INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

Performance Analysis of Buffered Random Multiple Access Protocols

by

Tao Wan, B.Sc., M.Sc.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

Ottawa-Carleton Institute for Electrical Engineering
Faculty of Engineering

Department of Systems and Computer Engineering
Carleton University
Ottawa, Ontario, Canada K1S 5B6

September 17, 1999

© 1999, Tao Wan

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-48339-8

Canada

The undersigned recommend to the Faculty of Graduate Studies and Research
acceptance of the thesis

Performance Analysis of Buffered Random Multiple Access Protocols

submitted by

Tao Wan, B.Sc, M.Sc.

in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Thesis Supervisor

Thesis Supervisor

Chair, Department of Systems and Computer Engineering

External Examiner

Carleton University

September 17, 1999

Abstract

This thesis deals with the performance analysis of random multiple access (MA) systems, particularly the systems where each user has a buffer capacity for more than one customer (packet or message). A unified approach, called the Tagged User Approach (TUA), is proposed in this thesis for the approximate analysis of the performance of buffered random MA systems.

In buffered MA systems, each user is a queueing system whose customer service time is determined by the behaviors of all users in the system. However, among the available literatures on the performance analysis of these systems, only few of them employ existing queueing theory because of the difficulty to find the customer service time distribution. This difficulty is overcome in TUA where the state flow graph is introduced to obtain the virtual customer service time distribution for each user efficiently, which is then applied to the user queueing analysis using the classical queueing theory. Consequently, the analyses on channel contention scheme and user queueing behaviors are decoupled. This leads to the significant simplification of the system analysis compared to the available approaches. Hence, TUA is able to handle very complicated systems where the available approaches may fail to work. Moreover, TUA is able to analyze the stability of buffered MA systems. In this thesis, the stability of buffered S-ALOHA system is analyzed, for the first time, using TUA.

TUA is an analytical framework. Particular techniques may be required when TUA is applied to the performance analysis of a random MA system. The techniques for analyzing

several typical protocols such as ALOHA, CSMA/CD, R-ALOHA and PRMA have been developed in this thesis. Finally, the accuracy of the TUA analysis has been verified to be very good by simulation.

Acknowledgments

The author wishes to express his sincere appreciation to his previous thesis advisor Professor Asrar Sheikh for his generous encouragement, continued support and invaluable guidance throughout this work. The author would like to acknowledge his current thesis supervisors Professor Ioannis Lambadaris and Professor Mihail Devetsikiotis for their supports and encouragements. Special thanks to Professor David Falconer for his help in the absence of Professor Asrar Sheikh. Thanks are due to Paul Tse, Yonghai Gu and Atweri Abdulbaset at Nortel Wireless for their assistance during my three year internship with Nortel Networks. Thanks also to Professor Oliver Yang in the Department of Electrical Engineering at University of Ottawa for his concern and encouragement.

Thanks to all the faculty and staff members of the Department of Systems and Computer Engineering, Carleton University, for their kindness and support.

Financial assistance provided by Nortel Networks, the National Science and Engineering Research Council of Canada, and the Department of Systems and Computer Engineering at Carleton University during the period of this research is gratefully acknowledged.

Finally, the author is highly indebted to his wife Hong-Hui Xiong for her enormous patience, constant support and encouragement.

Table of Contents

Acknowledgments	iii
Abstract.....	iv
Table of Contents	vi
List of Figures.....	xii
List of Tables	xvi
List of Abbreviations	xvii
List of Symbols	xix
CHAPTER 1. INTRODUCTION.....	1
1.1 Multiple Access Protocols	1
1.1.1 Fixed Assignment	1
1.1.2 Random Access.....	2
1.1.3 Demand Assignment.....	3
1.1.3.1 Demand Assignment with Centralized Control	3
1.1.3.2 Demand Assignment with Distributed Control.....	3
1.1.4 CDMA Protocols	5
1.1.5 Adaptive Strategies and Mixed Modes.....	6
1.2 Performance Measurements.....	6
1.2.1 Channel Throughput	7
1.2.2 Response Time.....	7
1.3 Motivations	8
1.4 Thesis Contributions	10

