of Eq.(5.50). Let $\bar{Q}_i^*$ be the mean queue length in user i calculated in this way:

$$\bar{Q}_i^* = \int_0^\infty x(-\omega_i)\exp(\omega_i x)\,dx = -\frac{1}{\omega_i} \qquad i=1,2,\cdots,M \tag{5.57}$$

Another way is to first discretize the exponential distributions in Eq.(5.50) and then take the average of the resulting discrete distribution. Let us follow [Koba74] which shows this technique. The discretized distribution of the queue length at user i is given by

$$p_i(n) \triangleq Prob\,[Q_i = n]$$

$$= \int_n^{n+1} (-\omega_i)\exp(\omega_i x)\,dx = (1-\rho_i)(\rho_i)^n \qquad n=0,1,2,\cdots \tag{5.58}$$

where

$$\rho_i = \exp(\omega_i) \qquad i=1,2,\cdots,M \tag{5.59}$$

From Eq.(5.58), the average queue length in user i (now denoted by $\bar{Q}_i^{**}$) is given by

$$\bar{Q}_i^{**} = \sum_{n=0}^{\infty} n\,p_i(n) = \frac{\rho_i}{1-\rho_i} \qquad i=1,2,\cdots,M \tag{5.60}$$

As usual, the mean packet delay in user i, D_i, is given as $\bar{Q}_i/\lambda_i$ by Little's result. Corresponding to the two $\bar{Q}_i$'s in Eqs.(5.57) and (5.60), let

$$D_i^* = \bar{Q}_i^*/\lambda_i\,; \qquad D_i^{**} = \bar{Q}_i^{**}/\lambda_i \qquad i=1,2,\cdots,M \tag{5.61}$$

In diffusion approximations to a G/G/1 queue, it is common to adjust the asymptotic value of mean queue length at zero input since we know it is the utilization of the server (i.e., the ratio of the mean arrival rate to the mean service rate) [Koba74,Gele80]. In our contention-type system, however, the queue length in a given user depends on the queue lengths in other users. Therefore, $\bar{Q}_i$ as $\lambda_i \to 0$ cannot be expressed in terms of parameters only for user i.

It is reasonable to assume that if the arrival rates at all users were negligibly small (let us denote this situation by $\lambda \to 0$) then the packets would be delayed only due to the randomized time before the first transmission (and the transmission time). Let us denote by $D_i^{(0)}$ the expected packet delay in user i when $\lambda \to 0$ ($i=1,2,\cdots,M$). They depend on the channel access protocol and transmission parameters. Noting that

$$\omega_i \simeq -\frac{2}{C^2} \quad \text{as} \quad \lambda \to 0 \qquad i=1,2,\cdots,M \tag{5.62}$$

we propose the following two formulas for the mean packet delay in accordance with the two expressions in Eqs.(5.57) and (5.60):

$$D_i^* = D_i^{(0)}\,\frac{1-1/\omega_i}{1+C^2/2} \qquad i=1,2,\cdots,M \tag{5.63}$$

$$D_i^{**} = D_i^{(0)}\,\frac{1-\exp(-2/C^2)}{1-\rho_i} \qquad i=1,2,\cdots,M \tag{5.64}$$

5.4.4 Discussion of the Numerical Examples

Let us look at some example systems for which we can examine the accuracy of our proposed formulas in Eqs.(5.63) and (5.64). We first confine our concern to the cases of statistically identical users (i.e., the same arrival processes and transmission parameter values for all users); these are the cases where the diffusion approximation is expected to work since all the users then tend to saturate in the same fashion. We then look at a case of nonidentical users. Our discussion refers to Figures 5.3 through 5.7. It is noteworthy in these figures that Eqs.(5.63) and (5.64) give similar numerical values despite their different appearance.

Our first and second examples are for slotted ALOHA with Bernoulli arrivals. For a system of M users each with a transmission probability p and an arrival probability λ in any slot, we have

$$S = Mp(1-p)^{M-1} \; ; \quad C^2 = 1 - S \; ; \quad c_a^2 = 1 - \lambda \; ; \quad D^{(0)} = 1/p \tag{5.65}$$

Thus, from Eqs.(5.56), (5.63) and (5.64), we can easily calculate D^* and D^{**}. Our first example is a system of 3 users of slotted ALOHA each with $p = 0.4$. The mean delay values D^* and D^{**} by Eqs.(5.63) and (5.64), respectively, are shown in Figure 5.3. The two upper bounds D_{HT} and D_U given in Section 5.3, and the lower bounds which are numerically obtained by assuming a finite buffer for each user are also displayed. Since the numerical solution for the 9-buffer case is expected to be close to the true values (so long as the imposed throughput is not near the allowable maximum), the accuracy of our formulas seems good.

The second example is a similar system of 10 users of slotted ALOHA each with $p = 0.1$. In Figure 5.4, our approximation D^* and D^{**} are compared with simulation results. We have here simulated 10,000 slots and shown the results by the bars centered at their sample means and with width equal to twice their sample standard deviations. We see that our approximation deviates further from the sample means (in the direction that gives larger values of mean delay than the simulation results) as the throughput is increased.

In the third example shown in Figure 5.5, we consider a system of $M = 5$ users of pure ALOHA each with a Poisson arrival stream. The interval between two successive transmissions at each nonempty user is assumed to be exponentially distributed with mean $1/g$ where $g = 0.1$ in this example. The values of S and C^2 for such a pure ALOHA system with parameters M and $G = gM$ can be approximately calculated by use of a procedure given in Section 4.3.1. In the present case, we have $S = 0.19$ and $C^2 = 0.74$. Because of the exponentially distributed interarrival times at each user, we use $c_a^2 = 1$ in Eq.(5.56). The packet delay at zero input is given by $D^{(0)} = 1 + (1/g) = 11$. The values of our approximate mean delays D^* and D^{**} in Eqs.(5.63) and (5.64) are plotted together with the simulation results (only the sample means are shown by circles) for 2,000 packets (i.e., 2,000 successful transmissions). Here too, our diffusion approximation appears to overestimate the mean delay as the arrival rates are increased.

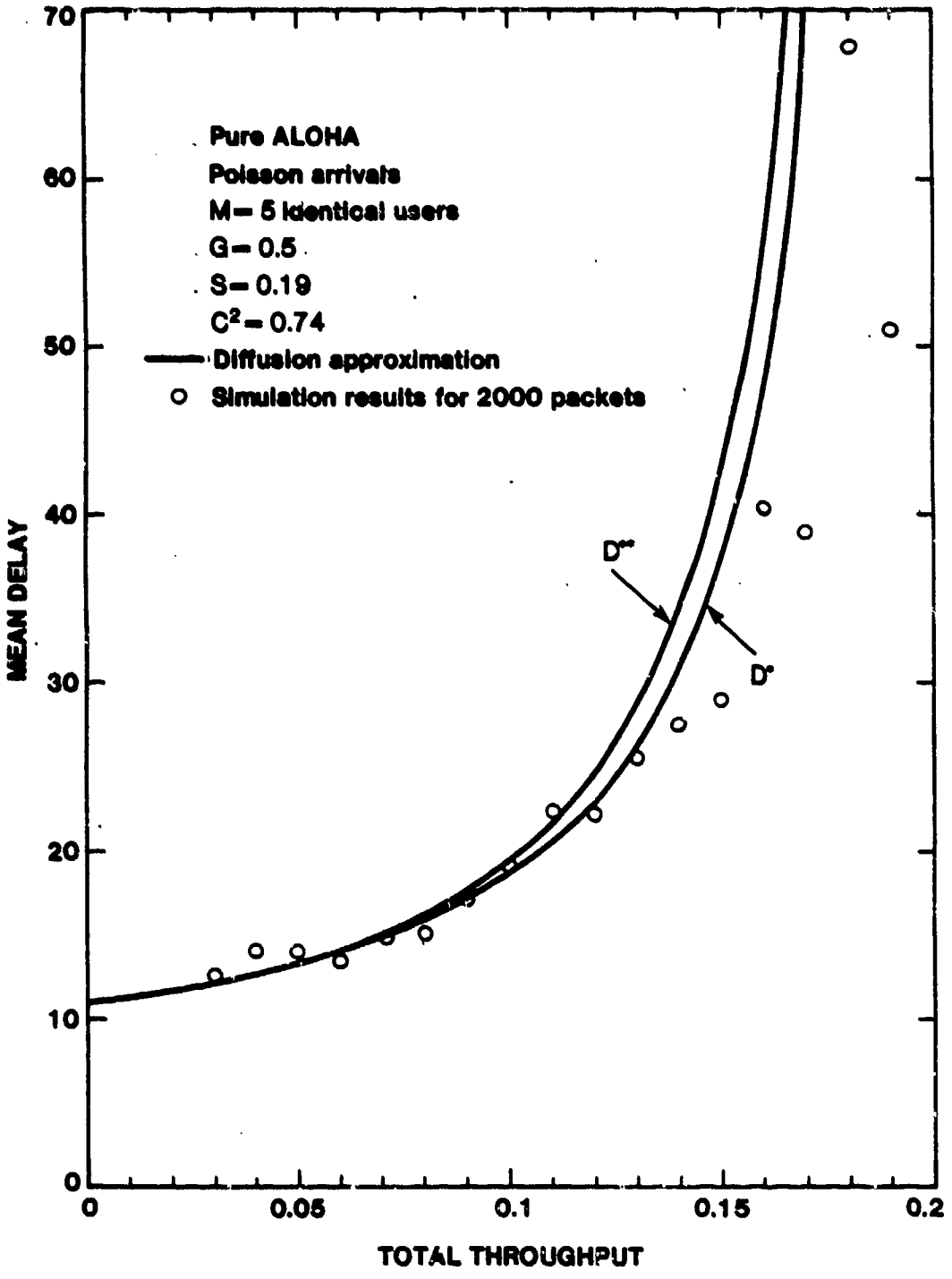


Figure 5.5 Comparison of the diffusion approximation with simulation results in pure ALOHA.

The fourth example deals with a symmetric hidden-user environment for a population of $M=20$ unslotted nonpersistent CSMA users each of whom can hear only $m=17$ other users's transmission. We assume zero propagation delay ($a=0$) and a Poisson arrival process at each user. The time until the next transmission is started by any user (who is not sensing a busy channel) is assumed to be exponentially distributed with mean $1/g$. According to our analysis in Section 4.3.2 (approximate but validated against simulation), for $G=gM=1.778$ which nearly makes S maximum, we have $S=0.45$ and $C^2=0.46$ (see Figures 4.4(a) and 4.4(b)). Again, $D^{(0)}=1+(1/g)=12.25$. In Figure 5.6 we show the values of D^* and D^{**} in Eqs.(5.63) and (5.64) with simulation results (sample means only) for 10,000 packets. The agreement is remarkable.

The last example illustrates a case where the agreement of our proposed formula and the simulation results is not so good for an asymmetric system configuration. Figure 4.3(b) shows a hearing graph (each node represents a user, and an edge is drawn between nodes i and j if users i and j hear each other) for a system of $M=10$ users forming a 'wall configuration.' We assume again unslotted nonpersistent CSMA with zero propagation delay, and Poisson arrivals at each user who schedules his next transmission at an exponentially distributed interval (with mean $1/g$). When $G=gM=1$ is chosen, an analysis in Section 4.3.2 yields approximately $S=0.3943$, $C^2=0.4507$; $q_1=q_{10}=0.0811$, $q_2=q_9=0.0896$, $q_3=q_8=0.0990$, $q_4=q_7=0.1094$, $q_5=q_6=0.1209$. The S_i's and C_i^2's are then calculated by Eq.(5.39). In Figure 5.7, D_i^* ($i=1,3,5$) computed by using these values are plotted together with the corresponding simulation results for 100,000 successful packet transmissions. (The values of D_i^{**} are not so different from those of D_i^*.) We see here that for users 3 and 5 our delay approximation gives lower values than the simulation results. This indicates a limitation for the applicability of our approximation.

5.5 Conclusion

In this chapter, we have studied the mean packet queueing delay for a finite population of random channel-access users with infinite buffers. For the case of two users of slotted ALOHA (due to [Sidi83]) and slotted CSMA with collision detection, the exact expressions for the mean delay have been shown. Then, for the cases of more than two users of slotted ALOHA with Bernoulli arrivals, some upper bounds on the mean delay have been obtained. Finally, a diffusion process approximation has been formulated for the joint queue length distribution, and based on its solution, two approximate mean delay formulas have been proposed.

Our diffusion approximation can be applicable to any single-hop system (including hidden-user configurations) for which we can calculate the first two moments of the distribution of the packet interdeparture times when all the users are assumed to be nonempty. However, the accuracy of this approximation appears to be good only for the case of statistically identical users (i.e., the same arrival process and transmission parameters) since they then saturate in a similar manner. Application of the diffusion process approximation to multi-hop packet radio networks may be possible if the hearing topology and traffic requirements are fairly homogeneous over the whole network.

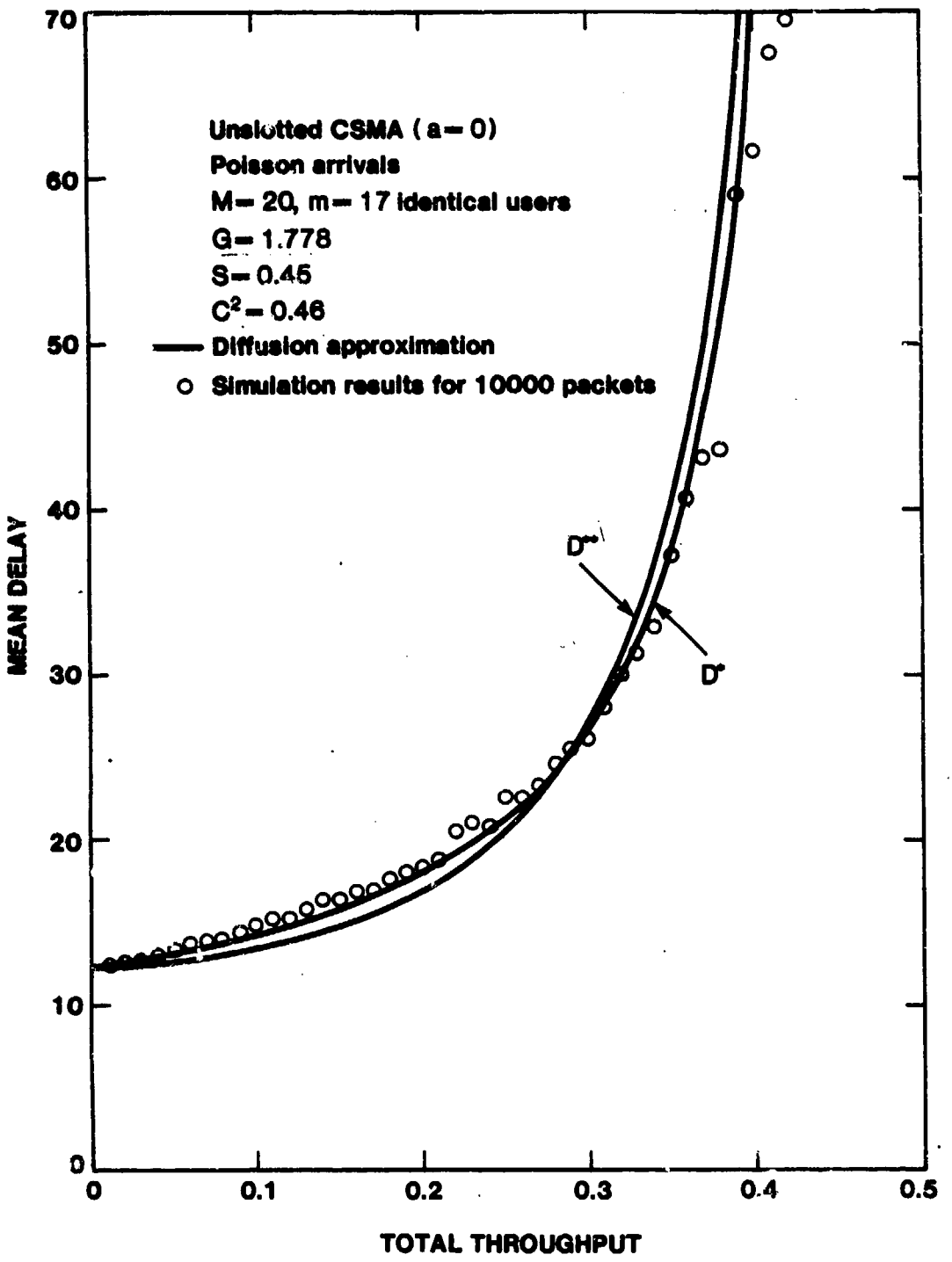


Figure 5.6 Comparison of the diffusion approximation with simulation results in hidden-user, unslotted CSMA.

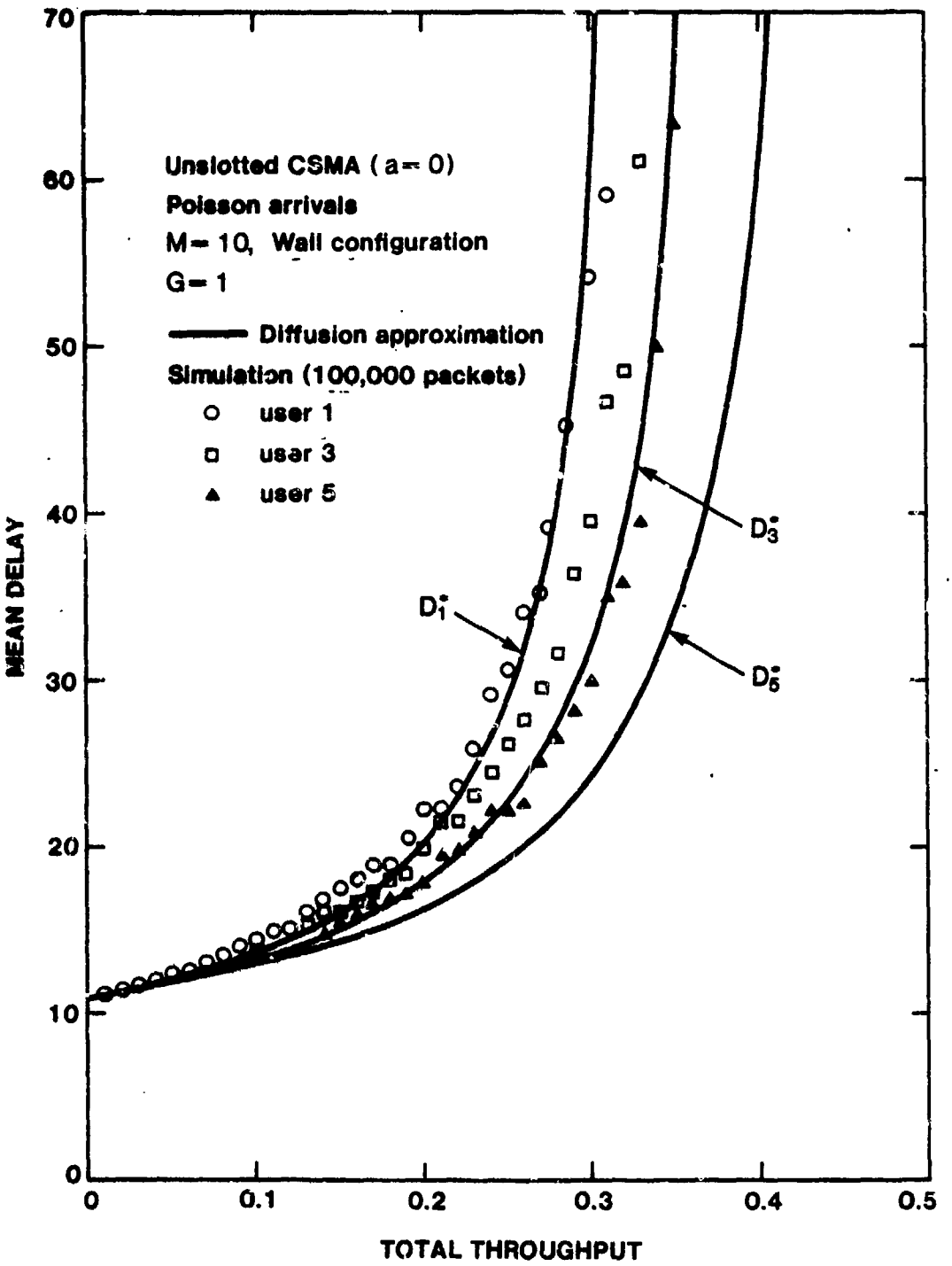


Figure 5.7 Comparison of the diffusion approximation with simulation results in wall-configuration, unslotted CSMA.

CHAPTER 6
Optimal Transmission Ranges for Randomly Distributed Packet Radio Terminals

In multi-hop packet radio networks with randomly distributed terminals, the optimal transmission radii to maximize the expected progress of packets in desired directions are determined with a variety of transmission protocols and network configurations. It is shown that the FM capture phenomenon with slotted ALOHA greatly improves the expected progress over the system without capture due to the more limited area of possibly interfering terminals around the receiver. The (mini)slotted nonpersistent carrier-sense-multiple-access (CSMA) only slightly outperforms ALOHA, unlike the single-hop case (where a large improvement is available), because of a large area of 'hidden' terminals and the long vulnerable period generated by them. As as example of an inhomogeneous terminal distribution, the effect of a gap in an otherwise randomly distributed terminal population on the expected progress of packets crossing the gap is considered. In this case, the disadvantage of using a large transmission radius is demonstrated.

6.1 Introduction

One of the key issues in providing efficient and cost-effective multi-hop packet radio networks is to find an adequate transmission power for each terminal in the network. The environment we have in mind is one in which communicating terminals are geographically distributed, possibly mobile, and require multi-access to a communication channel shared among themselves. It has been shown [Silv80] that the spatial reuse of the channel obtained by reducing the transmission power to a level such that only a few neighbors are within the range gives rise to an improved throughput (the average rate of successful transmissions) for the network. However, since the purpose of transmitting packets in a multi-hop environment is to advance them towards their destinations, a more appropriate measure of performance is the expected one-hop progress of a packet in the desired direction [Klei78,Silv80].

The optimal transmission power to maximize the expected progress involves the following trade-off. (Here we assume every terminal uses the same power.) A short-range transmission is favorable in terms of successful transmission because of its low possibility of collision (the overlapping of packet transmission periods from multiple transmitters) at the receiver. A long-range transmission is favorable because (i) it moves a packet far ahead in one hop if successful, and (ii) there is high probability of finding a candidate receiver in the desired direction. Roughly speaking, if we denote by N the average number of terminals within the transmission radius (N is clearly an increasing function of the radius), then the probability of successful

transmission is proportional to $1/N$ whereas the progress is proportional to $\sqrt{N}$ and the contribution from the receiver's angular position is expressed as a monotonically increasing function of N from 0 to some asymptotic value. Thus we see that there must exist an optimal value of N which maximizes the obtainable expected progress.

This chapter elaborates on these ideas with a variety of transmission protocols and network configurations. The protocols considered here include slotted ALOHA (with and without FM capture) [Robe72, and nonpersistent carrier-sense-multiple-access (CSMA) [Toba74,Klei75c]. Terminals are randomly located in the plane according to a two-dimensional Poisson distribution with homogeneous or inhomogeneous density. Each section below begins with the description of the model used in that section, followed by the formulation of the optimization problem. The optimal transmission range is found, and the performance is compared with other models. The results are summarized in the concluding section.

5.2 Optimum Transmission Radii for Slotted ALOHA

This section is concerned with the optimal transmission radii for randomly distributed terminals using slotted ALOHA as the transmission protocol. The same problem was considered by Kleinrock and Silvester [Klei78,Silv80] who provided the 'magic number' 6 as the optimal number of terminals to be covered by one transmission. However, there appears to be an inconsistency in their treatment. (In evaluating the probability of successful reception (Eq.(6.7) of [Silv80]), the number of terminals around the receiver is confused with that around the transmitter. As a matter of fact, the resultant optimal p, $p^* = 1/N$, could be greater than 1 (inconsistent with slotted ALOHA) for a very small transmission radius.) Therefore, we reconsider their problem and show a different magic number nearly equal to 8. The present section also serves to provide the most basic model among those considered in this chapter.

The basic assumptions and associated parameters used in this section are as follows:

Transmission protocol: slotted ALOHA. The slot length in time is equal to the transmission time of a packet. (All packets are assumed to be of the same length.) The propagation time is neglected (or considered to be included in the slot). We do not take into account the acknowledgment traffic. It is assumed that the successful reception of a packet is immediately made known to the transmitter (e.g., by using a different (free) channel of wide bandwidth).

Transmission probability: p. All terminals are supposed to have packets at all times (heavy-traffic assumption). For every slot, each terminal transmits a packet with probability p (and does not with probability $1-p$), where $0 < p \leqslant 1$.

Transmission radius: R. All terminals use the same transmission radius. This means that terminals within a circle of radius R centered at the transmitter hear the transmission, whereas others do not hear it at all. More than one transmission within a distance R of the receiver in the same slot bring about the collision of all packets at that receiver.

Spatial distribution of terminals: two-dimensional Poisson distribution with the average number of terminals per unit area λ. We assume that a new sample of the spatial distribution is given for every slot.

Distribution of the sources and destinations of packets: two-dimensional isotropic, i.e., uniform over the plane. For every slot, the direction of the final destination for a packet in each terminal is assumed to be distributed uniformly in angle.

Routing strategy: most forward within R (MFR). Each terminal is assumed to know the position of those terminals within s distance R. Given a packet and its final destination, a terminal transmits to the terminal most forward (among those whose positions it knows) in the direction of the final destination. If no terminals are in the forward direction, it transmits to the least backward terminal, if any. (A terminal cannot transmit to itself.) In case there are no terminals in the circle of radius R at all, it does not transmit in that slot. (Note that MFR may not be minimizing the remaining distance to be traveled to the destination. MFR is myopic routing.)

$N \triangleq \lambda\pi R^2$: the average number of terminals within a radius R, and also a measure of connectivity of the network.

In this environment, we have the following two measures of performance:

$S(p,N) \triangleq$ the one-hop throughput, defined as the average number of successful transmissions per slot from a terminal.

$Z(p,N) \triangleq$ the expected progress of a packet in the direction of its final destination per slot from a terminal. The progress x is attained when x is the distance between the transmitter and the receiver projected onto a line drawn towards the final destination, and the transmission to that receiver is successful.

We consider the progress that a given packet makes in the direction towards its final destination for a single (arbitrary) slot only, and do not discuss its behavior along the entire path. Note that $Z(p,N)$ has the dimension of length (e.g., miles). Therefore, $Z(p,N)\sqrt{\lambda}$ may conveniently be used as a dimensionless measure of the expected progress in the number of hopped-over terminals since $1/(2\sqrt{\lambda})$ is the average distance between two nearest terminals. (See Eq.(6.37) below.) We employ $Z(p,N)$ as the objective function for our optimization problem in accordance with the routing strategy MFR. A point in the (p,N) plane which maximizes $Z(p,N)$ is sought. However, the value of $S(p,N)$ at this optimal point is also interst-ing. It will turn out that the same $p = p^*(N)$ maximizes both $S(p,N)$ and $Z(p,N)$.

In order to evaluate $S(p,N)$, we first note that e^{-N} is the probability of there being no terminals within a distance R of the transmitter. In such a case no transmission can occur. Under the condition that there is at least one candidate receiver within R, let A_i be the event that there are i other terminals (excluding the transmitter P and receiver Q) within a distance R of Q. See Figure 6.1. Thanks to the memoryless property of the Poisson distribution, the distribution of the number of other terminals within R does not depend on the existence of P and Q. Thus we have

$$Prob\,[A_i] = \frac{N^i}{i!}e^{-N} \qquad i=0,1,2,\cdots \tag{6.1}$$

The transmission from P to Q is successful (let this event be denoted by $P\rightarrow Q$) if none of the terminals within a distance R of Q transmit (including Q itself). Thus,

$$Prob\,[P\rightarrow Q \mid A_i] = (1-p)^{i+1} \tag{6.2}$$

It follows that

$$S(p,N) = Prob\,[\text{ there is at least one terminal within } R\,]\cdot Prob\,[P \text{ transmits}\,]\cdot Prob\,[P\rightarrow Q]$$

$$= (1-e^{-N})\,p\sum_{i=0}^{\infty} Prob\,[P\rightarrow Q \mid A_i]\,Prob\,[A_i]$$

$$= p(1-p)\,e^{-pN}(1-e^{-N})\,, \tag{6.3}$$

$$\simeq p(1-p)N \quad \text{as } N\rightarrow 0, \qquad \simeq p(1-p)\,e^{-pN} \quad \text{as } N\rightarrow\infty$$

Given N, $S(p,N)$ is maximized by

$$p = p^*(N) \triangleq \frac{2}{N+2+\sqrt{N^2+4}}\,, \tag{6.4}$$

$$\simeq \tfrac{1}{2} \quad \text{as } N\rightarrow 0, \qquad \simeq 1/N \quad \text{as } N\rightarrow\infty$$

The maximum value itself is given by

$$S(p^*(N),N) = \frac{1-e^{-N}}{2+\sqrt{N^2+4}}\exp\left(-\frac{2N}{N+2+\sqrt{N^2+4}}\right), \tag{6.5}$$

$$\simeq (1/4)N \quad \text{as } N\rightarrow 0, \qquad \simeq 1/(Ne) \quad \text{as } N\rightarrow\infty$$

In the case $N\rightarrow 0$, at most only pairs of terminals can hear each other. So, they transmit with probability ½ (the well-known optimal p for two terminals). The optimized results for large N are the same as in [Klei78,Silv80] in which it is stated that they correspond to setting the average traffic load to be equal to one packet per slot within the transmission range. The optimized throughput of N terminals for large N, $1/e$, also conforms to Roberts' result [Robe72] for a

single-hop system of an infinite number of terminals.

We now proceed to find $Z(p,N)$. According to the MFR routing, $\tilde{z}$, the progress of a packet per transmission, is not greater than x if there are no terminals in the area A in Figure 6.2:

$$Prob[\tilde{z} \leqslant x] = e^{-\frac{N}{\pi}q(\frac{x}{R})}, \tag{6.6}$$

where

$$q(t) \triangleq \cos^{-1}(t) - t\sqrt{1-t^2} \tag{6.7}$$

Note that $q(t)$ is the area of A in Figure 6.2 when $R=1$ (unit circle) and $x=t$. Therefore, we have

$$\begin{aligned} Z(p,N) &= Prob[P\text{transmits}] \cdot Prob[P \to Q] \cdot E[\text{progress of a packet}] \\ &= p(1-p)\, e^{-pN} \int_{-R}^{R} x \cdot Prob[x < \tilde{z} \leqslant x+dx] \\ &= p(1-p)\, e^{-pN} \sqrt{\frac{N}{\lambda\pi}} [1 + e^{-N} - \int_{-1}^{1} e^{-\frac{N}{\pi}q(t)}\, dt] \end{aligned} \tag{6.8}$$

Thus, given N, $Z(p,N)$ is also maximized by $p = p^*(N)$ given by Eq.(6.4), and the normalized maximum is given by

$$Z(p^*(N),N)\sqrt{\lambda} = \frac{1}{2+\sqrt{N^2+4}} \exp\left(-\frac{2N}{N+2+\sqrt{N^2+4}}\right) \sqrt{\frac{N}{\pi}} [1 + e^{-N} - \int_{-1}^{1} e^{-\frac{N}{\pi}q(t)}\, dt], \tag{6.9}$$

$$\simeq \frac{16}{45} (\frac{N}{\pi})^{5/2} \text{ as } N \to 0, \quad \simeq \frac{1}{e\sqrt{\pi N}} \text{ as } N \to \infty$$

The functions $p^*(N)$, $S(p^*(N)$, and $Z(p^*(N),N)\cdot\sqrt{\lambda}$ as given by Eqs.(6.4), (6.5), and (6.9), respectively, are plotted in Figure 6.3. $Z(p^*(N),N)\sqrt{\lambda}$ has its maximum value at

$$N = N^* = 7.72$$

Thus, we propose the new magic number 8 as the optimal number of terminals to be covered in the transmission range. In terms of transmission radius, we have

$$R^* = 3.14(1/(2\sqrt{\lambda}))$$

The associated optimal values are

$$p^* \triangleq p^*(N^*) = 0.113, \quad S^* \triangleq S(p^*,N^*) = 0.0419, \quad Z^*\sqrt{\lambda} \triangleq Z(p^*,N^*)\sqrt{\lambda} = 0.0431$$

Therefore, the sketch of optimal transmission is described as follows. Each terminal transmits a packet in every ninth slot on the average ($1/p^* = 8.85$). The probability of success of such a transmission is $S^*/p^* = 0.37$ as slotted ALOHA predicts. It uses a transmission radius to span just about three (3.14) nearest neighbors in linear distance. Then, the expected progress of the

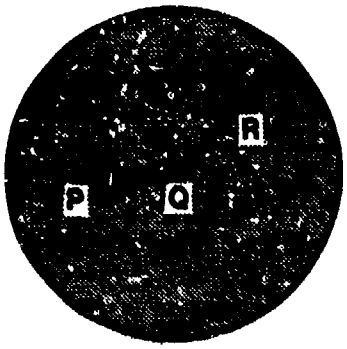


Figure 6.1 The area of interference with the transmission P→Q.

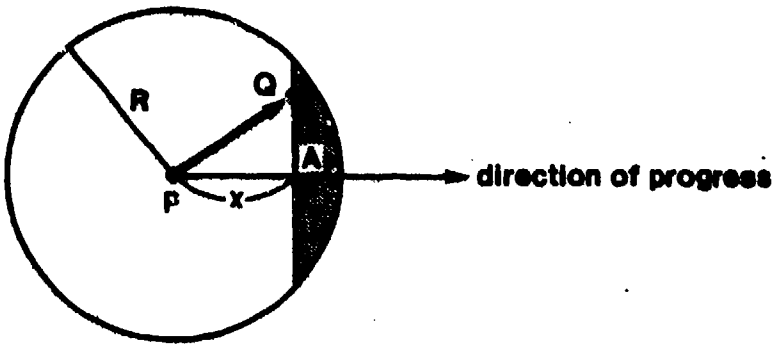


Figure 6.2 The position of the receiver Q .

packet is $Z^*/p^* = 0.76\,(1/(2\sqrt{\lambda})) \simeq \frac{2}{3}(R^*/e)$. Here the factor $1/e$ accounts for the probability of successful transmission, and $\frac{2}{3}R^*$ represents the effective distance that a packet is advanced by a successful transmission with radius R^*.

6.3 Optimal Transmission Radii for ALOHA with Capture

The analysis of the preceding section is here extended to the case of a slotted ALOHA system with FM capture. The observation that the capture phenomenon increases the throughput for a single receiver has been investigated by Roberts [Robe72] and Abramson [Abra77]. Fratta and Sant [Frat80] have shown how capture affects the throughput behavior of an ALOHA network which has multiple transmitters and receivers. They did not use the notion of the transmission radius as we have done in Section 6.2. Their work will be the basis of Section 6.6. In this section, we consider the optimization problem of the expected progress of packets through the MFR (most forward within the transmission radius R) routing in a capture environment. Similar work has been done by Nelson [Nels82] using a different routing strategy. (Specifically, in his routing, one of the (say) k terminals within a half circle (in the forward direction) of radius R is picked as a receivor with probability $1/k$. As a result, his optimized expected progress is somewhat smaller than ours. For example, (using the notations defined below) in the case of perfect capture he gives $Z^*/R^* = 0.0346$ while we give $Z^*/R^* = 0.0393$.)

The basic assumptions and parameters for the model we study here are the same as in Section 6.2, except for the conditions for successful transmission. They include the slotted ALOHA transmission protocol, transmission probability p, transmission radius R, Poisson distribution of terminals with parameter λ, MFR routing, isotropic distribution of source-destination pairs, and $N \triangleq \lambda\pi R^2$.

The concept of FM capture used in this section and Section 6.6 is the same as in the papers cited above. That is, a receiver will correctly receive a packet from a transmitter which is located at a distance r of the receiver, if none of the terminals within a distance αr of the receiver transmit simultaneously. The capture parameter α is related to the capture ratio CR in dB via $CR = 20\log_{10}\alpha$. $1 \leqslant \alpha < \infty$. The case $\alpha = 1$ is called perfect capture, whereas the case $\alpha \to \infty$ corresponds to the system without capture (i.e., the case considered in Section 6.2).

Under these circumstances, we evaluate the throughput, $S(p,N;\alpha)$, and the expected progress, $Z(p,N;\alpha)$. We employ $Z(p,N;\alpha)$ as the objective function of our optimization problem with respect to p and N. (Now the optimum $p^*(N;\alpha)$ for $S(p,N;\alpha)$ is different from that for $Z(p,N;\alpha)$.)

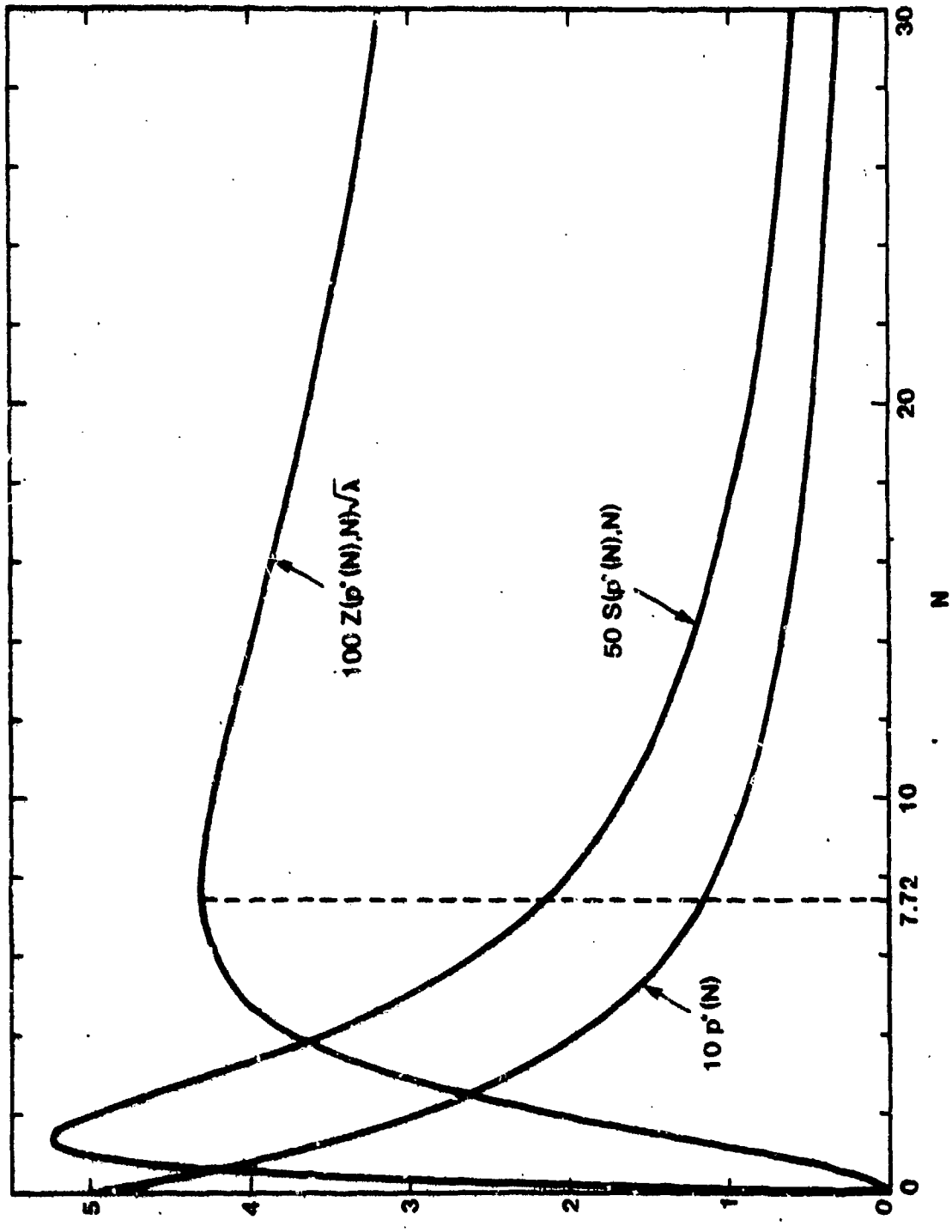


Figure 8.3 The optimal transmission for slotted ALOHA networks without capture

First, we state the conditions for successful transmission of a packet. Since all terminals are using the same transmission radius R, the transmission from the transmitter P to the receiver Q, under the condition that they are a distance r apart, is successful if no other terminals within a distance

$$r' \triangleq \min[\alpha r, R] \tag{6.10}$$

of Q (including Q itself) transmit at the same time [Robe72,Nels82]. Figure 6.4 shows the area of potential interfering terminals for the transmission from P to Q. Thus, unconditioning on the number of terminals in the area as in Section 6.2, we have

$$Prob\,[P \rightarrow Q \mid \tilde{r} = r] = (1-p)e^{-\lambda p \pi r'^2}, \tag{6.11}$$

where $P \rightarrow Q$ represents the event that the transmission from P to Q is successful, and $\tilde{r}$ is the distance between P and Q.

Secondly, we need the expression for the distribution of the positions of the receiver with respect to the transmitter. Let $(\tilde{r},\tilde{\theta})$ be the polar coordinates of the position of the receiver Q, where the origin of the coordinates is at the position of the transmitter P and $\tilde{\theta}$ is measured from the direction in which a packet at P is destined to proceed. See Figure 6.5 for the configuration. Let A be the shaded area. Due to the MFR routing, the receiver is located at (r,θ) if and only if there are no terminals in A and there is a terminal at (r,θ). Therefore,

$$Prob[\,r<\tilde{r}\leqslant r+dr,\ \theta<\tilde{\theta}\leqslant\theta+d\theta\,] = e^{-\lambda R^2(\psi-\sin\psi\cos\psi)}2\lambda r\,dr\,d\theta$$

$$= 2\lambda r e^{-\frac{N}{\pi}q(\frac{r}{R}\cos\theta)}\,dr\,d\theta \qquad 0\leqslant r\leqslant R,\ 0\leqslant\theta\leqslant\pi \tag{6.12}$$

where we have used the relation $r\cos\theta = R\cos\psi$ and the definition of $q(t)$ given in Eq.(6.7). Note that this is an expression for an event similar to the event defined for Eq.(6.6).

Thus, we have the throughput(similar to Eq.(6.3))

$$S(p,N;\alpha) = p\int_0^R\int_0^\pi (1-p)e^{-\lambda\pi r'^2}\,Prob\,[\,r<\tilde{r}\leqslant r+dr,\ \theta<\tilde{\theta}\leqslant\theta+d\theta\,]$$

$$= \frac{2}{\pi}pN(1-p)\int_0^1 te^{-pNt'^2}\,dt\int_0^\pi e^{-\frac{N}{\pi}q(t\cos\theta)}\,d\theta \tag{6.13}$$

where

$$t' = \min[\alpha t,1] \tag{6.14}$$

We see that Eq.(6.13) reduces to Eq.(6.3) when $\alpha\rightarrow\infty$. Also, we have

$$S(p,N;\alpha) \simeq p(1-p)N \qquad \text{for } N \ll 1 \text{ and a moderate value of } \alpha,$$

which is again the same as before.

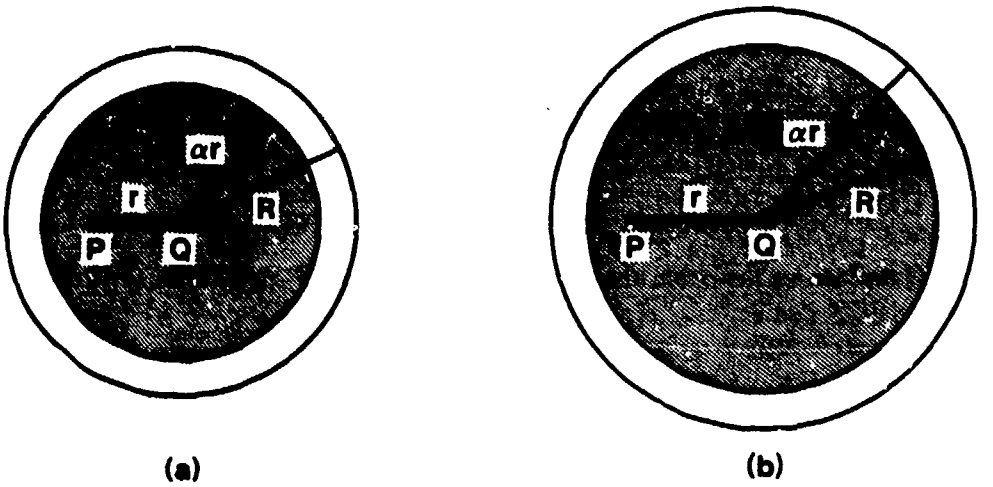


Figure 6.4 The area of terminals possibly interfering the transmission P→Q. (a) $\alpha r < R$. (b) $\alpha r > R$.