1.5	Organization of the Thesis	12
1.6	Summary	15
CHAPTER 2.	LITERATURE REVIEW.....	16
2.1	Introduction	16
2.2	Analysis on Buffered Random Access Protocols	18
2.3	Review of Related Research	20
2.4	Summary	27
CHAPTER 3.	TAGGED USER ANALYSIS	29
3.1	Introduction	29
3.2	Model of A General System.....	32
3.3	Queueing Analysis	35
3.3.1	User Queueing Model	35
3.3.2	Distribution of Virtual Packet Service Time	35
3.3.2.1	Probability of Channel Sensed To Be Idle.....	38
3.3.3	Queueing Analysis of Tagged Users.....	43
3.3.3.1	Finite Buffer Capacity.....	44
3.3.3.2	Infinite Buffer Capacity	45
3.4	Performance Evaluation	47
3.4.1	Throughputs	48
3.4.2	Queue Length of Users	49
3.4.3	Waiting Time, Response Time and Access Delay.....	49
3.4.4	Interdeparture Time	51
3.4.4.1	Output Process of Users.....	51
3.4.4.2	Output Process of The Channel	52
3.5	Accuracy of TUA.....	55
3.6	Summary	61

CHAPTER 4. S-ALOHA SYSTEMS.....	63
4.1 Introduction	63
4.2 System Model	63
4.3 System Analysis.....	64
4.3.1 S-ALOHA with Finite Buffer Capacity	67
4.3.1.1 Blocking Probability and Service Grade.....	68
4.3.1.2 Channel Throughput	68
4.3.1.3 Queue Length.....	69
4.3.1.4 Waiting Time and Response Time.....	69
4.3.1.5 Output Process of The Channel	70
4.3.2 S-ALOHA with Infinite Buffer Capacity	71
4.3.3 Stability Analysis of S-ALOHA with Infinite Buffer Capacity.....	72
4.3.4 Equilibrium Point of S-ALOHA with Finite Buffer Capacity.....	81
4.4 Numerical Algorithm	84
4.4.1 Finite Buffer S-ALOHA	84
4.4.2 Infinite Buffered S-ALOHA	85
4.5 Numerical Results	85
4.5.1 System with Infinite Buffer Capacity	85
4.5.2 System with Finite Buffer Capacity.....	87
4.5.3 Accuracy of TUA on The Analysis of S-ALOHA.....	89
4.6 Summary	93
CHAPTER 5. CDMA S-ALOHA AND SATELLITE S-ALOHA SYSTEMS.....	95
5.1 CDMA S-ALOHA Systems.....	95
5.1.1 System Model	96
5.1.2 Contention Scheme Analysis	98
5.1.3 Queueing Analysis.....	99
5.1.4 Performance Analysis.....	101
5.1.5 Numerical Examples.....	104

5.1.5.1	Example 1	105
5.1.5.2	Example 2	107
5.2	Satellite ALOHA System	110
5.2.1	System Model	111
5.2.2	Contention Scheme Analysis	112
5.2.3	Queueing Analysis	114
5.2.4	Numerical Examples	115
5.3	Summary	120
CHAPTER 6.	BUFFERED CSMA/CD SYSTEM	122
6.1	Introduction	122
6.2	System Model	122
6.3	System Analysis Using TUA	124
6.3.1	Distribution of Virtual Packet Service Time	124
6.3.2	Queueing Analysis of Tagged User	126
6.3.2.1	Finite Buffer Capacity	126
6.3.2.2	Infinite buffer Capacity	127
6.4	Performance Evaluation	128
6.4.1	Finite Buffer System	128
6.4.2	Infinite Buffer System	129
6.5	The Algorithm	130
6.6	Numerical Results	131
6.6.1	Example 1: Homogeneous Systems	131
6.6.2	Example 2: Optimal CSMA/CD System	135
6.6.3	Example 3: CSMA System	138
6.6.4	Example 4: Heterogeneous Systems	141
6.7	Summary	146
CHAPTER 7.	BUFFERED R-ALOHA SYSTEMS	147
7.1	Introduction	147

7.2	System Model	149
7.3	System Analysis	150
7.3.1	Message Service Time Distribution.....	150
7.3.2	Queueing Analysis	152
7.4	Performance Evaluation	157
7.5	Numerical Algorithm	158
7.6	Numerical Examples	159
7.6.1	Example 1	159
7.6.2	Example 2	160
7.7	Summary	168
CHAPTER 8. PACKET RESERVATION MULTIPLE ACCESS (PRMA).....		170
8.1	Introduction	170
8.2	PRMA System Model	172
8.3	System Analysis	174
8.3.1	Analysis of Data Users.....	174
8.3.2	Analysis of Voice Users.....	177
8.3.2.1	Packet dropping probability	180
8.3.2.2	Talkspurt dropping probability	182
8.3.3	Analysis of integrated systems.....	183
8.4	Performance Evaluation	187
8.4.1	Data Users	187
8.4.2	Voice Users.....	188
8.4.3	Channel Throughput	188
8.5	Numerical Algorithm	189
8.6	Numerical Results	190
8.7	Summary	200
CHAPTER 9. CONCLUSIONS		201
9.1	Summary of This Research	201