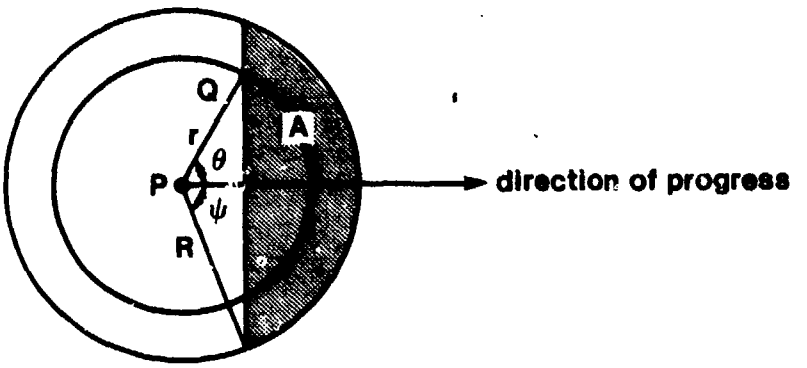


Figure 6.5 The position of the receiver Q with respect to the position of the transmitter P

The expected progress can be obtained similarly as

$$Z(p,N;\alpha) = p\int_0^R\int_0^{\pi}(1-p)e^{-\lambda\pi\rho r^2}r\cos\theta\, Prob\,[r<\tilde{r}\leqslant r+dr,\theta<\tilde{\theta}\leqslant\theta+d\theta]$$

$$= \frac{2}{\pi}pN(1-p)\sqrt{\frac{N}{\lambda\pi}}\int_0^1 t^2e^{-pNt^2}dt\int_0^{\pi}\cos\theta\, e^{-\frac{N}{\pi}q(t\cos\theta)}d\theta \qquad (6.15)$$

The maximum of $Z(p,N;\alpha)\sqrt{\lambda}$ is sought in the (p,N) plane for given α. Let $p^*(N;\alpha)$ be the p that achieves this maximum for given N and α. The optimal point for $\alpha=1$ (perfect capture) is found as follows:

$$N^* = 7.1 \text{ or } R^* = 3.0(1/(2\sqrt{\lambda})), \qquad p^* \triangleq p^*(N^*;1) = 0.17,$$
$$S^* \triangleq S(p^*,N^*;1) = 0.068, \qquad Z^*\sqrt{\lambda} \triangleq Z(p^*,N^*;1)\sqrt{\lambda} = 0.059$$

The expected progress is about 36% better than the system without capture. The optimal transmission is now sketched as follows. Each terminal transmits a packet in every sixth slot on the average $(1/p^* = 5.88)$. The probability of successful transmission is $S^*/p^* = 0.4 > 1/e$. The transmission radius used is three times the average distance between the two nearest neighbors. Then, the expected progress of a packet per transmission is $Z^*/p^* = 0.69(1/(2\sqrt{\lambda}))$.

Figure 6.6 displays the optimal values of parameters N and p, and resulting S^* and $Z^*\sqrt{\lambda}$ for various values of the capture parameter α. From [Robe72], good FM corresponds to $CR = 1.5$ while moderate FM corresponds to $CR = 3.0$ and poor FM corresponds to $CR = 6.0$. As we noted earlier, the results for $\alpha=\infty$ (no capture) coincide with those in Section 6.2. We first notice that $Z^*\sqrt{\lambda}$ with some capture is always greater than that without capture. Thus, a conclusion here is that the FM capture always helps the progress of packets. The reason for this is that we limit the area of interfering terminals within $\min[\alpha r,R]$ which is always no greater than R for the case without capture. This implies a smaller number of interfering terminals, thus giving higher throughput and greater expected progress. It is also interesting that as α increases, N^* first decreases and then increases to reach its final value. This might be explained as follows. For small α, the limitation of a conflicting area by αr is more effective than that by R, so N^* decreases with more conflict as α increases. On the other hand, for large α, the limitation by R is dominant, so N^* approaches the value without capture.

6.4 Optimal Transmission Radii for CSMA

In a single-hop network, another (great) improvement over ALOHA is made possible by CSMA. With this protocol, each terminal utilizes the information about channel status (busy or idle) obtained by listening to the channel. However, the existence of some terminals which are not in line-of-sight of others causes degradation in performance; this is called the hidden-terminal effect [Toba74,Toba75]. If we use the CSMA protocol in a multi-hop network, we expect a similar effect because the hearing range of the receiver is more or less different from the listening range of the transmitter. The purpose of this section is to estimate the effect of hidden terminals associated with CSMA, with the same terminal distribution and with the

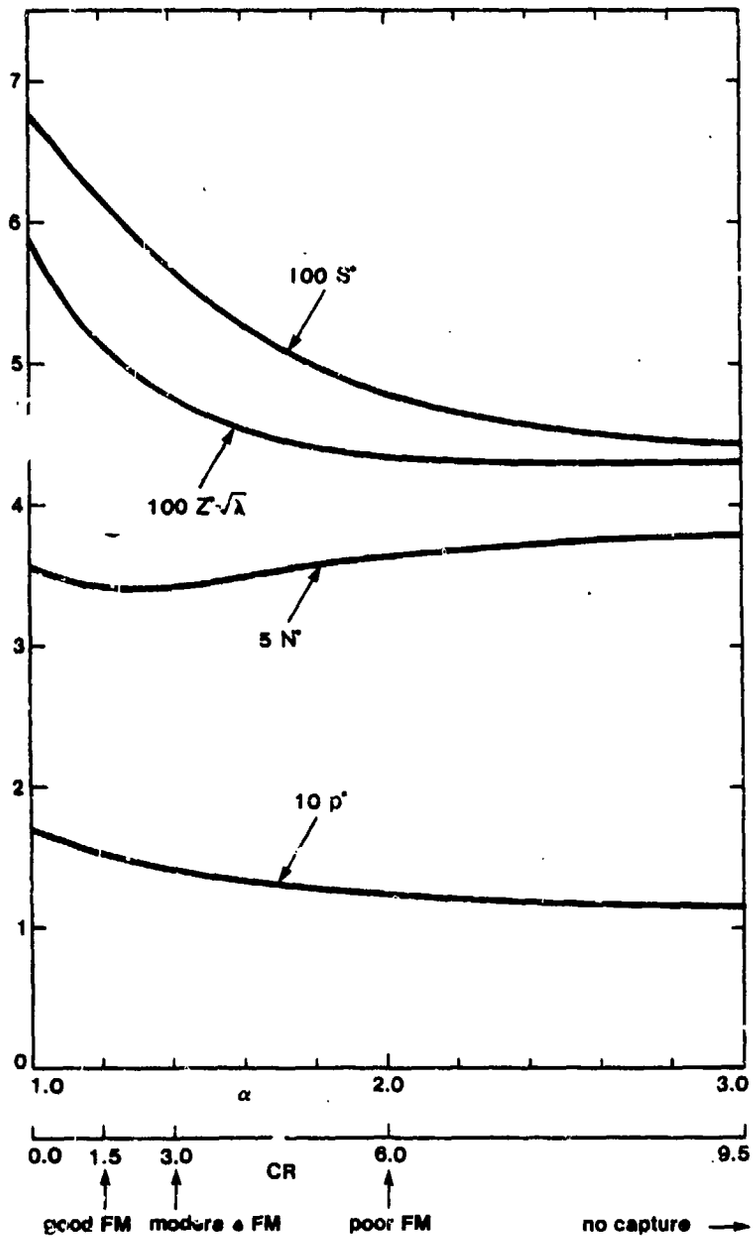


Figure 6.6 The optimal transmissions for slotted ALOHA with capture

same packet routing strategy as in the preceding sections. The basic assumptions and parameters carried over from Section 6.2 include the Poisson distribution of terminals with parameter λ, transmission radius R. MFR (most forward within R) routing, isotropic distribution of source-destination pairs, and $N \triangleq \lambda\pi R^2$.

We now explain the protocol of slotted nonpersistent CSMA. The constant packet transmission time is chosen as the unit of time, and the length of a (mini)slot, denoted by a, accounts for the signal propagation delay. In the derivation below, $\tau \triangleq 1/a$ is assumed to be an integer. (Propagation delay a is used to imply a time interval long enough for all the terminals in the transmission range to recognize the events that occurred time a before.) See Figure 6.7 for the illustration of the channel activity heard at the receiver. We assume that all the terminals within a distance R of the transmitter recognize the transmission in one slot and that they hear the transmission one slot more after the completion of transmission. Assuming that every terminal is ready to transmit at all times, the nonpersistent protocol is described as follows. In every slot, each terminal listens to the channel with probability p (and does not with probability $1-p$). That is, the channel-sensing behavior in a sequence of slots (except during the transmission) at each terminal constitutes independent Bernoulli trials. The parameter p is the sensing rate per slot. If the channel is sensed idle, it begins transmission in the same slot with probability 1. If the channel is sensed busy, it suppresses the transmission, and stops sensing the channel until the end of the current transmission. When the channel becomes idle, the above sensing procedure is repeated.

It is clear that the events whether an actual transmission occurs or not as a result of channel sensing in a sequence of slots at each terminal are no longer independent Bernoulli trials. However, we introduce the assumption that they are. That is, for every slot (except during the transmission), each terminal transmits a packet with probability p' (and does not with probability $1-p'$). A similar assumption was used in [Toba75], and the validity of results obtained was claimed by comparing the throughput values against simulation. The parameter p' is the transmission rate per slot. We leave the determination of p' in terms of p to the Appendix D since we formulate our optimization problem with only p'. Under these conditions, we will evaluate the throughput of transmission $S(p',N;a)$, and the expected progress $Z(p',N;a)$. We employ $Z(p',N;a)$ as the objective function of our optimization problem with respect to p' and N. (Again $S(p',N;a)$ and $Z(p',N;a)$ are not optimized by the same $p'(N)$.)

A particular transmission is successful when no other terminals within a distance R of the receiver transmit during the transmission period $1+a$. Let us consider the conditions for the successful transmission from the transmitter P to receiver Q referring to Figure 6.8. In Figure 6.8(a), the shaded area A and B shows the area of terminals whose transmission may collide with the transmission from P to Q at Q. Since the terminals in area A recognize the transmission in one slot, a collision will be avoided if they do not begin transmission in the same slot. On the other hand, since the transmissions from the terminals in area B occur independently, it is sufficient that they keep silent throughout the entire vulnerable period of length $2+a$ or $2\tau+1$ slots shown in Figure 6.8(b) (the first τ slots are included so as to prevent any interference with the ongoing transmissions and the second $\tau+1$ slots are included not to be interfered with newly started transmissions). (Two packets whose transmissions start

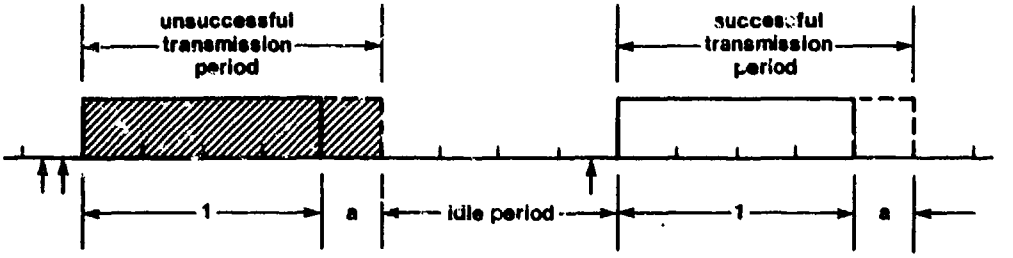


Figure 6.7 Slotted nonpersistent CSMA (transmission and idle periods)

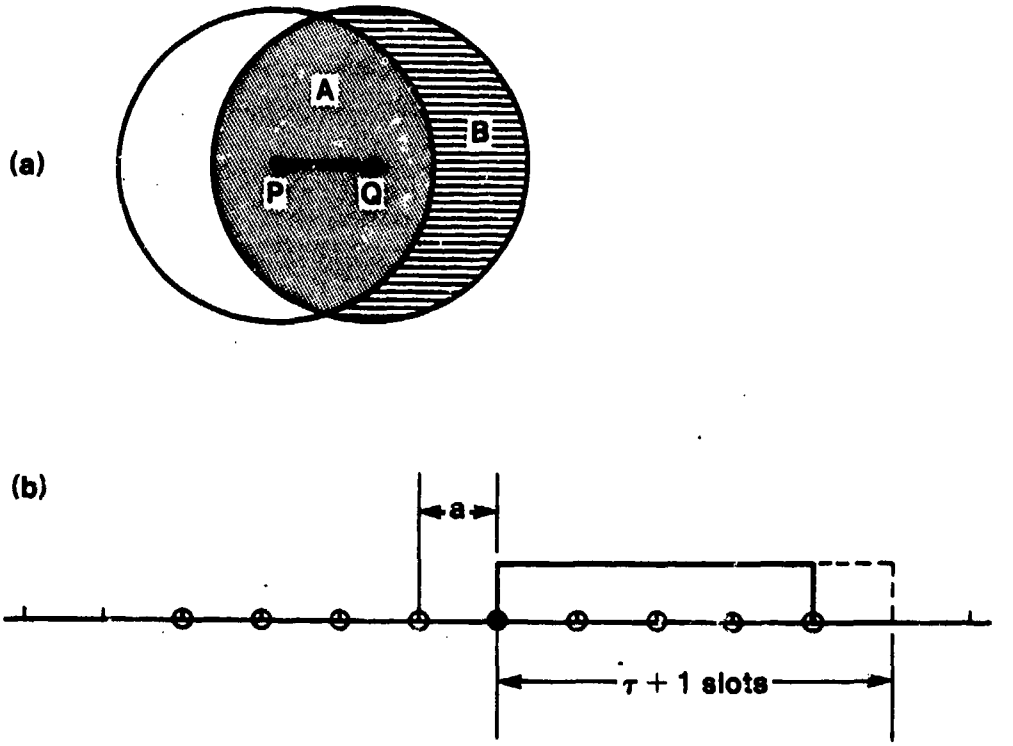


Figure 6.8 The period for the transmission P to Q vulnerable to the transmissions from areas A and B (a) Configuration (b) Time line (● : vulnerable points to A and B, ○ : vulnerable points to B)

with τ slots apart may or may not be received successfully; however, we exclude such a case to pessimistically evaluate the probability of success.) Therefore, if $\tilde{r}$ denotes the distance between P and Q, and $P \rightarrow Q$ denotes the successful transmission from P to Q, then we have

$$\begin{aligned} Prob\,[P \rightarrow Q \mid \tilde{r}=r] &= Prob\,[\,Q \text{ does not start transmission in the same slot}\,] \\ &\quad \cdot Prob\,[\,\text{no transmission from } A \text{ during a slot} \mid \tilde{r}=r\,] \\ &\quad \cdot Prob\,[\,\text{no transmission from } B \text{ during } 2\tau+1 \text{ slots} \mid \tilde{r}=r\,] \end{aligned}$$

Since the area of A is $2R^2q(\frac{r}{2R})$ and the area of B is $\pi R^2 - 2R^2q(\frac{r}{2R})$, we get

$$\begin{aligned} Prob\,[P \rightarrow Q \mid \tilde{r}=r] &= (1-p')\, e^{-p'\lambda \cdot 2R^2 q(\frac{r}{2R})} \cdot e^{-(2\tau+1)p'\lambda \cdot [\pi R^2 - 2R^2 q(\frac{r}{2R})]} \\ &= (1-p')\, e^{-p'N[1+2\tau[1-\frac{2}{\pi}q(\frac{r}{2R})]]}, \end{aligned} \tag{6.16}$$

where $q(t)$ is defined in Eq.(6.7). An assumption involved here is that an independent sample of terminal distributions is given afresh for every slot throughout the vulnerable period from B. Based on this assumption we have evaluated the probability of success in each slot independently.

Since the assumptions about routing are the same as in Section 6.3, the distribution of the position $(\tilde{r},\tilde{\theta})$ of the receiver Q with respect to the transmitter P is given by Eq.(6.12). It follows that the one-hop throughput is given by

$$\begin{aligned} S(p',N;a) &= \frac{p'}{a} \int_0^R \int_0^\pi Prob\,[P \rightarrow Q \mid \tilde{r}=r] \cdot Prob\,[r < \tilde{r} \leqslant r+dr,\ \theta \leqslant \tilde{\theta} < \theta+d\theta] \\ &= \frac{2}{\pi} p'\tau N(1-p')\, e^{-p'(2\tau+1)N} \int_0^1 t e^{\frac{4p'\tau N}{\pi} q(\frac{t}{2})} dt \int_0^\pi e^{-\frac{N}{\pi} q(t\cos\theta)} d\theta, \end{aligned} \tag{6.17}$$

and similarly the expected progress is given by

$$Z(p',N;a) = \frac{2}{\pi} p'\tau N(1-p') \sqrt{\frac{N}{\lambda\pi}}\, e^{-p'(2\tau+1)N} \int_0^1 t^2 e^{\frac{4p'\tau N}{\pi} q(\frac{t}{2})} dt \int_0^\pi e^{-\frac{N}{\pi} q(t\cos\theta)} d\theta \tag{6.18}$$

(We note that the above $S(p',N;a)$ and $Z(p',N;a)$ are not the long-time average values because we have not taken into account the channel activity cycles (idle and busy) whose duration is variable. Thus, Eqs.(6.17) and (6.18) may be viewed as giving the instantaneous values at transmission start times; note that S and Z in slotted ALOHA cases are overall means and the instantaneous values at the same time. Thus the comparison between CSMA and ALOHA is meaningful.)

The maxima of the function $Z(p',N;a)\sqrt{\lambda}$ are determined numerically in the (p',N) plane for various values of a. The optimal point for $a = 0$ (zero propagation delay) is found as follows:

$$N^* = 5.3 \quad \text{or} \quad R^* = 2.6(1/(2\sqrt{\lambda})), \qquad \lim_{a \to 0} \frac{p^*}{a} = \lim_{\tau \to \infty} \tau p^* = 0.20,$$

$$S^* \triangleq S(p^*, N^*; 0) = 0.077, \qquad Z^* \sqrt{\lambda} \triangleq Z(p^*, N^*; 0)\sqrt{\lambda} = 0.050$$

Therefore, the optimized expected progress is only about 16% ($\doteq (0.050-0.0431)\times 100/0.0431$) better than ALOHA system without capture. This small improvement in performance, unlike the single-hop case, appears due to the large area of hidden terminals (about half of the hearable range for $N = N^*$) and the long period (twice as long as the packet transmission time) vulnerable to their transmission.

In Figure 6.9, the optimized expected progress with CSMA is plotted for various values of a, together with those for ALOHA systems without capture and with perfect capture. (For proper comparison, the optimized expected progress with slotted ALOHA should be divided by $1 + a$ to include the propagation time in a slot.) It is seen that the performance of CSMA lies between ALOHA without capture and ALOHA with perfect capture. With reference to Figure 6.6, CSMA's performance turns out to be comparable to that of ALOHA with capture ratio about 1.5 *dB* which corresponds to good FM. The degradation of the expected progress with increasing a is due to the longer vulnerable period.

6.5 Optimal Transmission Radii in an Inhomogeneous Density of Terminals

So far we have considered only the Poisson distribution of terminals with the same spatial density everywhere. However, it is of importance in our multi-hop packet radio studies to extend the analysis to inhomogeneous structures. For example, how should the transmission power be controlled as one passes from a region of low density terminals to higher density terminals and then back out again to lower density terminals; this corresponds to a kind of geographical bottleneck. Another configuration is what we call the 'dumbbell' configuration in which we have high density regions (say, two cities) connected together with an extremely low density region (say, a desert). Here one inquires whether the low density region helps the transmission or not. These are some of the motivations for our study of inhomogeneous configurations of packet radio terminals.

Specifically, the configuration of terminals we consider in this section is a vacant strip of width b in an otherwise Poisson-distributed terminal population with uniform average density λ. Taking the x-axis perpendicular to the gap length, the average density of terminals at x is given by

$$g(x) \triangleq \begin{cases} 0 & 0 \leqslant x \leqslant b \\ \lambda & \text{elsewhere} \end{cases} \tag{6.19}$$

We introduce the 'intensity' of the gap by

$$\beta \triangleq \lambda b^2 \tag{6.20}$$

which is the average number of terminals that would be in the gap of length b if it were not for the gap. The dimensionless quantity β will be used as a characteristic parameter below.

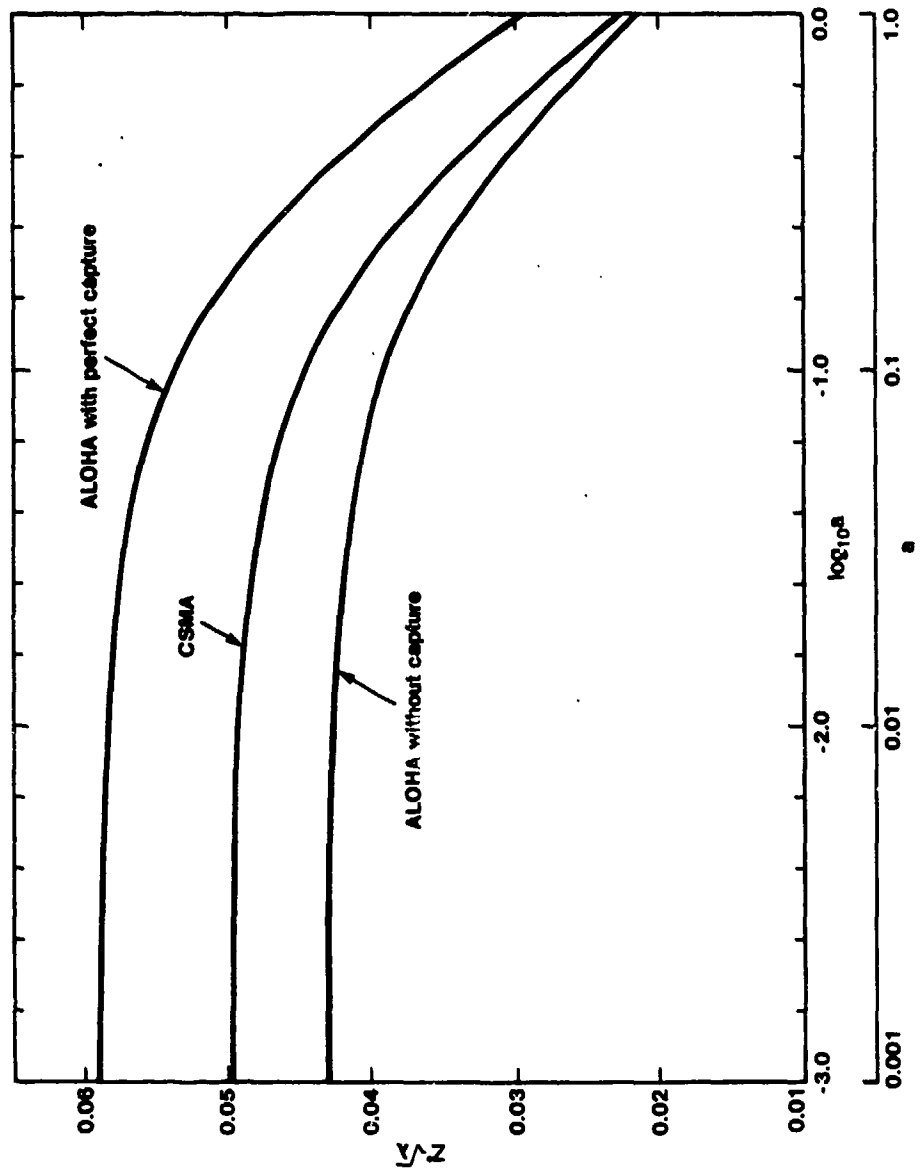


Figure 6.9 Comparison of the optimized expected progress among ALOHA with and without capture, and CSMA networks

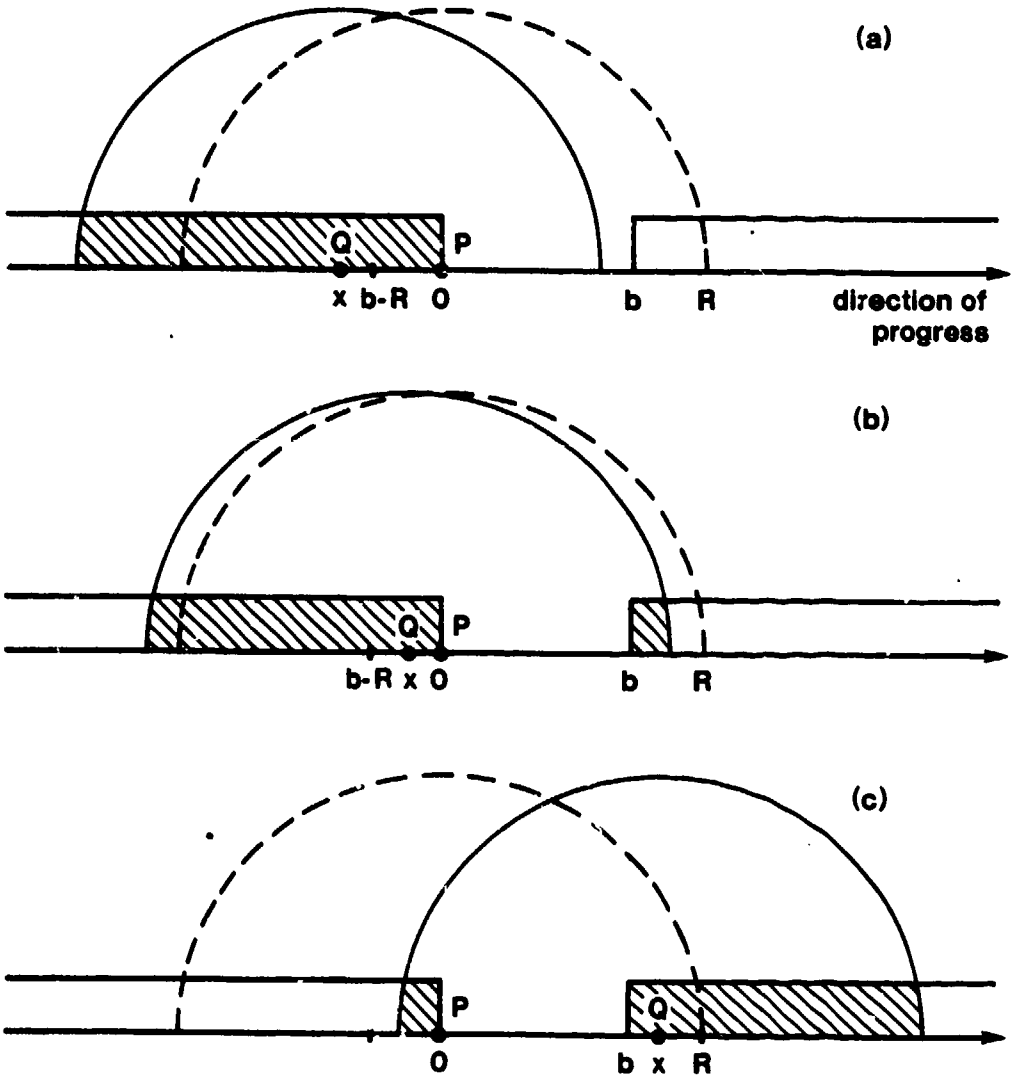


Figure 6.10 Three cases of the position of the receiver Q (P: the transmitter, ////: the area of possibly interfering terminals) (a) $x \leqslant b-R$ (b) $b-R \leqslant x \leqslant 0$ (c) $b \leqslant x \leqslant R$

In the following we evaluate the expected progress of a packet residing at the terminal P on the left bank $(x=0)$ and destined to cross the gap. See Figure 6.10 for the configuration. We assume slotted ALOHA protocol and the transmission radius R $(>b)$ for all terminals. For simplicity, we do not optimize the transmission probability p but will use the value $p=0.113$ which has been found optimal for the case of homogeneous Poisson distribution (see Section 6.2). We recognize the terminals being within a distance R of the receiver as those which may cause conflict with our transmission. Then, our usual procedure yields the probability of successful transmission to the receiver at x as

$$p(1-p)\, e^{-pn(x;R)} \tag{6.21}$$

where $n(x;R)$ is the average number of terminals within a circle of radius R around the receiver at x, given by

$$n(x;R) = \begin{cases} \lambda R^2[\pi - q(-\frac{x}{R})] & -R \leqslant x \leqslant b-R \\ \lambda R^2[q(\frac{b-x}{R}) + \pi - q(-\frac{x}{R})] & b-R \leqslant x \leqslant 0,\ b \leqslant x \leqslant R \end{cases} \tag{6.22}$$

where $q(t)$ is defined in Eq.(6.7). The probability distribution function, $F(x;R)$, of the position of the receiver is given by

$$F(x;R) \triangleq Prob\,[\text{ no terminal in } (x,R) \mid R\,]$$

$$= \begin{cases} \exp\left\{-\lambda R^2[q(\frac{b}{R}) + q(\frac{x}{R}) - \frac{\pi}{2}]\right\} & -R \leqslant x \leqslant 0 \\ \exp\left\{-\lambda R^2 q(\frac{x}{R})\right\} & b \leqslant x \leqslant R \end{cases} \tag{6.23}$$

Using these expressions, the expected progress of our packet is calculated as

$$Z(R) = p(1-p) \int_{-R}^{R} e^{-pn(x;R)} x\, dF(x;R) \tag{6.24}$$

or, in a normalized form,

$$Z(N;\beta)\sqrt{\lambda} = 2p(1-p)(\frac{N}{\pi})^{3/2} \left\{ \int_r^1 t\sqrt{1-t^2} \exp\left\{-\frac{pN}{\pi}[q(r-t) + \pi - q(-t)] - \frac{N}{\pi} q(t)\right\} dt \right.$$

$$+ \int_{r-1}^{0} t\sqrt{1-t^2} \exp\left\{-\frac{pN}{\pi}[q(r-t) + \pi - q(-t)] - \frac{N}{\pi}[q(r) + q(t) - \frac{\pi}{2}]\right\} dt$$

$$\left. + \int_{-1}^{r-1} t\sqrt{1-t^2} \exp\left\{-\frac{pN}{\pi}[\pi - q(-t)] - \frac{N}{\pi}[q(r) + q(t) - \frac{\pi}{2}]\right\} dt \right\}, \tag{6.25}$$

where

$$N = \lambda \pi R^2; \quad r = b/R = \sqrt{\pi\beta/N} \tag{6.26}$$

Figure 6.11 shows the optimal radii N^* and R^* and the expected progress $Z(N;\beta)\sqrt{\lambda}$ for various values of gap intensity β. We see that, for fixed λ

$$N^* \simeq \begin{cases} 7.7 & \text{as} \quad \beta \to 0 \\ \infty & \text{as} \quad \beta \to \infty \end{cases} \tag{6.27}$$

and

$$Z(N^*)\sqrt{\lambda} \simeq \begin{cases} 0.043 & & \text{as } \beta \to 0 \\ 0.050 & \text{(maximum)} & \text{at about } \beta = 1.0 \\ 0 & & \text{as } \beta \to \infty \end{cases} \tag{6.28}$$

The results for narrow gaps reduce to the case with homogeneous density. As the gap width increases the expected progress increases because some of the possibly interfering terminals are removed by the gap. However, for too wide a gap, the transmission radius must be accordingly larger in order to cross it, which causes more conflicts at the receiver; thus the expected progress decreases. It is interesting that the optimized expected progress achieves its maximum at about $\beta = 1$. We can also see that, for fixed b

$$R^*/b \simeq \begin{cases} \infty & \text{as} \quad \beta \to 0 \\ 1 & \text{as} \quad \beta \to \infty \end{cases} \tag{6.29}$$

and

$$Z(R^*)/b \simeq \begin{cases} \infty & \text{as} \quad \beta \to 0 \\ 0 & \text{as} \quad \beta \to \infty \end{cases} \tag{6.30}$$

Therefore, for large λ, the optimal transmission radius is just large enough to reach the other bank. However, the higher possibility of interference with the terminals behind the receiver (i.e., those in the area $x > b$) diminishes the value of $Z(R^*)/b$. For small λ since there is almost no interference, the packet can proceed as far as an arbitrarily large transmission radius.

From Figure 6.11, we see that the existence of the gap such that $\beta < 2$ helps the transmission. Notice, however, that a gap has an effect on performance only when $\beta >> 1$. Thus, a conclusion here is that, to cross the gap we should not use a large transmission power with the same channel; rather we had better use a separate channel (or wire) to avoid possible collisions.

6.6 Optimal Number of Neighbors for ALOHA with Capture

In this section, we extend the model of a partially connected packet radio network with capture proposed by Fratta and Sant [Frat80] to the context of our optimization problem. The reason for doing this is that the packet routing algorithm possibly implemented in each terminal is more suitably handled with their model than with the aforementioned MFR which assumes each terminal knows the position of an indefinite number of terminals within a distance R. However, without a notion of transmission radius, their model has a drawback of having an

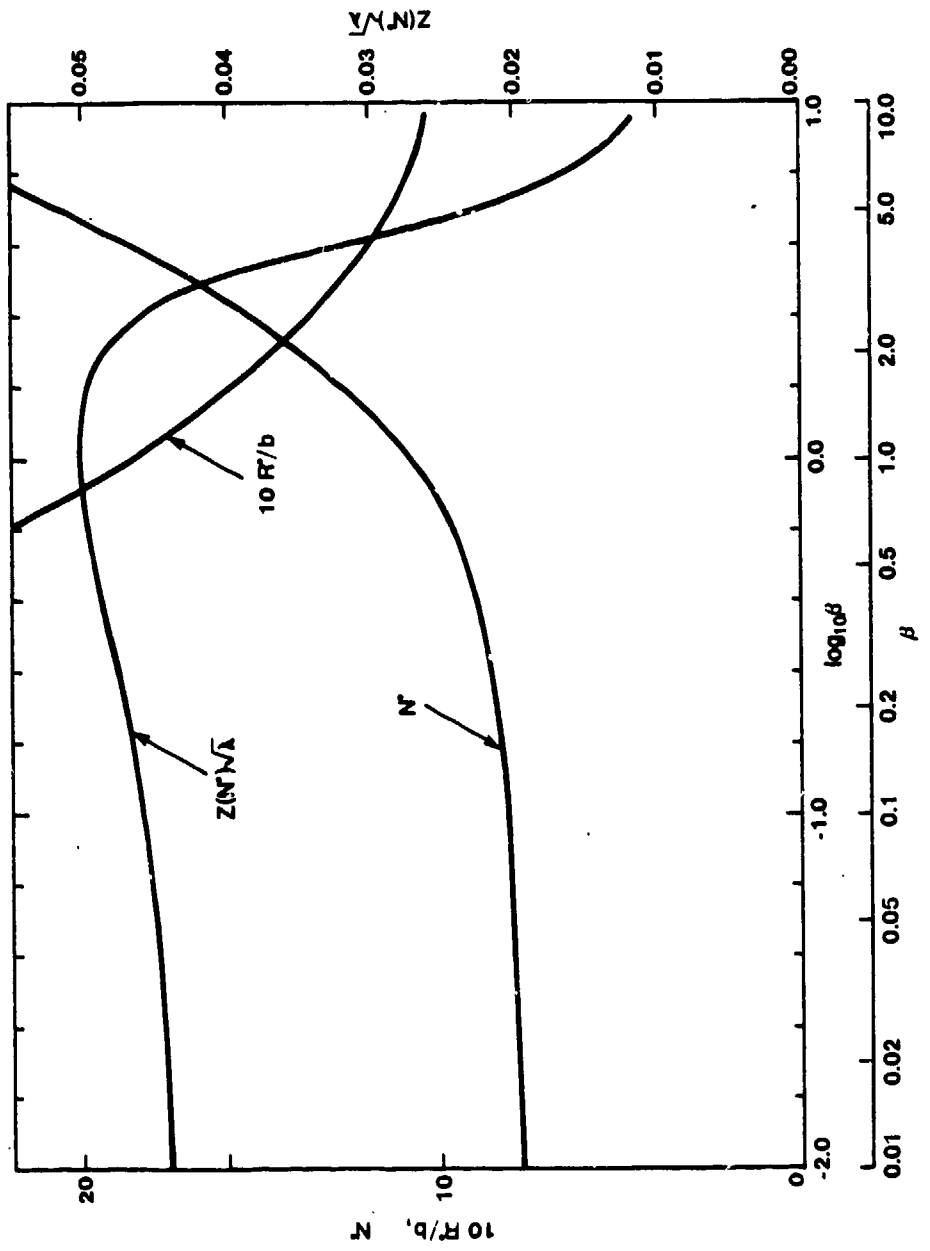


Figure 6.11 The optimal transmission radii and expected progress for packet crossing a gap

unrealistically wide area of interfering terminals for the case of poor capture (large capture ratio). Therefore, the results obtained here should be applicable only to the case of good capture.

The present model assumes a slotted ALOHA transmission protocol with transmission probability p in each slot, a Poisson distribution of terminals with homogeneous density λ, and an isotropic distribution of source-destination pairs. The concept of capture is described in Section 6.2 with capture parameter α. Every terminal is assumed to use the same transmission power. We do not use the notion of transmission radius, which implies that a transmission over a distance r is successful if none of the other terminals within the distance αr of the receiver transmit in the same slot.

We now explain the routing strategy employed here. Each terminal is assumed to know all the positions of its N nearest neighbors. Given a packet and its final destination, a terminal transmits to the most forward terminal in the direction of the final destination among those N neighbors whose positions are known. In case no terminals exist ahead, it transmits to the least backward neighbor. We call this routing as MFN (most forward within N). MFN assumes that each terminal keeps the positions of only a fixed number of terminals, lending itself to easy implementation.

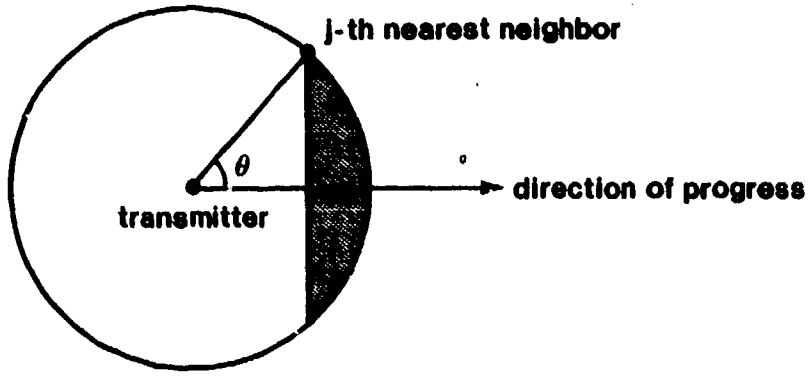


Figure 6.12 The angular position of the j-th nearest neighbor.

The routing algorithm at each terminal is formally stated as follows:

$j \leftarrow N$

L: Consider j nearest neighbors. If the jth nearest one is the most forward, transmit to it. Otherwise, $j \leftarrow j-1$ and go to L.

This algorithm always terminates in at most N cycles. We can evaluate the routing probability

$a_j(N)$ that the jth nearest neighbor is chosen as the receiver when considering N neighbors. To this end, let θ be the angular position of the jth nearest neighbor measured from the direction of progress, as shown in Figure 6.12. Since the jth nearest neighbor is selected as the receiver only when it is the most forward among j neighbors, its probability is given by

$$(1-\frac{1}{\pi}\,|\,\theta-\sin\theta\cos\theta\,|\,)^{j-1} \qquad -\pi \leqslant \theta \leqslant \pi \tag{6.31}$$

(Notice that the distribution of the positions of up to the $j-1$st nearest neighbors is no longer Poisson but uniform since we have specified j.) Unconditioning on θ with the isotropic assumption gives

$c_j \triangleq$ *Prob* [A terminal transmits to the jth nearest neighbor among j of them]

$$= \frac{1}{\pi}\int_0^{\pi} [1-\frac{1}{\pi}(\theta-\sin\theta\cos\theta)]^{j-1}\,d\theta \qquad j=1,2,\cdots,N \tag{6.32}$$

Using the above definition of c_j's, we may finally write the routing probability as

$$a_j(N) = c_j \prod_{k=j+1}^{N} (1-c_k) \qquad j=1,2,\cdots,N \tag{6.33}$$

Clearly

$$\sum_{j=1}^{N} a_j(N) = 1 \tag{6.34}$$

In Table 6.1, we show some values of $a_j(N)$.

Table 6.1 The routing probabilities $a_j(N)$

j	$N=1$	2	3	4	5	6	7	8	9	10
1	1.0000	.5000	.3017	.1976	.1359	.0966	.0704	.0523	.0395	.0302
2		.5000	.3017	.1976	.1359	.0966	.0704	.0523	.0395	.0302
3			.3967	.2598	.1787	.1271	.0926	.0688	.0519	.0397
4				.3450	.2373	.1687	.1230	.0913	.0689	.0527
5					.3122	.2220	.1618	.1202	.0907	.0693
6						.2890	.2106	.1564	.1180	.0902
7							.2712	.2015	.1520	.1162
8								.2571	.1940	.1483
9									.2455	.1876
10										.2356

From this point, we follow the derivation in [Frat80]. First, the probability density function of the distance r_j to the jth nearest neighbor is given by

$$P_{r_j}(r) = \frac{2(\lambda\pi r^2)^{j-1}}{r(j-1)!}\, e^{-\lambda\pi r^2} \qquad r > 0 \tag{6.35}$$

It follows that the mean distance to the jth nearest neighbor is

$$E[r_j] = \frac{(2j-1)!!}{2\sqrt{\lambda}(2j-2)!!} \tag{6.36}$$

where $(2j-1)!! = (2j-1)\cdot(2j-3)\cdots 3\cdot 1$ and $(2j)!! = (2j)\cdot(2j-2)\cdots 4\cdot 2$. Particularly, the mean distance between the two nearest neighbors is given by

$$E[r_1] = 1/(2\sqrt{\lambda}) \tag{6.37}$$

which (without the factor ½) we have used extensively to normalize the expected progress in the preceding sections.

Next, let S_j be the event that a packet transmitted to the jth nearest neighbor is successfully received. As shown in [Frat80],

$$Prob[S_j \mid r_j = r] = (1-p)(1-pq)^{j-1} e^{-\lambda p\pi r^2(\alpha^2-q)} \tag{6.38}$$

where

$$q \triangleq \begin{cases} \dfrac{2}{3} - \dfrac{\sqrt{3}}{2\pi} \doteq 0.391 & \alpha = 1 \\ \dfrac{1}{\pi}\left\{\pi + (\alpha^2-2)\cos^{-1}\left(\dfrac{\alpha}{2}\right) - \alpha\sqrt{1-\dfrac{1}{4}\alpha^2}\right\} & 1 \leqslant \alpha \leqslant 2 \\ 1 & \alpha \geqslant 2 \end{cases} \tag{6.39}$$

Unconditioning on r_j with Eq.(6.35) gives

$$Prob[S_j] = \frac{(1-p)(1-pq)^{j-1}}{(1+\alpha^2 p - pq)^j} \tag{6.40}$$

It follows that the throughput of transmission is given by

$$S(p,N;\alpha) = p\sum_{j=1}^{N} a_j(N)\, Prob[S_j] \tag{6.41}$$

This completes our quotation from [Frat80]. Similarly, we obtain the expected progress as

$$Z(p,N;\alpha) = \frac{p}{\sqrt{1+\alpha^2 p - pq}} \sum_{j=1}^{N} b_j(N)\, E[r_j]\, Prob[S_j] \tag{6.42}$$

where

$$b_j(N) \triangleq \prod_{k=j+1}^{N} (1-c_k) \cdot \frac{1}{\pi}\int_0^{\pi} \cos\theta\left[1 - \frac{1}{\pi}(\theta - \sin\theta\cos\theta)\right]^{j-1} d\theta \qquad j = 1,2,\cdots,N \tag{6.43}$$

Our optimization problem is to find the maximum of $Z(p,N;\alpha)\sqrt{\lambda}$ in the (p,N) plane. In Figure 6.13, the optimal values of N, N^*, and the maximum values of $Z(p,N;\alpha)\sqrt{\lambda}$, $Z^*\sqrt{\lambda}$, are plotted for various values of capture parameter α. It can be seen that the expected progress decreases rapidly as α increases, which does not agree with the result shown in Figure 6.6. This comes from the present assumption that there is no fixed transmission radius. However, for small α, the results that $N^* = 7$ and $Z^*\sqrt{\lambda} \simeq 0.05$ agree with the previous results. Therefore, we may conclude that 7 is suitable for the number of known terminals when the MFN routing is adopted.

6.7 Conclusion

We have solved for the maximum expected progress per hop (Z), provided by the optimal transmission probability (p) and transmission radius (expressed in terms of the number of terminals in the range, N), in some models of randomly distributed packet radio terminals (with average density λ) under the assumption of heavy traffic (all terminals always have ready packets). The quantity $Z\sqrt{\lambda}$ has been used consistently as the dimensionless objective function for optimization problems with respect to p and N. Major conclusions about the performance of each model are as follows:

The optimal transmission with slotted ALOHA without capture is attained by $N=7.72$ and $p=0.113$ which gives $Z\sqrt{\lambda}=0.0431$. Therefore, each terminal transmits once in every nine slots on the average with the transmission radius covering just about eight nearest neighbors in the direction of packet's final destination. The probability of success of such a transmission is nearly equal to $1/e$. The expected progress per transmission is about two thirds of R/e, where R is the optimal transmission radius ($N=\lambda\pi R^2$).

FM capture improves the performance of slotted ALOHA systems due to the more limited area of possibly interfering terminals around the receiver. The expected progress in a system with perfect capture (optimized with $N=7.1$ and $p=0.17$) is about 36% greater than that in the system without capture. The probability of successful transmission is also higher than $1/e$. A model which is more amenable to implementation (each terminal knows the positions of only a fixed number of its neighbors) has shown similar results.

The slotted nonpersistent CSMA provides a nominal improvement in performance over the ALOHA system (16% improvement in the optimized expected progress for the zero propagation delay), which is not as large an improvement as we have obtained in the single-hop case. The reason for this is the large area of 'hidden' terminals (about half of the interfering area) which cannot hear the transmission, and the long vulnerable period (twice as long as the packet transmission time) due to those terminals. The performance of (slotted nonpersistent) CSMA is comparable to that of ALOHA with good FM capture (capture ratio about 1.5 dB). The degradation occurs as the ratio of propagation delay to the transmission time increases.

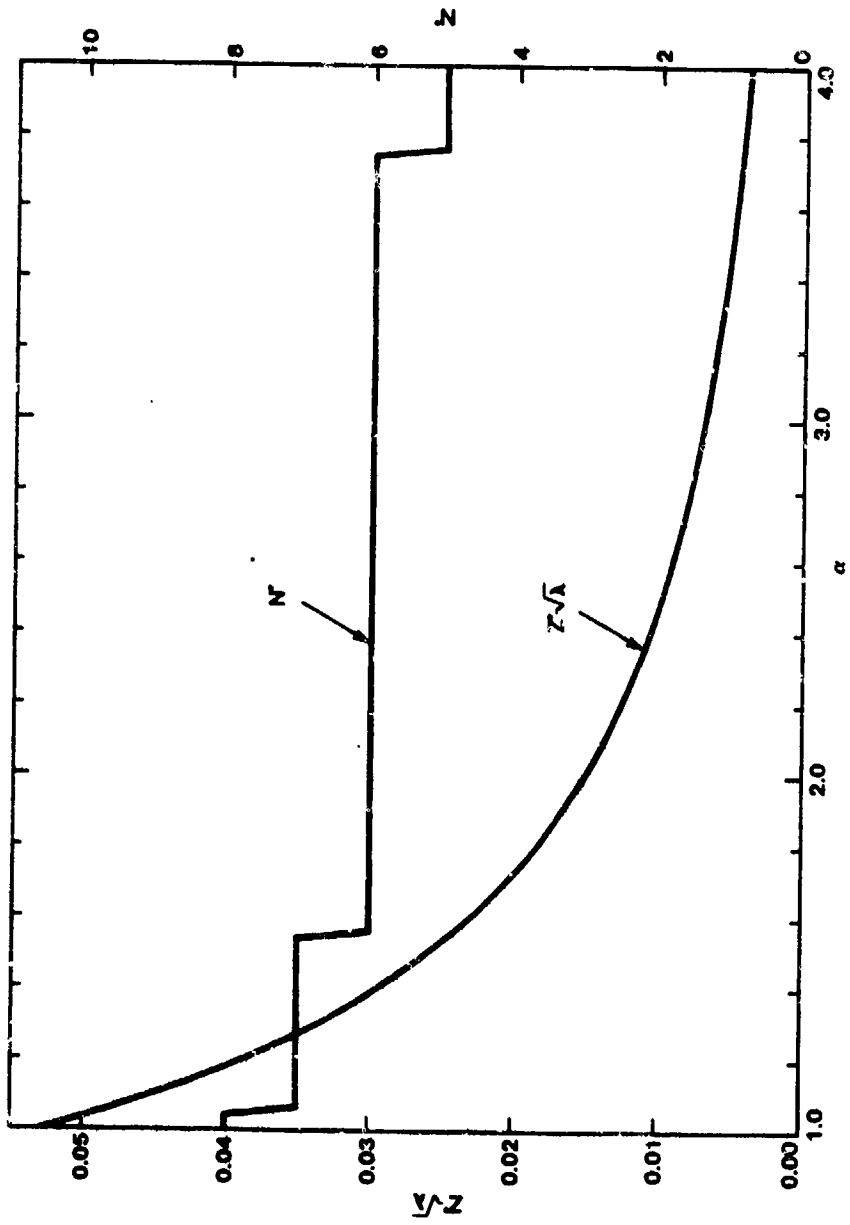


Figure 6.13 The optimal transmissions for ALOHA with capture and without transmission radius

As an example of an inhomogeneous terminal distribution, the effect of a gap of width b in an otherwise uniformly Poisson-distributed terminal population on the optimal transmission has been considered. The expected progress of a packet residing at the terminal on the bank and destined to cross the gap is evaluated with parameter $\beta = \lambda b^2$, called gap intensity. For fixed λ, the existence of the gap helps the progress for $\beta < 2$, because some of the possibly interfering terminals are removed by the gap. The maximum in the optimized expected progress occurs at about $\beta = 1$. Thus, to cross most gaps wider than the average inter-terminal distance, one had better not use a large transmission radius, but should more sensibly use a separate channel or wire.

CHAPTER 7
Throughput-Delay Characteristics of Some Slotted-ALOHA Multi-Hop Packet Radio Networks

A Markovian model is formulated to find the throughput-delay performance for slotted-ALOHA multi-hop packet radio networks with fixed configuration of packet radio units (terminals and repeaters) and fixed source-to-sink paths of packets. Improvements in performance effected by the adjustment of transmission parameters (suppression/acceleration) according to the states of nearby units and/or by having repeaters equipped with multiple buffers are demonstrated. An efficient way to numerically solve a large Markov chain problem is also shown in Appendix E.

7.1 Introduction

The packet radio network considered in this chapter is a ground-based minicomputer communication network using a shared multiple-access radio channel. One of its potential usages will be providing real-time computer-based communication for packet radio-equipped military users both in garrison and in the battlefield. Another application is to replace regional wired packet-switching networks without the need for cable extension.

Although some intensive experimental research on packet radio networks has taken place at several locations during the last few years (e.g., PRNET in [Kahn78]), little theoretical work about their performance evaluation seems to have been published so far. As compared to the analysis of one-hop broadcast networks for which extensive literature has appeared, one of the difficulties in dealing with multi-hop networks resides in the fact that the issue of routing comes into play as in the wire-based store-and-forward networks. However, because of colliding transmissions from multiple packet radio units, we have not found any exact solution — whether in a product form or not — for evaluating the performance of a general class of multi-hop packet radio networks. One of the reasons that a discrete-time queueing network (modeling the slotted-ALOHA system) does not lend itself to a product-form solution is that more than one event can occur in a single renewal interval [Bhar80].

As for the approximate evaluation of the average packet delay and the optimal routing with respect to it, some contributions may be noted. Leiner [Lein80] showed an approximate way to get the delay at each link given its traffic requirement using Kleinrock's ZAP approximation [Klei77] for the throughput-delay curves for a variety of channel access protocols in single-hop systems. Kung [Kung81] speculates that the average delay is a convex function in the space over the traffic requirements on all links, on the basis of the ZAP approximation

[Klei77] of the throughput-delay curves. Thus he adapts the flow deviation method, originally developed in [Frat73] for wire-based store-and-forward networks, to the multihop packet radio networks. Some other authors [Boor80,Sidi81] create more or less idealistic assumptions (such as zero propagation delay and perfect delay capture) to inhibit interference of transmissions and discuss the resulting throughput and optimal routing. Some two-hop networks are analyzed by Tobagi [Toba80a,Toba80b].

In this chapter, we take a Markov-chain approach to find the throughput-delay characteristics for a general class of slotted-ALOHA multi-hop packet radio networks which consist of a relatively small number of packet radio units. In Section 7.2 we describe our basic model of packet radio networks in detail. This is followed in Section 7.3 by the Markov-chain formulation to calculate the throughput and average end-to-end packet delay for a given network. The trade-offs between them are shown for two example networks.

In the following sections, we propose and analyze three ways (and their combinations) to reduce the average packet delay for a given throughput requirement and to increase the maximum supportable throughput. In the Appendix E we show some computational techniques to solve a Markov-chain problem with a large state space.

7.2 Basic Model

In this section, we describe in detail our basic model of packet radio networks. Consider a network consisting of a fixed number of packet radio units, each having an omnidirectional antenna and thereby being capable of either transmitting or receiving a packet but not both simultaneously. We distinguish the two kinds of packet radio units: terminal and repeater. A terminal is defined to be a unit which can be a source and/or a sink of packets but does not relay any packets in transit. A repeater is defined to be a unit which neither generates nor absorbs any packets but only relays them.

We assume that every unit is within the transmission range of some other units but not necessarily of all others; this hearing topology is given and fixed. Let us represent the hearing configuration by a matrix (h_{ij}) defined by

$$h_{ij} = h_{ji} = \begin{cases} 1 & : \text{ if units } i \text{ and } j \text{ hear each other} \\ 0 & : \text{ otherwise} \end{cases}$$

$$h_{ii} = 1 \tag{7.1}$$

We also assume a given set of fixed paths for packets which connect pairs of specific source and sink terminals via a number of repeaters. Thus packets originating at the source terminal of a particular path are sent (with specific destination id for each link) in a store-and-forward manner through several repeaters along a unique path down to the sink terminal and

absorbed there. Let these paths be numbered as $k = 1, 2, 3, \cdots$. Figure 7.1 shows two such networks which we will call networks 1 and 2. (Notice that the paths in these networks are carefully laid out in order to prevent deadlock situations inherent in uncontrolled finite-buffer systems. For example, the reversal of direction of path 1 in network 2 would bring about the so-called indirect store-and-forward deadlock [Gerl80].) We will use them as examples for evaluating throughput and average source-to-sink packet delay.

All units in the network are assumed to use the common radio channel band. The reference time is slotted, and the slot size is such that it includes the transmission time of a packet, its propagation delay, and the time needed to notify the transmitter of the results of transmission (successful or not). This acknowledgement is assumed to be given for free. We employ the slotted-ALOHA protocol by which we mean that the channel slot is used or idle. Also we neglect channel noise and assume no channel errors for single transmissions.

We now proceed to describe in more detail the properties of our terminals and repeaters. Let us begin with a terminal. A terminal can be a source of at most one path (for simplicity in our model) and/or a sink of multiple paths, and it possesses buffer space for a single packet only. The packets received at proper sink terminals are consumed immediately so that they do not claim any buffer space. Let us represent the state of terminal i, s_i, by the state of its buffer: that is, it is in the 'empty' state when the buffer is empty and in the 'k-backlogged' state when the buffer contains a packet which belongs to path k. Thus

$$s_i = \begin{cases} 0 & : \text{empty} \\ k\ (\neq 0) & : k\text{-backlogged} \end{cases} \tag{7.2}$$

Notice that since a terminal can be a source of at most one path, every source terminal has exactly two possible states. On the other hand, sink-only terminals are always in the empty state.

A source terminal of path k in the empty state generates a new packet at the beginning of a slot instantaneously with probability $\lambda(k)$, and in such a case it transmits the packet with probability 1 in the same slot. (Thus $1-\lambda(k)$ is the probability of no action in any given slot when in the empty state.) Suppose that the destination of the first transmission from a source terminal i is unit j (repeater or sink terminal). The conditions for this transmission to be successful are: (i) for all units which can be heard by j, including j and excluding i, that such a unit does not transmit in the same slot, and that (ii) unit j is in the empty state. (The states of a repeater are explained shortly.) If the first transmission is successful, the terminal remains in the empty state for the next slot. If it is unsuccessful because either (i) or (ii) is not met (we call these 'collision' and 'buffer blockage', respectively) or both, the terminal goes into the k-backlogged state.

A k-backlogged terminal transmits its packet with probability $p(k)$ in any slot and delays action until the next slot with probability $1-p(k)$. The conditions for successful transmission are the same as above (i) and (ii). Immediately following its successful transmission, a backlogged terminal switches to the empty state. In the cases of no action and

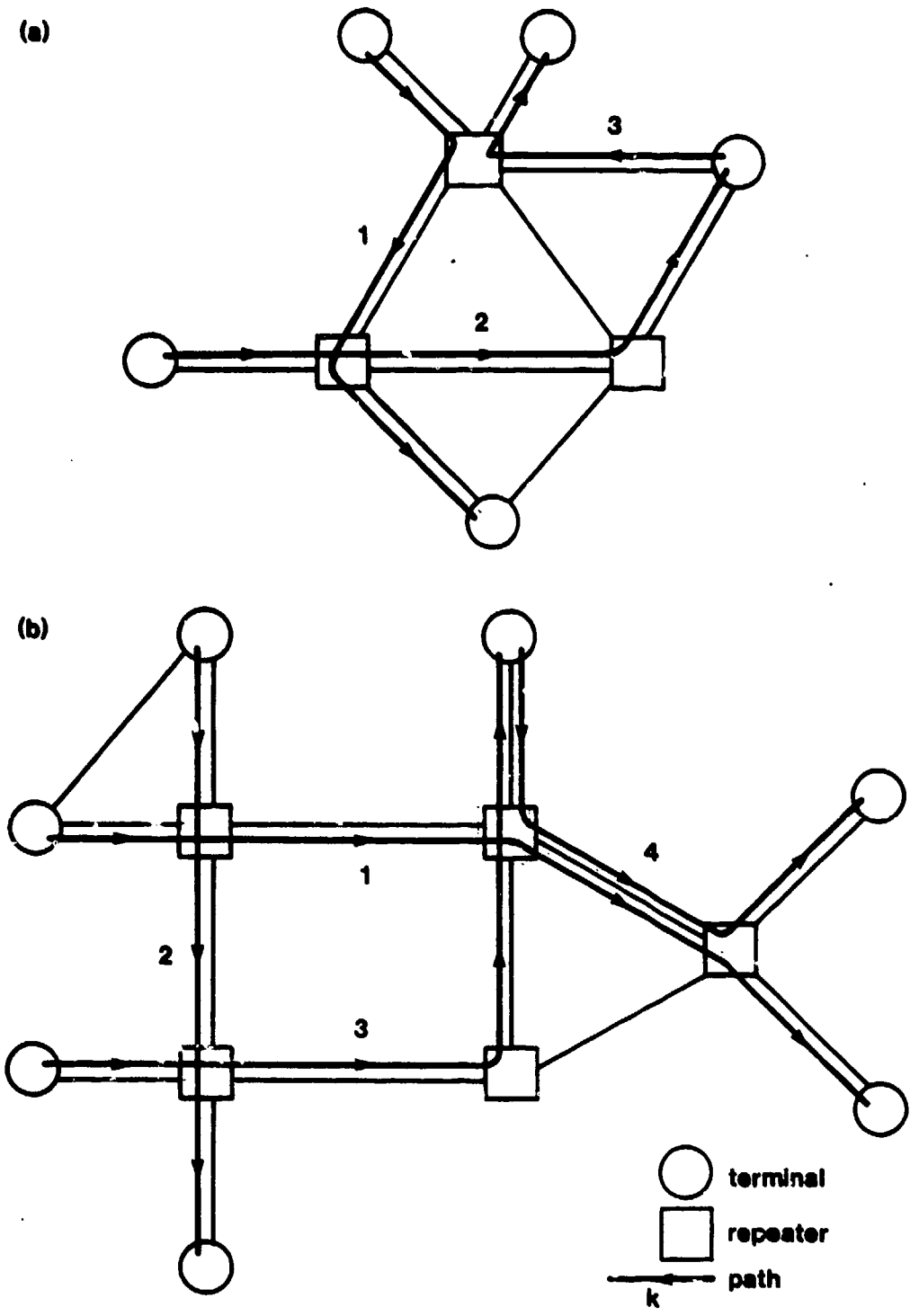


Figure 7.1 Examples of networks (a) Network 1 — 5 terminals, 3 repeaters and 3 paths (b) Network 2 — 7 terminals, 5 repeaters and 4 paths

unsuccessful transmission, it remains in the same backlogged state for the next slot. We assume that a sink terminal receives a packet with success only on the condition of no collision, irrespective of whether the buffer is empty or occupied by an outgoing packet. The successful reception at a sink terminal does not affect its state.

Next we describe the operation of repeaters. As said before, a repeater does not generate or absorb any packets but simply relays them. Let each repeater be equipped with a single packet buffer. The state of repeater i, denoted by s_i, is identified by the contents of its buffer, just as for terminals. Thus, we have the same representation:

$$s_i = \begin{cases} 0 & : \text{empty} \\ k\ (\neq 0) & : k\text{-backlogged} \end{cases} \tag{7.3}$$

However, since a repeater can be used by multiple paths as exemplified by the networks depicted in Figure 7.1, the number of distinct states for a given repeater is equal to the number of paths which pass that repeater plus one (the empty state).

A repeater in the empty state takes no action in any slot with probability 1. A repeater in the k-backlogged state behaves just like a terminal in the k-backlogged state: i.e., in any slot it transmits a packet according to a Bernoulli process with parameter $p(k)$. The conditions for successful transmission from a repeater are similar to those from a terminal mentioned above. We note that the successful transmission of a packet from a repeater in the k-backlogged state changes its state into the empty state. On the other hand, the successful reception of a packet belonging to path k at a repeater in the empty state changes its state into a the k-backlogged state.

We note that the above-mentioned slotted-ALOHA transmission protocol is an extension to the multi-hop environment of the immediate-first-transmission (IFT) protocol referred to by Tobagi [Toba80a].

7.3 Formulation

We now turn our attention to the formulation of the procedure to calculate the throughput and the average source-to-sink packet delay for networks such as described in Section 7.2. Note that the packet transmission process at each unit in any slot is based only on its current state and not on the past states. This memoryless property makes slot boundaries Markov points. Thus we follow the usual formulation of discrete-time homogeneous Markovian systems.

First, let us represent the state of the whole network in a given slot, s, by the Cartesian concatenation of the states of all units in the network: $\mathbf{s} = (s_1, s_2, \cdots, s_M)$, where M is the total number of units involved. Also, represent the behavior of the network for the slot, e, by the Cartesian concatenation of the actions of all units: $\mathbf{e} = (e_1, e_2, \cdots, e_M)$, where

$$e_i = \begin{cases} 0 & : \text{unit } i \text{ does not transmit} \\ k\ (\neq 0) & : \text{unit } i \text{ transmits a packet of path } k \end{cases} \tag{7.4}$$

The behavior of the network is a stochastic phenomenon given the current state of the network. Since each unit behaves independently of others, we may write

$$p(\mathbf{e}\,|\,\mathbf{s}) \triangleq Prob\,[\text{behavior} = \mathbf{e}\,|\,\text{current state} = \mathbf{s}] = \prod_{i=1}^{M} Q_i(s_i,e_i) \tag{7.5}$$

where each factor $Q_i(s_i,e_i)$ for unit i is given as follows. For terminal i which is the source of path k,

$$Q_i(s_i,e_i) = \begin{cases} 1-\lambda(k) & : \ s_i = e_i = 0 \\ \lambda(k) & : \ s_i = 0,\ e_i = k\ (\neq 0) \\ 1-p(k) & : \ s_i = k\ (\neq 0),\ e_i = 0 \\ p(k) & : \ s_i = e_i = k\ (\neq 0) \end{cases} \tag{7.6}$$

For terminal i where no paths originate, $Q_i(s_i,e_i) = 1$. For repeater i,

$$Q_i(s_i,e_i) = \begin{cases} 1 & : \ s_i = e_i = 0 \\ 1-p(k) & : \ s_i = k\ (\neq 0),\ e_i = 0 \\ p(k) & : \ s_i = e_i = k\ (\neq 0) \end{cases} \tag{7.7}$$

Given the current state s and behavior e of the network, it is not difficult to determine the state of the system for the next slot, $\mathbf{s}' = (s'_1, s'_2, \cdots, s'_M)$. For example, non-transmission at unit i does not affect the states of any other units. The successful transmission from unit i to unit j of a packet belonging to path k brings

$$s'_j = \begin{cases} s_j & : \ j \text{ is a sink terminal of path } k \\ k & : \ j \text{ is a repeater} \end{cases}$$

$$s'_i = 0 \tag{7.8}$$

The unsuccessful transmission from unit i of a packet of path k simply gives

$$s'_i = k \tag{7.9}$$

Thus by examining all possible events for each state, we can construct the transition probabilities of our homogeneous Markovian system:

$$P(\mathbf{s}'\,|\,\mathbf{s}) \triangleq Prob\,[\text{next state} = \mathbf{s}'\,|\,\text{current state} = \mathbf{s}]$$

$$= \sum_{\mathbf{e} \text{ such that it gives } \mathbf{s}'} p(\mathbf{e}\,|\,\mathbf{s}) \tag{7.10}$$

Let us denote by $\pi(\mathbf{s})$ the equilibrium probability that the network is in state s:

$$\pi(\mathbf{s}) \triangleq Prob\,[\text{state} = \mathbf{s}] \tag{7.11}$$

Then we have the equilibrium state equations

$$\begin{cases} \pi(\mathbf{s}') = \sum_{\mathbf{s}} \pi(\mathbf{s})\, P(\mathbf{s}' \mid \mathbf{s}) & \text{for all } \mathbf{s}' \\ \sum_{\mathbf{s}} \pi(\mathbf{s}) = 1 \end{cases} \tag{7.12}$$

This system of linear simultaneous equations may be solved numerically given values of $\lambda(k)$ and $p(k)$. We have some comments on the computational aspects involved in solving a large system of linear equations of above type in the Appendix E.

Once the solution $\pi(\mathbf{s})$ is obtained, we can compute the following quantities of interest. First, the average backlog of packets along path k, $Q(k)$, is given by

$$Q(k) = \sum_{\mathbf{s}} Q(\mathbf{s};k)\, \pi(\mathbf{s}) \tag{7.13}$$

where $Q(\mathbf{s}:k)$ denotes the number of k-backlogged units when the network is in state s. The total average backlog, Q, is given by

$$Q = \sum_{k} Q(k) \tag{7.14}$$

Second, the throughput of path k, $S(k)$, is defined as the average number of successfully delivered packets per slot from source to sink of path k. Since no packets disappear on their way, they can be counted at the sink terminal:

$$S(k) = \sum_{\mathbf{s}} \pi(\mathbf{s}) \sum_{\mathbf{e}} p(\mathbf{e} \mid \mathbf{s})\, S(k \mid \mathbf{s},\mathbf{e}) \tag{7.15}$$

where

$$S(k \mid \mathbf{s},\mathbf{e}) = \begin{cases} 1 : \text{successful transmission to the sink terminal} \\ \quad \text{of path } k, \text{ given s and e} \\ 0 : \text{otherwise} \end{cases} \tag{7.16}$$

Thus the total throughput of the network, S, is given by

$$S = \sum_{k} S(k) \tag{7.17}$$

We call the maximum attainable throughput with respect to changing the values of $\lambda(k)$ and $p(k)$ the capacity of the network. Lastly, the average packet delay of path k, $D(k)$, is defined as the average time, in slots, that a packet of path k takes to go from its source to sink terminal. Applying Little's result [Litt61] to each path, this is given by

$$D(k) = 1 + \frac{Q(k)}{S(k)} \tag{7.18}$$

Note that the first term accounts for the first transmission from the source. Applying Little's result to the whole network, the overall average packet delay, D, is given by

$$D = 1 + \frac{Q}{S} \tag{7.19}$$

This concludes the formulation of our basic model.

We have calculated these quantities for networks 1 and 2 depicted in Figure 7.1, and the results are shown in Figures 7.2 (for network 1) and 7.3 (for network 2) in the form of the overall average packet delay (D) (for the basic protocol) versus the total throughput (S). We have assumed throughout this chapter that $\lambda(k) = \lambda$ and $p(k) = p$ for all paths, not only for simplicity but also for fairness among peths. The displayed curves are actually optimum envelopes in the sense that given λ the value of p is adjusted in order to minimize D. (The optimization procedure is based on the Fibonacci search method for a unimodal function [Kues73]. The unimodality of D in p has been assumed.) The curves in Figure 7.4 show the throughput-delay relations for individual paths in network 2. Note that these curves have been obtained for overall optimal values of p, not subject to individual optimization.

Now let us look at the behavior for small values of λ. The throughput of each path is nearly given by λ, irrespective of the values of p, because collisions and buffer blockage are rare. Also the average packet delays for paths 1, 2 and 3 of network 1 are given by $1+2/p$, $1+2/p$, and $1+1/p$, respectively, and thus the overall average packet delay for network 1 is given by

$$\frac{1}{3}[(1+\frac{2}{p})+(1+\frac{2}{p})+(1+\frac{1}{p})] = 1+\frac{5}{3p}$$

Similarly, the overall average packet delay for network 2 for small λ is given by

$$\frac{1}{4}[(1+\frac{3}{p})+(1+\frac{2}{p})+(1+\frac{3}{p})+(1+\frac{2}{p})] = 1+\frac{5}{2p}$$

Thus it is clear that $p = 1$ brings the least delays 8/3 and 7/2 for networks 1 and 2, respectively.

Under moderate values of λ, we see from Figure 7.4 that the throughputs are rather fairly distributed among all paths while the delays are nearly proportional to the individual path lengths.

For larger values of λ, we reach the maximum in throughput, the capacity. The capacities of networks 1 and 2 are about 0.16 and 0.19 packets per slot, respectively, for the present transmission protocol. As the throughput approaches its maximum, the average packet delay increases rapidly for a marginal increase in throughput.

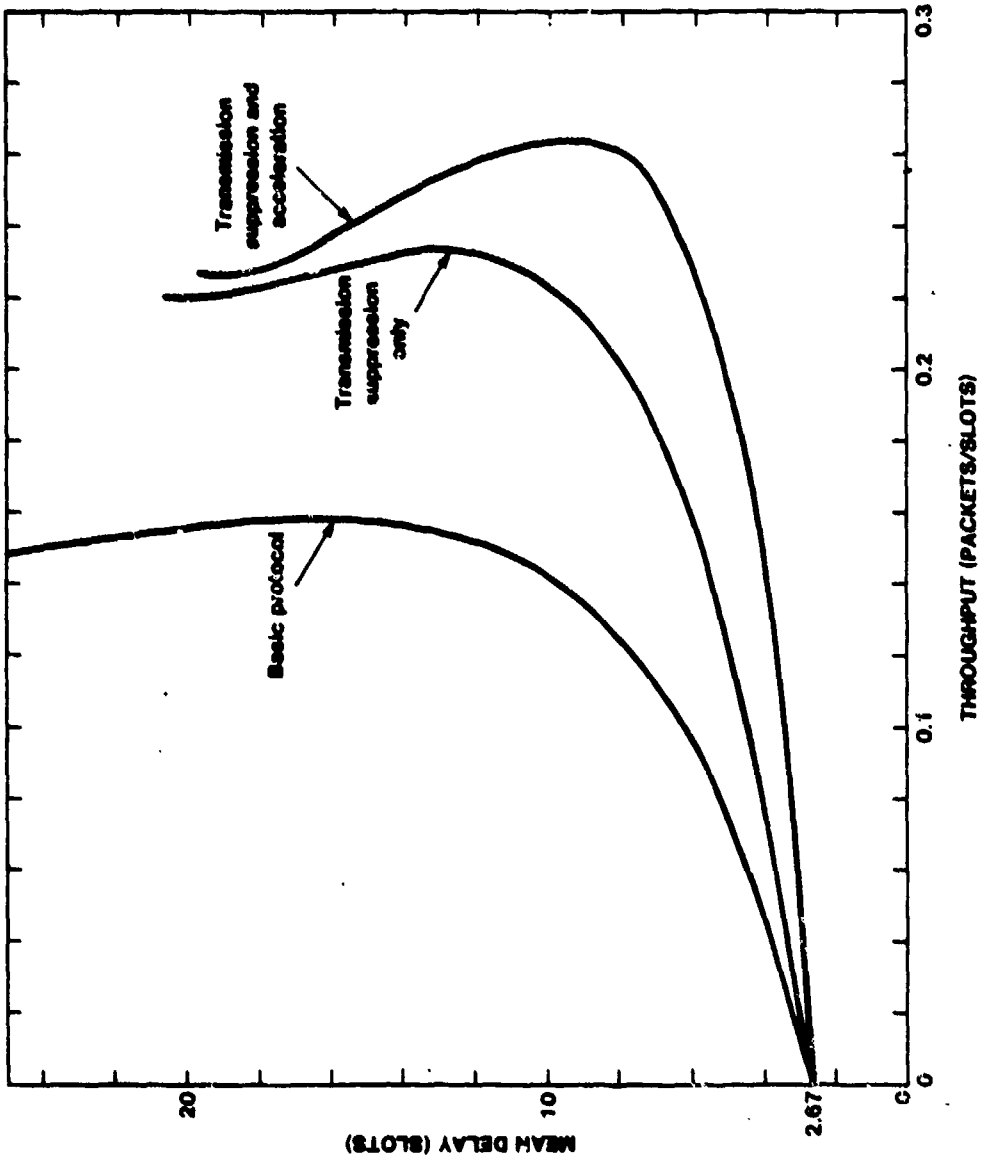


Figure 7.2 Throughput-delay characteristics for network 1 with single-buffered repeaters

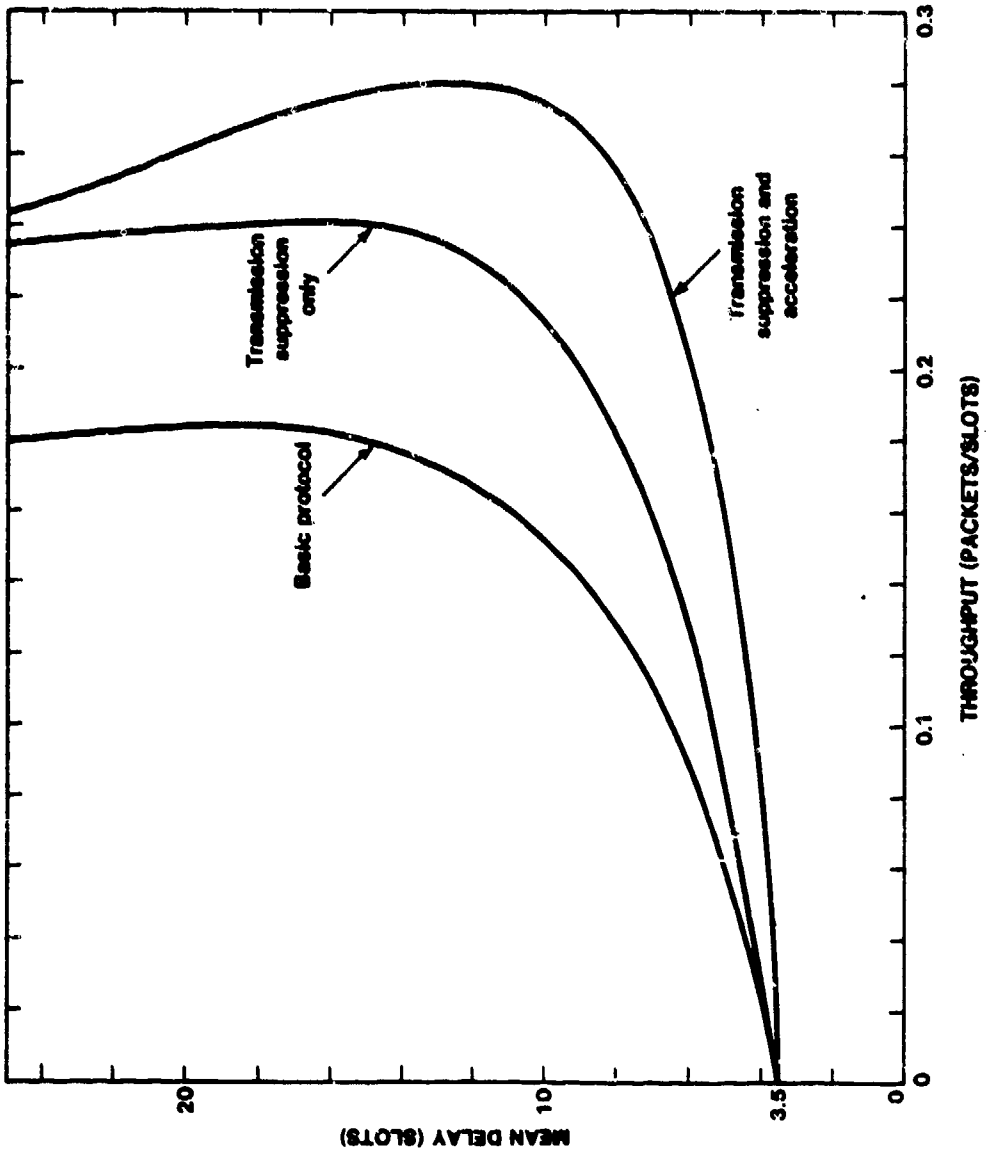


Figure 7.3 Throughput-delay characteristics for network 2 with single-buffered repeaters

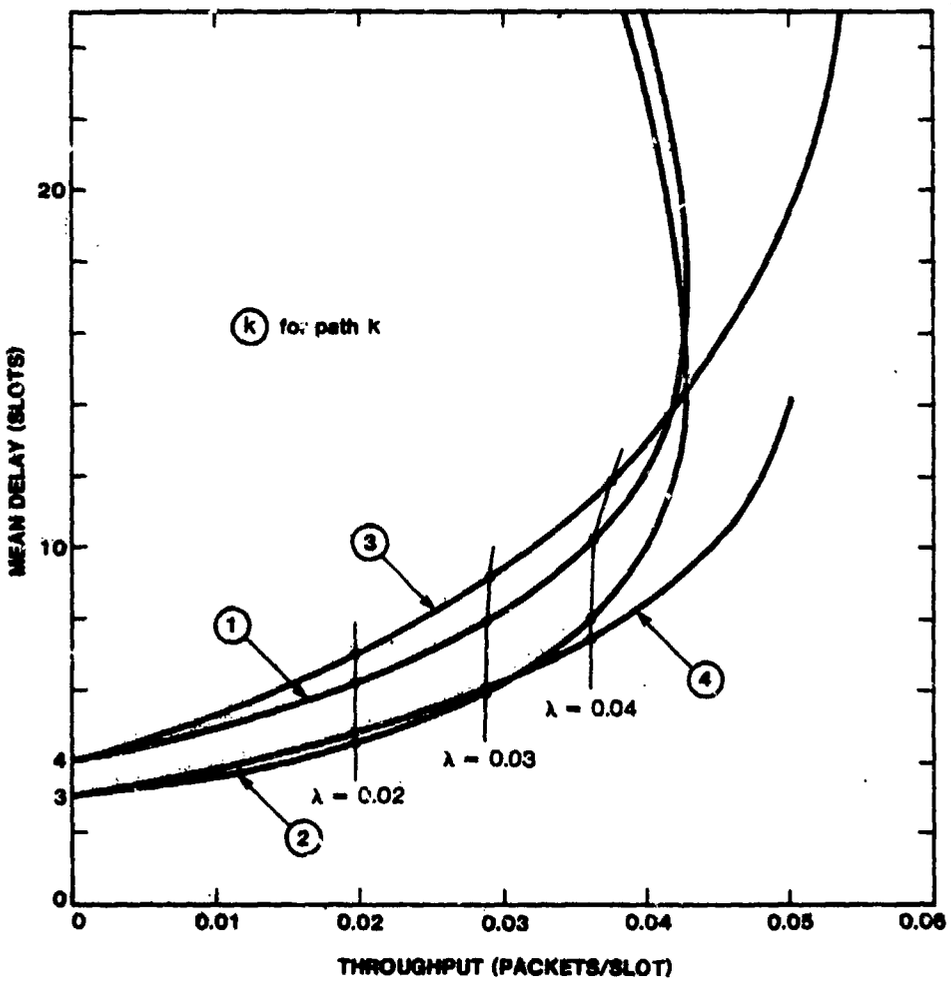


Figure 7.4 Throughput-delay characteristics for individual paths in network 2 (basic protocol and single buffer)

7.4 Improvement by Transmission Suppression

In the basic protocol described in Section 7.2, we have assumed that the behavior of each unit in any slot is based on its current state only (isolated strategy). Thus it may happen that a unit transmits a packet to a unit having no available buffer, resulting for sure in failure. If information about the buffer state of the destination unit were available to the transmitter, it could avoid this foreseeable wastage of channel capacity by suppressing the transmission and make it available for others. We exploit this possibility in this section.

Suppose that repeaters with no available buffer broadcast a 'busy tone', in a different channel, towards their neighbors at the beginning of any slot. (This busy tone should not be confused with the one introduced in [Toba75] to solve the hidden terminal problem. The busy tone in [Toba75] is emitted by a receiver when it is receiving a packet.) It is assumed that the hearing topologies on both channels are identical. Also suppose that the busy tones from multiple repeaters do not collide and necessary information is always captured correctly by potential transmitters instantaneously. Knowing that the destination has no available buffer, a unit with a packet will suppress its transmission with probability 1 (i.e., p is set to 0) for the current slot. Otherwise the value of p is unchanged.

Thus the probability of behavior e given the current state s is expressed as

$$p(\mathbf{e}\,|\,\mathbf{s}) = \prod_{i=1}^{M} Q_i(s_i, s_j, e_i) \tag{7.20}$$

where j is the destination id of the transmission from unit i (which is unique given s_i) and each factor $Q_i(s_i,s_j,e_i)$ is given as follows. For terminal i which is a source of path k,

$$Q_i(s_i,s_j,e_i) = \begin{cases} 1-\lambda(k) & : \ s_i = e_i = 0,\ s_j = 0 \\ \lambda(k) & : \ s_i = 0,\ e_i = k\ (\neq 0),\ s_j = 0 \\ 1-p(k) & : \ s_i = k\ (\neq 0),\ s_j = e_i = 0 \\ p(k) & : \ s_i = e_i = k\ (\neq 0),\ s_j = 0 \\ 1 & : \ s_j \neq 0,\ e_i = 0 \end{cases} \tag{7.21}$$

For terminal i where no paths originate, $Q_i(s_i,s_j,e_i) = 1$. For repeater i,

$$Q_i(s_i,s_j,e_i) = \begin{cases} 1 & : \ s_i = 0 \text{ or } s_j \neq 0,\ e_i = 0 \\ 1-p(k) & : \ s_i = k\ (\neq 0),\ s_j = e_i = 0 \\ p(k) & : \ s_i = e_i = k\ (\neq 0),\ s_j = 0 \end{cases} \tag{7.22}$$

Since the above modification of transmission parameters is again based on the current system state only, we still have the Markovian property for the network behavior. Thus the formulation proceeds as in Section 7.3. With respect to a source terminal in the empty state which has been forced to suppress its transmission, we can distinguish two models: (i) it then goes into the backlogged state retaining the packet (retransmission model) ; and (ii) it remains in the empty state dropping the packet (loss model). Since the delay due to buffer blockage at

entry to the network must also be counted as much as one slot for a user, we opt for the model (i) in this chapter. If we were to evaluate the delay for only those packets that are accepted in the network, we would employ the model (ii). Thus, in the case of transmission suppression as well as unsuccessful transmission from unit i of a packet of path k, the next state of unit i will be $s'_i = k$.

In Figures 7.2 and 7.3 we show the throughput-delay characteristics improved by the transmission suppression scheme for networks 1 and 2, respectively. We see that the capacities of the networks are greatly increased. A close comparison of what is happening at each unit and along each path between the basic protocol and the transmission suppression protocol for the same values of λ and p has revealed that the present scheme gives rise to much fewer collisions and thus much fewer backlogs with only slightly higher throughput. This results in much lower average packet delay due to Little's result. Also it has turned out that the utilizations (fraction of time when the buffer is occupied) of units are much lower, especially for source terminals. Thus, it seems that, in case of congestion, newly entering packets are likely to be blocked at source terminals which prevent them from entering the network. Therefore, we think that this transmission suppression scheme provides a natural flow control at the network access level [Gerl80], and speculate that the buffer-full condition at entry repeaters could be a good indication of network congestion. The fashion in which the buffer-full condition propagates to neighbors is analogous to the 'backpressure bit' scheme for the virtual circuit (hop level) flow control employed in Tymnet [Gerl80].

7.5 Improvement by Transmission Acceleration

In this section we seek another way to take advantage of information about the network state. In addition to knowledge of the state of the immediate destination, let us assume that each unit with a packet knows the states of all hearing neighbors of the destination. Then, when the destination unit (which could be a sink terminal) and its neighbors (which must not be a source terminal) are all in the empty state, it is foolish to flip a coin to decide whether to transmit or not. Since we know for sure there are no other transmissions around the destination, we simply raise the value of p to 1 and with probability 1 transmit our packet. We call this operation, combined with the transmission suppression scheme described in Section 7.4, as the transmission suppression/acceleration protocol. Of course the above-mentioned information is not obtained for free. However it is interesting to examine the resulting improvement in the throughput-delay performance as an ideal limit. The delay vs. throughput curves based on this protocol are shown in Figures 7.2 and 7.3 for networks 1 and 2, respectively. As expected we see the increased capacity as well as the reduced delay for given throughput.

7.6 Improvement by Multiple Buffers for Repeaters

So far we have assumed that terminals and repeaters are equipped only with single packet buffers, and one of the conditions for successful transmission is that the buffer of the destination unit be empty (except for the final delivery to sink terminals). Therefore, some improvement in throughput-delay performance is expected by letting repeaters have more than one buffer. (Multiple buffers for terminals do not help performance in our model.) In this section, we examine this effect.

Let m be the number of buffers (each being capable of containing a single packet) provided for each repeater. The packets in buffers $1, 2, \cdots$ form a queue waiting for transmission in this order. We express the state of repeater i, s_i, as

$$s_i = (s_i^{(1)}, s_i^{(2)}, \cdots, s_i^{(m)}) \tag{7.23}$$

where

$$s_i^{(l)} = \begin{cases} 0 & : \text{ buffer } l \text{ is empty} \\ k\ (\neq 0) & : \text{ buffer } l \text{ contains a packet of path } k \end{cases}$$

$$l = 1, 2, \cdots, m \tag{7.24}$$

Note that the state $s_i^{(l)} = 0$ for some l implies that $s_i^{(l')} = 0$ for all l' such that $l < l' \leq m$.

Now a repeater having at least one packet transmits one in buffer 1 (say, of path k), again according to a Bernoulli process with parameter $p(k)$. The condition for the successful transmission to a repeater is, besides no collision, that there be at least one empty buffer at the receiver. If the transmission is successful, the transmitted packet joins the tail of the queue (i.e., it is placed in the lowest-numbered available buffer) at the receiver, while at the transmitter (if it is a repeater) other packets, if any, are moved toward the head of queue by one position. If the transmission is unsuccessful the failed packet remains in buffer 1. In other aspects, the transmission protocols are the same as before.

For networks with multibuffer repeaters, we can still formulate a Markov chain problem and solve it numerically to obtain the delay vs. throughput curves. In Figure 7.5 we plot the optimum delay curves for $m = 1, 2$ and 3 for network 1 without adjustment of transmission parameters. From this figure we see that the capacity of the network for $m = 2$ (0.21 packets per slot) is about a 30% increase over the single-buffer case. This is also accompanied by a reduced delay for a given throughput. However, the improvement by going from $m = 2$ to $m = 3$ is not so outstanding as increasing m from 1 to 2. A similar observation has been made by Tobagi [Toba80a] for some two-hop networks, where he comments that the lack of any important improvement experienced by increasing m is mainly explained by the fact that the system, at optimum, is mostly 'channel bound' as opposed to 'storage bound.'

The effect of the transmission suppression/acceleration scheme, described in Sections 7.4 and 7.5, in the case of multibuffer repeaters is demonstrated in Figure 7.6 for network 1 with $m = 2$. Here the transmission suppression is in effect when all the buffers at the destination are occupied. The transmission acceleration takes place when all neighbors of the destination (which must not be source terminals) have empty buffers. In Figure 7.6 we still see some improvement brought about by these schemes although they are not as large as in the single-buffer case shown in Figure 7.2.

7.7 Conclusion

We have analyzed the throughput-delay characteristics for slotted-ALOHA multi-hop packet radio networks where the hearing configuration of packet radio units (terminals and repeaters) and source-to-sink paths of packets are given and fixed. The problems are formulated as discrete-time Markov chains and then solved numerically.

Besides the basic model — characterized by isolated transmission behavior and single-buffered repeaters — three ways to improve the throughput-delay performance have been exploited. They include (i) transmission suppression when the destination's buffer is occupied, (ii) transmission acceleration when the buffers of all neighbors of the destination are empty, and (iii) multiple buffers for repeaters.

It has been shown that the transmission suppression scheme provides a natural flow control at the network access level to prevent packets from entering the 'communication subnet.' This brings about significantly lower delay for a given throughput, and achieves a much higher maximum throughput. The transmission acceleration combined with appropriate suppression gives further improvement in the throughput-delay trade-offs at the cost of necessitating more information about the network state.

With more than one buffer for repeaters we have fewer chances of failure of transmission due to a buffer shortage at destinations. It has been shown that increasing the number of buffers from 1 to 2 offers more performance enhancement than going from 2 buffers to 3 buffers. The effect of transmission suppression/acceleration in the multibuffer case was also demonstrated.

Although the Markov chain approach used in this chapter is not suitable for large-size networks due to too much computational time and storage required, it may be useful for examining the effect of any particular heuristic protocol in prototype (small) network models. Also it can provide a benchmark against which simulation models are validated.

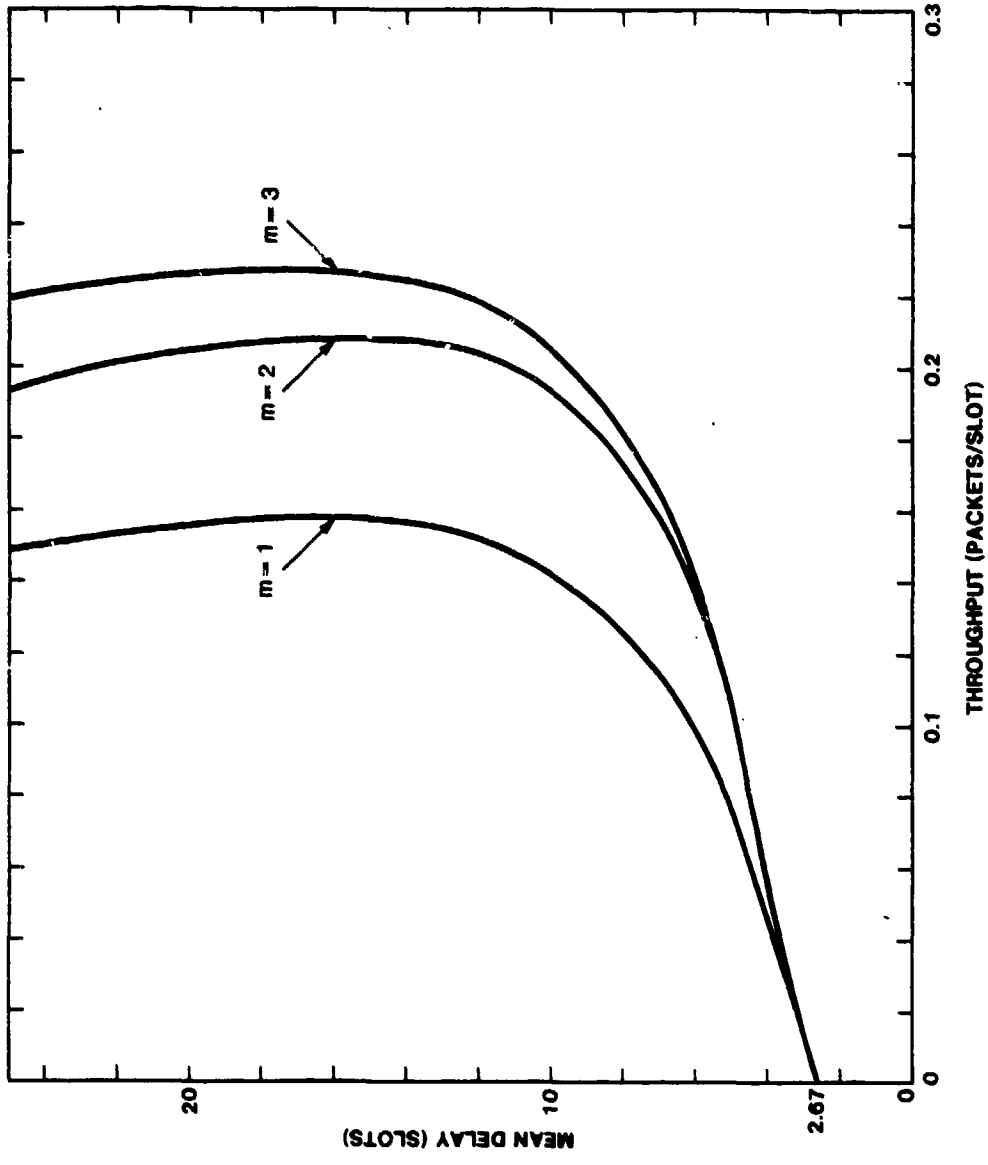


Figure 7.5 Throughput-delay characteristics for network 1 (basic protocol) with m buffers for repeaters (m = 1, 2, 3)

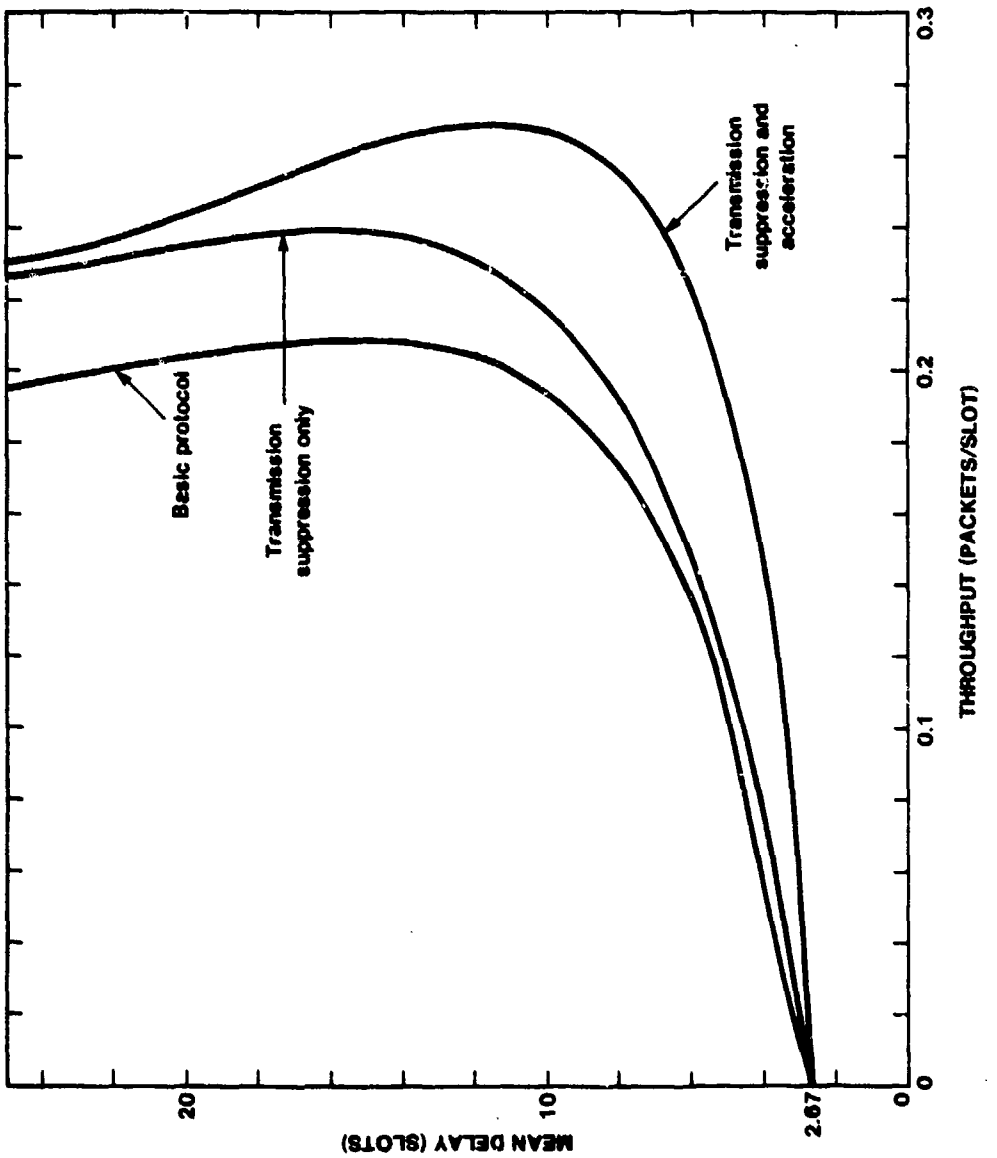


Figure 7.6 Throughput-delay characteristics for network 1 with double-buffered repeaters

CHAPTER 8
Concluding Remarks

8.1 Application to Local Area Networks

Our study of performance evaluation for packet radio networks has close relationship with its counterpart in local area computer networks such as ETHERNET [Metc76]. In fact, one of the three access protocols recommended by IEEE Computer Society Project 802 Local Area Networks Standards Committee is CSMA with collision detection, whose throughput, output processes and packet delay have been extensively studied in this dissertation. Usually, a local area network environment corresponds to a single-hop, fully-connected system (such as we considered in Chapters 2 and 3) since every user shares common information on the channel state. However, the case involving some malfunctioning equipment can be viewed as a hidden-user situation (such as we treated in Chapter 4) in the sense that some users ignore the ongoing channel activity. Then, for example, the curves in Figure 4.4(a) for the throughput of hidden-user CSMA may be thought of as showing the throughput degradation in the imperfect environment of local area networks.

8.2 Considerations on the Model and the Methodology

In order to show how modern mathematics fails to solve a simple problem, consider as an example a system of two pure ALOHA users. Let each of them independently alternate between the transmission period of fixed duration and the idle period of exponentially distributed duration. Suppose as usual that a transmission is successful if the complete transmission time is not overlapped by the other user's transmission. Although we can obtain the throughput for this system (as in Eq.(2.8)), we have been unable to find the higher moments or the distribution for the interval between two successive successful transmissions (called packet interdeparture time). Incidentally, this problem has its equivalent in a vehicle traffic system [Yeo78]. A complete knowledge of the distribution of the packet interdeparture time in our system (corresponding the service time distribution in traditional queueing systems) is prerequisite if we are to analyze the packet queueing delay. This is an example of a very simple but fundamental problem for which we do not know the (exact) solution. We have many more difficult problems in the performance analysis of packet radio systems.

In Chapter 4 of this dissertation, we have shown an approximation approach to the above-mentioned problem. Methodologicaliy, it approximates a point process (inter-event times are dependent) by a renewal process (inter-event times are independent and identically distributed). By doing so, the extant theory of renewal processes is at hand to apply. A similar trend can be seen in the analysis of general queueing networks [Sevc77,Kueh79,Whit82]. It is this author's opinion that this direction is one practical way to attack the problem.

8.3 Suggestions for Future Research in Multi-Hop Systems

The theoretical performance study of multi-hop packet radio networks is still in the early stages, and almost all the important aspects of network operation are virtually untouched. Even the very first problem of determining an efficient and reliable channel access protocol in general does not seem to have a clear answer. At present we do not know whether we should consider adaptation from existing local area network protocols or invent something new for multi-hop systems. A promising protocol may be the busy-tone multiple-access (BTMA) [Toba75] with which collisions at receivers are mostly avoided. However, the assessment of its potential drawback, namely the suppression of other transmissions which would coexist successfully without BTMA, has not been studied.

Given a channel access protocol and a set of source-to-destination paths, the evaluation of the average end-to-end packet delay at given throughput requirement is the next problem. We prefer an analytic formula for the average delay rather than numerical procedures such as in [Lein80,Lee82,Sliv82a-c] and in Chapter 7. Our proposed mean delay formulas for single-hop systems (in Chapter 5) is one such attempt.

After the analysis of throughput and delay, the issues of routing and flow control will be studied (in terms of performance). Only an abstract of the contribution by Humblet on the iterative algorithm to find the optimal routing is available [Humb82]. An attempt by Kung [Kung81] combining the ZAP approximation and flow deviation technique for optimal routing appear to still need more validation with backup simulation.

Finally, we must address the issues of mobile operation inherent in multi-hop packet radio networks. So far, no theoretical work seems to have appeared on the effects of mobility on the network capacity, let alone the packet delay. We have not identified what features involved in the mobility are decisive to the network performance. The speed of a PRU may be related with the probability that a station loses its track which induces re-initialization delay. On the other hand, if the emission of ROP occurs too often, these packets will flood the network, causing the degradation of effective throughput. It is of interest to investigate the trade-offs involved in the speed of PRUs, the frequency of ROP emission and the density of monitoring repeater or stations.

APPENDICES

A. Derivation of Eq.(3.4)

Let $F(x)$ be the z-transform of the distribution of $D(t)$, the number of departures from all users during $[0,t]$. The joint probability that during this interval user i attains $D_i(t) = k_i$ departures ($i = 1,2,\cdots,M$) given that $D(t) = k$ is given by a multinomial distribution

$$\frac{k!}{k_1!\, k_2! \cdots k_M!}\, q_1^{k_1} q_2^{k_2} \cdots q_M^{k_M} \tag{A.1}$$

where $k = k_1 + k_2 + \cdots + k_M$, and q_i is the probability that a departure is achieved by user i. Thus, the conditional joint z-transform of $\mathbf{D}(t) = [\, D_1(t), D_2(t), \cdots, D_M(t)\,]$ is given by

$$(q_1 z_1 + q_2 z_2 + \cdots + q_M z_M)^k \tag{A.2}$$

and the unconditional joint z-transform of $\mathbf{D}(t)$ is given by a compound distribution

$$G(\mathbf{z}) = F(q_1 z_1 + q_2 z_2 + \cdots + q_M z_M) \tag{A.3}$$

where $\mathbf{z} = [\, z_1, z_2, \cdots, z_M\,]$.

Now, from Eq.(A.3) and the definition of $F(z)$, the means and covariances of $\mathbf{D}(t)$ can be formally calculated. First, the mean is given by

$$\overline{D_i(t)} = \frac{\partial G(\mathbf{z})}{\partial z_i}\bigg|_{\mathbf{z}=\mathbf{1}} = q_i \frac{dF(z)}{dz}\bigg|_{z=1} = q_i\, \overline{D(t)} \qquad i = 1,2,\cdots,M \tag{A.4}$$

where $\mathbf{1} = [1,1,\cdots,1]$. Next, from

$$\overline{D_i(t)^2} - \overline{D_i(t)} = \frac{\partial^2 G(\mathbf{z})}{\partial z_i^2}\bigg|_{\mathbf{z}=\mathbf{1}} = q_i^2\, [\, \overline{D(t)^2} - \overline{D(t)}\,]\,, \tag{A.5}$$

the variance is given by

$$\begin{aligned} Var[D_i(t)] &\triangleq \overline{D_i(t)^2} - \overline{D_i(t)}^2 \\ &= q_i^2\, [\, \overline{D(t)^2} - \overline{D(t)}\,] + q_i \overline{D(t)} - q_i^2 \overline{D(t)}^2 \\ &= q_i^2\, Var[D(t)] + q_i(1-q_i)\, \overline{D(t)} \qquad i = 1,2,\cdots,M \end{aligned} \tag{A.6}$$

Also, for $i \neq j$, from

$$\overline{D_i(t)\cdot D_j(t)} = \frac{\partial^2 G(\underline{z})}{\partial z_i \partial z_j}\Big|_{\underline{z}=1} = q_i q_j [\overline{D(t)^2} - \overline{D(t)}] , \tag{A.7}$$

the covariance is given by

$$\begin{aligned} Cov[D_i(t), D_j(t)] &\triangleq \overline{D_i(t)\cdot D_j(t)} - \overline{D_i(t)}\cdot\overline{D_j(t)} \\ &= q_i q_j [\overline{D(t)^2} - \overline{D(t)}] - q_i q_j \overline{D(t)}^2 \\ &= q_i q_j [Var[D(t)] - \overline{D(t)}] \qquad i \neq j, \quad i,j = 1,2,\cdots,M \end{aligned} \tag{A.8}$$

Eqs.(A.4), (A.7) and (A.8) give Eq.(3.4).

B. Derivation of Eq.(3.22)

In an infinite population of pure ALOHA users, the duration of an unsuccessful transmission period F consists of an indefinite number (L, say) of packet interarrival times whose durations are less than 1 (denoted by $f^{(1)}, f^{(2)}, \cdots, f^{(L)}$) terminated by a full length of 1; see Figure 3.2(b):

$$F = f^{(1)} + f^{(2)} + \cdots + f^{(L)} + 1 \tag{B.1}$$

All $f^{(n)}$'s are independent and identically distributed (let their generic representation be f with its pdf's Laplace transform $f^*(s)$) as

$$Prob[f \leq t] = \frac{1 - e^{-Gt}}{1 - e^{-G}} \qquad 0 \leq t \leq 1$$

$$f^*(s) = \frac{G[1 - e^{-(s+G)}]}{(s+G)(1-e^{-G})} , \tag{B.2}$$

$$\bar{f} = \frac{1}{G} - \frac{e^{-G}}{1-e^{-G}} \; ; \quad Var[f] = \frac{1}{G^2} - \frac{e^{-G}}{(1-e^{-G})^2}$$

where G is the rate of arrivals. The number of such interarrival times L (with its distribution's z-transform $L^*(z)$) is independent of $f^{(n)}$'s and is geometrically distributed as

$$Prob[L = n] = (1 - e^{-G})^{n-1} e^{-G} \qquad n = 1,2,\cdots$$

$$L^*(z) = \frac{z\, e^{-G}}{1 - z(1 - e^{-G})} ,$$

$$\bar{L} = e^G \; ; \quad Var[L] = e^G (e^G - 1) \tag{B.3}$$

From Eqs.(B.1)-(B.3), we have

$$F^*(s) = e^{-s} L^*[f^*(s)] = \frac{G\, e^{-(s+G)}[1 - e^{-(s+G)}]}{(1-e^{-G})[s + G\, e^{-(s+G)}]} ,$$

$$\bar{F} = \bar{L}\,\bar{f} + 1 = \frac{e^G - 1 - G\,e^{-G}}{G\,(1-e^{-G})},$$

$$Var[F] = \bar{L}\,Var[f] + \bar{f}^2\,Var[L] = \frac{e^{2G}}{G^2} - \frac{2\,e^G}{G} - \frac{e^{-G}}{(1-e^{-G})^2} \qquad \text{(B.4)}$$

Eq.(B.4) is identical to Eq.(3.22).

C. Derivation of Eqs.(3.45) and (3.47)

For the sake of brevity, let us call the start of transmission as the arrival of a packet, and let time $t = 0$ be the arrival time of an idle-period-breaking packet. Let n_1 and n_2 be the numbers of arrivals during time intervals $[0,c]$ and $[c,a]$, respectively ($0 \leqslant n_1 + n_2 \leqslant M-1$). Note that during $[0,a]$ each user behaves independently of all others because (by definition of CSMA) he does not hear the transmission until time a passes. Our objective is to find $Prob[\,Y \leqslant y \mid n_1 = 0\,]$ (successful case) and $Prob[\,Y \leqslant y \mid n_1 \geqslant 1\,]$ (unsuccessful case) for $0 \leqslant y \leqslant a$, where Y is the arrival time of the last overlapping packet; see Figure 3.4.

We first consider the case of a successful transmission which occurs with probability $Prob[\,n_1 = 0\,] = e^{-g(M-1)} = \gamma$ (see Eq.(3.43)). We consider the two subcases: case (i) $n_2 = 0$ with probability $e^{-g(a-c)(M-1)}$, and case (ii) $1 \leqslant n_2 \leqslant M-1$ with probability $1 - e^{-g(a-c)(M-1)}$. In case (i), we let $Y = 0$:

$$Prob[\,Y \leqslant y \mid n_1 = n_2 = 0\,] = 1 \qquad 0 \leqslant y \leqslant a \qquad \text{(C.1)}$$

In case (ii), definitely $Y > c$. The probability that n_2 arrivals fall in $[c,y]$, $c < y < a$, while the other $M-1-n_2$ do not arrive during $[c,a]$ is given by

$$\binom{M-1}{n_2}\,[1-e^{-g(y-c)}]^{n_2}\,e^{-g(a-c)(M-1-n_2)}$$

Unconditioning this expression on $n_2 \geqslant 1$, we have

$$Prob[\,Y \leqslant y \mid n_1 = 0,\, n_2 \geqslant 1\,] = \begin{cases} 0 & 0 \leqslant y \leqslant c \\ \dfrac{[1-e^{-g(y-c)}+e^{-g(a-c)}]^{M-1} - e^{-g(a-c)(M-1)}}{1-e^{-g(a-c)(M-1)}} & c \leqslant y \leqslant a \end{cases} \qquad \text{(C.2)}$$

Further unconditioning of Eqs.(C.1) and (C.2) over the cases (i) and (ii) gives

$$Prob[\,Y \leqslant y \mid n_1 = 0\,] = \begin{cases} e^{-g(a-c)(M-1)} & 0 \leqslant y \leqslant c \\ [1-e^{-g(y-c)}+e^{-g(a-c)}]^{M-1} & c \leqslant y \leqslant a \end{cases} \qquad \text{(C.3)}$$

We next consider the case of an unsuccessful transmission which occurs with probability $Prob[n_1 \geq 1] = 1-\gamma$. For $y \leq c$, the probability that n_1 arrivals fall in $[0,y]$ while none of $M-1-n_1$ arrive in $[0,a]$ is given by

$$\binom{M-1}{n_1}(1-e^{-\pi y})^{n_1} e^{-\pi a(M-1-n_1)}$$

Unconditioning this expression on $n_1 \geq 1$, we have

$$Prob[Y \leq y \mid n_1 \geq 1] = \frac{(1-e^{-\pi y}+e^{-\pi a})^{M-1} - e^{-\pi a(M-1)}}{1-\gamma} \qquad 0 \leq y \leq c \qquad \text{(C.4)}$$

For $c \leq y \leq a$, assuming $n_1 \geq 1$ arrivals in $[0,c]$ with probability

$$\binom{M-1}{n_1}(1-e^{-\pi c})^{n_1} e^{-\pi c(M-1-n_1)},$$

we consider the two cases: case (i) $n_2 = 0$ with probability $e^{-\pi(a-c)(M-1-n_1)}$, and case (ii) $n_2 \geq 1$. In case (i), clearly

$$Prob[Y \leq y \mid n_2 = 0, \text{ given } n_1] = 1 \qquad \text{(C.5)}$$

In case (ii), the probability that n_2 arrivals fall in $[c,y]$ while none of $M-1-n_1-n_2$ arrive during $[c,a]$ is given by

$$\binom{M-1-n_1}{n_2}[1-e^{-\pi(y-c)}]^{n_2} e^{-\pi(a-c)(M-1-n_1-n_2)}$$

Unconditioning this expression on $n_2 \geq 1$ gives

$$Prob[Y \leq y \mid n_2 \geq 1, \text{ given } n_1] = \frac{[1-e^{-\pi(y-c)}+e^{-\pi(a-c)}]^{M-1-n_1} - e^{-\pi(a-c)(M-1-n_1)}}{1-e^{-\pi(a-c)(M-1-n_1)}} \qquad \text{(C.6)}$$

Further unconditioning of Eqs.(C.5) and (C.6) over the two cases (i) and (ii) given $n_1 \geq 1$ yields

$$Prob[Y \leq y \mid \text{given } n_1] = [1-e^{-\pi(y-c)}+e^{-\pi(a-c)}]^{M-1-n_1} \qquad \text{(C.7)}$$

Finally, by unconditioning Eq.(C.7) on $n_1 \geq 1$, we get

$$Prob[Y \leq y \mid n_1 \geq 1] = \frac{(1-e^{-\pi y}+e^{-\pi a})^{M-1} - (e^{-\pi c}-e^{-\pi y}+e^{-\pi a})^{M-1}}{1-\gamma} \qquad c \leq y \leq a \qquad \text{(C.8)}$$

Eqs.(C.3), (C.4) and (C.8) complete the derivation.

D. Determination of the Transmission Rate for CSMA

This appendix is associated with Section 6.4. Here we derive the relation between transmission rate p' and channel-sensing rate p for slotted nonpersistent CSMA. As in [Toba75], we assume that

$$p' = p P_I \tag{D.1}$$

where P_I is the probability that the channel is sensed idle. Since the probability of an empty slot is given by $e^{-p'N}$, the expected value of the idle period I (see Figure 6.7) is

$$\bar{I} = \sum_{k=1}^{\infty} kae^{-kp'N}(1-e^{-p'N}) = \frac{ae^{-p'N}}{1-e^{-p'N}} \tag{D.2}$$

On the other hand, the transmission period is $1+a$. Therefore,

$$P_I = \frac{\bar{I}}{1+a+\bar{I}} = \frac{a\,e^{-p'N}}{1+a-e^{-p'N}} \tag{D.3}$$

Thus we have obtained an equation which determines p' in terms of p:

$$p' = \frac{ap\,e^{-p'N}}{1+a-e^{-p'N}} \tag{D.4}$$

For $a \ll 1$, we have

$$p' = \frac{p}{1+(p'/a)N} \tag{D.5}$$

which explicitly gives p'.

Using the optimal values in this case ($N=5.3$ and $p'/a=0.20$), we see that the actual transmission rate is 41% of the sensing rate.

E. A Numerical Technique to Solve a Large Sparse Markov Chain

This appendix is associated with Chapter 7. Here we discuss certain computational aspects of solving a large system of linear simultaneous equations for a Markov chain. The chain equations are of the form

$$\pi_i = \sum_{j=1}^{N} \pi_j P_{ji} \qquad i=1,2,\cdots,N \tag{E.1}$$

$$\sum_{i=1}^{N} \pi_i = 1 \tag{E.2}$$

where P_{ij} is the steady-state transition probability from state i to j, and π_i is the steady-state probability of the system being at state i ($i,j=1,2,\cdots,N$). Note that one of the equations in Eq.(E.1) is redundant.

A fundamental method for solving the above type of system of equations is the Gauss elimination method [Gera78]. However, since this method requires $O(N^3)$ operations and $O(N^2)$ storage cells, it is infeasible for systems with large N. On the other hand, it turns out in our problems that most elements of the matrix $\underline{P} \equiv (P_{ij})$ are null, i.e., $\underline{P}$ is a very sparse matrix. In Table E.1 below, we show the number of distinct states (N) and the percentage of nonzero elements in matrix $\underline{P}$ for some of the cases we have dealt with in Chapter 7. We note that the Gauss elimination method which loses this sparsity due to additions of rows of $\underline{P}$ cannot take advantage of the sparse matrix representation which saves computational time and storage. Thus we are led to an iteration method; specifically we have adapted the Gauss-Seidel iteration method [Gera78] to our problems in the following way. (Some other methods may be found in [Wall74].)

Let $\{\pi_i^{(n)}\}$ be the nth iterative solution to the system given by Eqs.(E.1) and (E.2). First put

$$\pi_i^{(0)} = 0 \qquad i = 2, 3, \cdots, N$$

$$\pi_1^{(0)} = 1 \tag{E.3}$$

Choose k such that

$$k \triangleq \{i \mid P_{kk} = \max_{1 \le i \le N} P_{ii}\} \tag{E.4}$$

Then, for $n \geq 1$, for every $i (\neq k)$,

$$\pi_i^{(n+1)} = \frac{1}{1 - P_{ii}} \left[\sum_{j=1}^{i-1} P_{ji} \pi_j^{(n+1)} + \sum_{j=i+1}^{N} P_{ji} \pi_j^{(n)} \right] \tag{E.5}$$

and for $i = k$,

$$\pi_k^{(n+1)} = 1 - \sum_{j=1}^{k-1} \pi_j^{(n+1)} - \sum_{j=k+1}^{N} \pi_j^{(n)} \tag{E.6}$$

A sufficient condition for this iteration to converge whatever $\{\pi_i^{(0)}\}$ may be used initially is given by

$$\sum_{j=1}^{N} P_{ji} < 1 \qquad i = 1, 2, \cdots, N \tag{E.7}$$

Actually this condition is not always met in our problems. However, since this is a sufficient condition and not a necessary condition, there are cases where the iteration converges when this condition is not met. In fact, we have had convergence for most cases in our calculation. Our first selection of the index k is intended to approach the above sufficient condition.

Note that by using the Gauss-Seidel iteration we have only to store the nonzero values of $\underline{P}$. This saves storage (which provides fast computation in virtual storage systems) as well as reducing the number of necessary operations. In fact, we estimate the latter as

O(Number of nonzero elements in $\underline{P}$) × O(Number of iterations)

Here the number of needed iterations depends on the starting values $\{\pi_i^{(0)}\}$ and the criterion for stopping the iteration. Using the initial values $\pi_1^{(0)} = 1$ and $\pi_i^{(0)} = 0$ for $2 \leqslant i \leqslant N$ (state 1 corresponds to the state that all the buffers in the network are empty) and the stopping criterion $\max_i |\pi_i^{(n+1)} - \pi_i^{(n)}| < 10^{-4}$, we have obtained convergence with fewer than 100 iterations in most cases.

Thus we believe that the above method is quite time- and storage-efficient for our problems. As a matter of fact, for $N = 3456$ (network 2, $m = 1$), it took less than one minute to solve a system of equations of the form of Eqs.(E.1) and (E.2) on a VAX-11/780 at the UCLA Computer Science Department. So it seems that a major time-consuming part in our calculation of throughputs and delays is now constructing the transition probability matrix $\underline{P}$ which involves enumerating all possible events which can occur for every state of the network and determining the resulting next state for each of these events.

Table E.1. The number of states and the sparseness of $\underline{P}$ for some network models depicted in Figure 7.1. Suppression or acceleration of transmission is not employed. m = number of buffers in each repeater.

cases		number	nonzero elements in $\underline{P}$	
network	m	of states	number	percentage
1	1	144	1190	5.74
1	2	1176	12424	0.89
1	3	7200	83412	0.16
2	1	3456	76377	0.64

REFERENCES

[Abra70] N.Abramson, "The ALOHA System — Another Alternative for Computer Communications," *AFIPS Conference Proceedings, 1970 Fall Joint Computer Conference,* Vol.37, pp.281-285.

[Abra73a] N.Abramson, "The Aloha System," in *Computer-Communication Networks,* N.Abramson and F.F.Kuo (eds.), Prentice-Hall, 1973.

[Abra73b] N.Abramson, "Packet Switching with Satellites," *AFIPS Conference Proceedings, 1973 National Computer Conference,* Vol.42, pp.695-702.

[Abra77] N.Abramson, "The Throughput of Packet Broadcasting Channels," *IEEE Transactions on Communications,* Vol.COM-25, No.1, pp.117-128, January 1977.

[Akav79] G.Akavia and L.Kleinrock, "Performance Tradeoffs and Hierarchical Designs of Distributed Packet-Switching Communication Networks," Computer Science Department, School of Engineering and Applied Science, University of California, Los Angeles, Engineering Report UCLA-ENG-7952, September, 1979.

[Arth82] E.Arthurs and B.W.Stuck, "A Theoretical Performance Analysis of Polling and Carrier Sense Collision Detection Communication Systems," in *Local Computer Networks,* P.C.Ravasio, G.Hopkins and N.Naffah (eds.), North-Holland Publishing Company, 1982.

[Bake81] D.J.Baker and A.Ephremides, "The Architectural Organization of a Mobile Radio Network via Distributed Algorithm," *IEEE Transactions on Communications,* Vol.COM-29, No.11, pp.1694-1701, November 1981.

[Beve70] G.S.Beveridge and R.S.Schecter, *Optimization: Theory and Practice,* McGraw-Hill Book Company, 1970.

[Bhar80] K.Bharath-Kumar, "Discrete-Time Queueing Systems and Their Networks," *IEEE Transactions on Communications,* Vol.COM-28, No.2, pp.260-263, February 1980.

[Boor80] R.R.Boorstyn and A.Kershenbaum, "Throughput Analysis of Multihop Packet Radio," *ICC'80 Conference Record,* IEEE International Conference on Communications, 80CH1505-6, Seattle, WA., pp.13.6.1-13.6.6, June 8-12, 1980.

[Carl75] A.B.Carleial and M.E.Hellman, "Bistable Behavior of ALOHA-Type Systems," *IEEE Transactions on Communications,* Vol.COM-23, No.4, pp.401-409, April 1975.

[Cox54] D.R.Cox and W.L.Smith, "On the Superposition of Renewal Processes," *Biometrika,* Vol.41, Parts 1 and 2, pp.91-99, June 1954.

[Davi80] D.H.Davis and S.A.Gronemeyer, "Performance of Slotted ALOHA Random Access with Delay Capture and Randomized Time of Arrival," *IEEE Transactions on Communications,* Vol.COM-28, No.5, pp.703-710, May 1980.

[Dubn68] H.Dubner and J.Abate, "Numerical Inversion of Laplace Transforms by Relating Them to the Finite Fourier Cosine Transforms," *Journal of the Association for Computing Machinery,* Vol.15, No.1, pp.115-123, January 1968.

[Ehrl79] N.Ehrlich, "The Advanced Mobile Phone Service," *IEEE Communications Magazine,* Vol.17, No.2, pp.9-15, March 1979.

[Elsa82] M.Elsanadidi and W.W.Chu, "Study of Acknowledgment Schemes in a Star Multiaccess Network," *IEEE Transactions on Communications,* Vol.COM-30, No.7, pp.1657-1667, July 1982.

[Ferg77a] M.J.Ferguson, "A Bound and Approximation of Delay Distribution for Fixed-Length Packets in an Unslotted ALOHA Channel and a Comparison with Time Division Multiplexing (TDM)," *IEEE Transactions on Communications,* Vol.COM-25, No.1, pp.136-139, January 1977.

[Ferg77b] M.J.Ferguson, "An Approximate Analysis of Delay for Fixed and Variable Length Packets in Unslotted ALOHA Channel," *IEEE Transactions on Communications,* Vol.COM-25, No.7, pp.644-654, July 1977.

[Frat73] L.Fratta, M.Gerla and L.Kleinrock, "The Flow Deviation Method: An Approach to Store-and-Forward Communication Network Design," *Networks,* Vol.3, Issue 2, pp.97-133, 1973.

[Frat80] L.Fratta and D.Sant, "Some Models of Packet Radio Networks with Capture," *Proceedings of the Fifth International Conference on Communication,* pp.155-161, October 27-30, 1980.

[Fuku78] A.Fukuda, K.Mukumoto and T.Hasegawa, "Adaptive Retransmission Randomization Schemes for a Packet-Switched Random-Access Broadcast Channel," *Proceedings of the Fourth International Conference on Computer Communication,* pp.543-548, Kyoto, September 1978.

[Gelo80] E.Gelenbe and I.Mitrani, *Analysis and Synthesis of Computer Systems,* Academic Press, 1930.

[Gera78] C.F.Gerald, *Applied Numerical Analysis,* Second edition, Addison-Wesley Publishing Company, Inc., 1978.

[Gerl80] M.Gerla and L.Kleinrock, "Flow Control: A Comparative Survey," *IEEE Transactions on Communications,* Vol.COM-28, No.4, pp.553-574, April 1980.

[Gitm76] I.Gitman, R.M.Van Slyke and H.Frank, "Routing in Packet-Switching Broadcast Radio Networks," Broadcast Radio Networks,' *IEEE Transactions on Communications,* Vol.COM-24, No.8, pp.925-930, August 1976.

[Hale80] W.K.Hale, "Frequency Assignment: Theory and Application," *Proceedings of the IEEE,* Vol.68, No.12, pp.1497-1514, December 1980.

[Hans79] L.W.Hansen and M.Schwartz, "An Assigned-Slot Listen-Before-Transmission Protocol for a Multiaccess Data Control," *IEEE Transactions on Communications,* Vol.COM-27, No.6, pp.846-857, June 1979.

[Heym82a] D.P.Heyman and M.J.Sobel, *Stochastic Models in Operations Research, Volume I: Stochastic Processes and Operating Characteristics,* McGraw-Hill Book Company, 1982.

[Heym82b] D.P.Heyman, "An Analysis of the Carrier-Sense Multiple-Access Protocol," *The Bell System Technical Journal,* Vol.61, No.8, pp.2023-2051, October 1982.

[Humb82] P.Humblet, "Rational Routing in PR Networks," UCLA Packet Radio Analytical Workshop, August 10-11, 1982.

[Ito77] Y.Ito, "Superposition of Distinguishable Point Processes," *Journal of Applied Probability,* Vol.14, No.1, pp.200-204, March 1977.

[Ito78] Y.Ito, "Superposition and Decomposition of Stationary Point Processes," *Journal of Applied Probability,* Vol.15, No.3, pp.481-493, September 1978.

[Jenq82] Y.C.Jenq, "Theoretical Analysis of Slotted ALOHA, CSMA and CSMA-CD Protocols," *Conference Record, GLOBECOM '82,* IEEE Global Telecommunications Conference, pp.A2.1.1-A2.1.5, Miami, November 29 - December 2, 1982.

[Jubi81] J.C.Jubin, N.E.Gower and T.J.Quilici, "Improved Packet Radio Level 6 Protocol Program," Collins Communications Systems Division, Rockwell International, Dallas, Texas 75207, May 22, 1981.

[Kahn78] R.E.Kahn, S.A.Gronemeyer, J.Burchfiel and R.C.Kunzelman, "Advances in Packet Radio Technology," *Proceedings of the IEEE,* Vol.66, No.11, pp.1468-1496, November 1978.

[Karl75] S.Karlin and H.W.Taylor, *A First Course in Stochastic Processes*, Second Edition, Academic Press, 1975.

[Klei73] L.Kleinrock and S.S.Lam, "Packet Switching in a Slotted Satellite Channel," *AFIPS Conference Proceedings*, 1973 National Computer Conference, Vol.42, pp.703-710, June 1973.

[Klei75a] L.Kleinrock, *Queueing Systems, Volume 1: Theory*, John Wiley and Sons, 1975.

[Klei75b] L.Kleinrock and F.A.Tobagi, "Packet Switching in Radio Channels: Part I - Carrier Sense Multiple-Access Modes and Their Throughput-Delay Characteristics," *IEEE Transactions on Communications*, Vol.COM-23, No.12, pp.1400-1416, December 1975.

[Klei76] L.Kleinrock, *Queueing Systems, Volume 2: Computer Applications*, John Wiley and Sons, 1976.

[Klei77] L.Kleinrock, "Performance of Distributed Multi-Access Computer-Communication Systems," *Information Processing 77, Proceedings of IFIP Congress 77*, pp.547-552, Toronto, August 8-12, 1977, North-Holland Publishing Company.

[Klei78] L.Kleinrock and J.A.Silvester, "Optimum Transmission Radii for Packet Radio Networks or Why Six is a Magic Number," *Conference Record, National Telecommunications Conference*, pp.4.3.1-4.3.5, December 1978.

[Klei80] L.Kleinrock and Y.Yemini, "Interfering Queueing Process in Packet Switched Broadcast Communication," *Information Processing 80, Proceedings of IFIP Congress 80*, Tokyo, October 1980, pp.557-562, North-Holland Publishing Company, Amsterdam, 1980.

[Koba74] H.Kobayashi, "Application of the Diffusion Approximation to Queueing Networks I: Equilibrium Queue Distributions," *Journal of the Association for Computing Machinery*, Vol.21, No.2, pp.316-328, April 1974.

[Koba77] H.Kobayashi, Y.Onozato and D.Huynh, "An Approximation Method for Design and Analysis of of an ALOHA System," *IEEE Transactions on Communications*, Vol.COM-25, No.1, pp.148-157, January 1977.

[Kueh79] P.J.Kuehn, "Approximate Analysis of General Queuing Networks by Decomposition," *IEEE Transactions on Communications*, Vol.COM-27, No.1, pp.113-126, January 1979.

[Kues73] J.L.Kuester and J.H.Mize, *Optimization Techniques with Fortran*, McGraw-Hill Book Company, 1973.

[Kung81] K.C.Kung, "Flow Deviation Algorithm in a Multihop Packet Radio Network," Private communication. Working paper 81006, 1981.

[Kunz81] R.C.Kunzelman, V.D.Cone, K.S.Klemba, J.E.Mathis, D.L.Nielson and P.Spilling, "Packet Radio Network and Internetwork Research and Development," Final Technical Report Covering the Period November 1, 1977 to February 28, 1981, SRI International, May 1981.

[Lam74] S.S.Lam, "Packet Switching in a Multi-Access Broadcast Channel with Applications to Satellite Communication in a Computer Network," Computer Science Department, School of Engineering and Applied Science, University of California, Los Angeles, Engineering Report UCLA-ENG-7429, March 1974.

[Lam75] S.S.Lam and L.Kleinrock, "Packet Switching in a Multiaccess Broadcast Channels: Dynamic Control Procedures," *IEEE Transactions on Communications,* Vol.COM-23, No.9, pp.891-904, September 1975.

[Lam80] S.S.Lam, "A Carrier-Sense Multiple Access Protocol for Local Networks," *Computer Networks,* Vol.4, No.1, pp.21-32, 1980.

[Lawr73] A.J.Lawrance, "Dependency of Intervals Between Events in Superposition Processes," *Journal of the Royal Statistical Society,* Ser.B, Vol.35, No.2, pp.306-315, 1973.

[Laws80] M.Lawson and S.Smith, "Reconfiguration Techniques of a Mobile Network," *Proceedings 1980 International Zurich Seminar Digital Communication,* Zurich, Switzerland, IEEE 80CH1521-4 COM, pp.B10.1-B10.5.

[Lee82] I.Lee and J.A.Silvester, "An Iterative Scheme for Performance Modeling of Slotted ALOHA Packet Radio Networks," *Proceedings of the International Conference on Communications,* pp.6G.1.1-6G.1.5, Philadelphia, June 1982.

[Lein80] B.M.Leiner, "A Simple Model for Computation of Packet Radio Network Communication Performance," *IEEE Transactions on Communications,* Vol.COM-28, No.12, pp.2020-2023, December 1980.

[Litt61] J.D.C.Little, "A Proof for the Queuing Formula: $L = \lambda W$," *Operations Research,* Vol.9, No.3, pp.383-387, May-June 1961.

[Metc76] R.M.Metcalfe and D.R.Boggs, "Ethernet: Distributed Packet Switching for Local Computer Networks," *Communications of the ACM,* Vol.19, No.7, pp.395-404, July 1976.

[Mino78] D.Minoli and I.Gitman, "Combinatorial Issues in Mobile Packet Radio Networks," *IEEE Transactions on Communications,* Vol.COM-26, No.12, pp.1821-1826, December 1978.

[Mino79a] D.Minoli and I.Gitman, "Monitoring Mobile Packet Radio Devices," *IEEE Transactions on Communications*, Vol.COM-27, No.2, pp.509-517, February 1979.

[Mino79b] D.Minoli, "Exact Solution for the Initialization Time of Packet Radio Networks with Two Station Buffers," *AFIPS Conference Proceedings*, Vol.48, 1979 National Computer Conference, pp.875-885, June 1979.

[Mino79c] D.Minoli, I.Gitman and D.H.Walters, "Analytical Model for Initialization of Single-Hop Packet Radio Networks," *IEEE Transactions on Communications*, Vol.COM-27, No.12, pp.1959-1967, December 1979.

[Moll81] M.L.Molle, "Unifications and Extensions of the Multiple Access Communications Problem," Computer Science Department, School of Engineering and Applied Science, University of California, Los Angeles, Engineering Report UCLA-ENG-8118, CSD-810730, July 1981

[Nels82] R.Nelson, "Channel Access Protocols for Multi-Hop Broadcast Packet Radio Networks," Computer Science Department, School of Engineering and Applied Science, University of California, Los Angeles, Engineering Report UCLA-ENG-8259, CSD-820731, July 1982.

[Nils80] A.Nilsson, W.Chou and C.J.Graff, "A Packet Radio Communication System Architecture in a Mixed Traffic and Dynamic Environment," *Proceedings Computer Networking Symposium*, IEEE Computer Society Press, CH1586-7/80, pp.51-66, 1980.

[Otne78] R.K.Otnes and L.Enochson, *Applied Time Series Analysis, Volume I: Basic Techniques*, John Wiley and Sons, 1978.

[Perl80] R.Perlman (revised by J.Westcott and J.Jubin), "Multistation and CAP6 Routing Implementation Specification," Packet Radio Temporary Note 289 (Revision 1), October 1980.

[Quil79] T.J.Quilici and M.H.Cisco, *Experimental Packet Radio (EPR) Users Guide*, Collins Communications Systems Division, Rockwell International, Dallas, Texas 75207, December 5, 1979.

[Robe72] L.G.Roberts, "ALOHA Packet System with and without Slots and Capture," ARPA Satellite System Note 8 (NIC 11290), ARPA Network Information Center, Stanford Research Institute, Menlo Park, California, June 26, 1972. Reprinted in ACM SIGCOMM *Computer Communication Review*, Vol.5, No.2, pp.28-42, April 1975.

[Saad81] T.N.Saadawi and A.Ephremides, "Analysis, Stability, and Optimization of Slotted ALOHA with a Finite Number of Buffered Users," *IEEE Transactions on Automatic Control*, Vol.AC-26, No.3, pp.680-689, June 1981.

[Sevc77] K.C.Sevcik, A.I.Levy, S.K.Tripathi and J.L.Zahorjan, "Improving Approximations of Aggregated Queueing Network Systems," *Computer Performance*, K.M.Chandy and M.Reiser (eds.), North-Holland, 1977.

[Shac82] N.Shacham, "Throughput Delay Performance of Packet-Switching Multiple-Access Channel with Power Capture," submitted to *Performance Evaluation*. Also, "Performance Analysis of a Multiple Access Channel with Heterogeneous Traffic and Capture," *Conference Record, GLOBECOM '82*, IEEE Global Telecommunications Conference, pp.B2.4.1-B2.4.5, Miami, November 29 - December 2, 1982.

[Sidi81] M.Sidi and A.Segall, "A Busy-Tone-Multiple-Access-Type Scheme for Packet-Radio Networks," *Performance of Data Communication Systems and Their Applications*, G.Pujolle (ed.), pp.1-10, North-Holland Publishing Company, 1981.

[Sidi83] M.Sidi and A.Segall, "Two Interfering Queues in Packet-Radio Networks," *IEEE Transactions on Communications*, Vol.COM-31, No.1, pp.123-129, January 1983.

[Silv80] J.A.Silvester, "On the Spatial Capacity of Packet Radio Networks," Computer Science Department, School of Engineering and Applied Science, University of California, Los Angeles, Engineering Report UCLA-ENG-8021, May 1980.

[Silv82a] J.A.Silvester and I.Lee, "Performance Modeling of Multi-Access Computer Communications Network," *Proceedings of Compcon*, September 1982.

[Silv82b] J.A.Silvester and I.Lee, "Performance Modeling of Buffered CSMA — An Iterative Approach," *Conference Record, GLOBECOM '82*, IEEE Telecommunications Conference, pp.F1.6.1-F1.6.5, Miami, November 29 - December 2, 1982.

[Silv82c] J.A.Silvester and I.Lee, "An Approximate Model for Delay Performance in Multi-Hop Broadcast Networks," UCLA Packet Radio Analytical Workshop, August 10-11, 1982.

[Tasa82] S.Tasaka and Y.Ishibashi, "Stability and Performance of a Packet Reservation System for Satellite Broadcast Channels," *Conference Record, GLOBECOM '82*, IEEE Global Telecommunications Conference, pp.A2.3.1-A2.3.8, Miami, November 29 - December 2, 1982.

[Toba74] F.A.Tobagi, "Random Access Techniques for Data Transmission over Packet Switched Radio Networks," Computer Science Department, School of Engineering and Applied Science, University of California, Los Angeles, Engineering Report UCLA-ENG-7499, December 1974.

[Toba75] 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 Communications*, Vol.COM-23, No.12, pp.1417-1433, December 1975.

[Toba76] F.A.Tobagi and L.Kleinrock, "On the Analysis of Buffered Packet Radio Systems," *Proceedings of the Ninth Hawaii International Conference on System Sciences,* Honolulu, Hawaii, pp.42-45, January 1976.

[Toba77] 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 Communications,* Vol.COM-25, No.10, pp.1103-1120, October 1977.

[Toba80a] F.A.Tobagi, "Analysis of a Two-Hop Centralized Packet Radio Network — Part I: Slotted ALOHA," *IEEE Transactions on Communications,* Vol.COM-28, No.2, pp.196-207, February 1980.

[Toba80b] F.A.Tobagi, "Analysis of a Two-Hop Centralized Packet Radio Network — Part II: Carrier Sense Multiple Access," *IEEE Transactions on Communications,* Vol.COM-28, No.2, pp.208-216, February 1980.

[Toba80c] F.A.Tobagi, "Multiaccess Protocols in Packet Communication Systems," *IEEE Transactions on Communications,* Vol.COM-28, No.4, pp.468-488, April 1980.

[Toba80d] F.A.Tobagi and V.B.Hunt, "Performance Analysis of Carrier Sense Multiple Access with Collision Detection," *Computer Networks,* Vol.4, No.5, pp.245-259, October-November 1980.

[Toba82a] F.A.Tobagi, "Distributions of Packet Delay and Interdeparture Time in Slotted ALOHA and Carrier Sense Multiple Access," *Journal of the Association for Computing Machinery,* Vol.29, No.4, pp.907-927, October 1982.

[Toba82b] F.A.Tobagi and J.M.Brazio, "Throughput Analysis of Multihop Packet Radio Networks under Various Channel Access Schemes," UCLA Packet Radio Analytical Workshop, August 10-11, 1982.

[Wall74] V.L.Wallace, "Algebraic Techniques for Numerical Solution of Queueing Networks," *Mathematical Methods in Queueing Theory,* A.B.Clarke, (ed.), Springer-Verlag, pp.295-305, 1974.

[West79] J.Westcott, "Issues in Congestion Control: Detection and Current Routing Design," Packet Radio Temporary Note 265, January 1979.

[West80] J.Westcott, "CAP5.6 Labeler," Bolt Beranek and Newman Inc., 50 Moulton Street, Cambridge, Massachusetts 02238, December 1980.

[Whit82] W.Whitt, "Approximating a Point Process by a Renewal Process, I: Two Basic Methods," *Operations Research,* Vol.30, No.1, pp.125-147, January-February 1982.

[Yemi80] Y.Yemini, "On Channel Sharing in Discrete-Time, Multi-Access Broadcast Communication," Computer Science Department, School of Engineering and Applied Science, University of California, Los Angeles, Engineering Report UCLA-ENG-8061, September 1980.

[Yeo78] G.F.Yeo, "On a Superposition of Alternating Renewal Processes in Traffic," *Advances in Applied Probability*, Vol.10, No.2, p.304, June 1978.