9.2	Summary of Contributions.....	204
9.3	Suggestions for Future Research.....	205
9.4	Conclusion	206
BIBLIOGRAPHY		207
 CHAPTER A. FLOW GRAPH ANALYSIS OF DISCRETE-TIME MARKOV		
CHAINS AND SEMI-MARKOV CHAINS.....		215
A.1	Flow Graph Analysis of Markov Chains	215
A.2	Flow Graph Analysis of Semi-Markov Chains.....	218
 CHAPTER B. DISCRETE-TIME QUEUEING SYSTEMS.....		220
B.1	Discrete-time Models	220
B.2	Notation.....	221
B.3	Analysis of Geo/Geo/1/K System	222
B.4	Analysis of Geo/Geo/1 System	226
B.5	Analysis of Geo/G/1/K System.....	227
B.6	Analysis of Geo/G/1 System.....	230

List of Figures

CHAPTER 1.	1
CHAPTER 2.	16
CHAPTER 3.	29
Figure 3.1. State flow graph of tagged user in Group i	36
Figure 3.2. State flow graph of the channel	53
Figure 3.3. Channel throughput and packet response time of a S-ALOHA with $N = 100$ and $\lambda = 0.0042$	56
Figure 3.4. Channel throughput and packet response time of a S-ALOHA with $N = 60$ and $\lambda = 0.007$	57
Figure 3.5. Channel throughput and packet response time of a S-ALOHA with $N = 20$ and $\lambda = 0.021$	57
Figure 3.6. Channel throughput and packet response time of a S-ALOHA with $N = 20$	59
CHAPTER 4.	63
Figure 4.1. State flow graph of tagged user for a S-ALOHA system	65
Figure 4.2. Time-varying packet arrival rate $\lambda(t)$	75
Figure 4.3. Fluid Approximation Trajectories for $N = 100$ and $p = 0.025$	76
Figure 4.4. Throughput lines for different transmission probabilities given $N = 100$	77
Figure 4.5. User throughput versus $y = p(1 - p_0)$	79
Figure 4.6. Fluid approximation trajectories of finite buffer system for $p = 0.008 \leq 1/N$, $L = 5$ and $N = 100$	82
Figure 4.7. Fluid approximation trajectories for $p > 1/N$	83
Figure 4.8. Mean packet response time versus transmission probability	86
Figure 4.9. User idle probabilities at equilibrium points versus transmission probability	87
Figure 4.10. Channel throughput versus transmission probability	88
Figure 4.11. Mean response time versus transmission probability	89
Figure 4.12. Mean buffer length versus transmission probability	90
Figure 4.13. Channel throughput versus permission probability	90
Figure 4.14. Mean packet response time versus transmission probability	91

CHAPTER 5.	95
Figure 5.1. Channel throughput versus offered load of a CDMA S-ALOHA system	105
Figure 5.2. Mean response time versus offered load.....	106
Figure 5.3. Packet blocking probability versus offer Load	106
Figure 5.4. Channel throughput versus offer Load.....	107
Figure 5.5. Mean packet response time versus offer Load.....	108
Figure 5.6. Mean message response time versus offer Load.....	109
Figure 5.7. Packet blocking probability versus offer Load	109
Figure 5.8. Message blocking probability vs. offer Load.....	110
Figure 5.9. State flow graph of tagged user for a S-ALOHA system in satellite communications channel.....	112
Figure 5.10. Channel throughput versus transmission probability of a satellite S-ALOHA system with $\lambda = 0.0035$	115
Figure 5.11. The mean packet response time versus transmission probability of a satellite S-ALOHA system with $\lambda = 0.0035$	116
Figure 5.12. Channel throughput versus transmission probability of a satellite S-ALOHA system with $\lambda = 0.002$	117
Figure 5.13. The mean packet response time versus transmission probability of a satellite S-ALOHA system with $\lambda = 0.002$	117
Figure 5.14. Channel throughput versus packet transmission probability of System 1 and 2.....	118
Figure 5.15. Mean packet response time versus packet transmission probability of System 1 and 2.....	118
Figure 5.16. Mean queue length versus packet transmission probability of System 1 and 2.....	119
Figure 5.17. User contention probability versus packet transmission probability of System 1 and 2.....	119
CHAPTER 6.	122
Figure 6.1. State flow graph of tagged user.....	124
Figure 6.2. Channel throughput of a CSMA/CD system	132
Figure 6.3. The system response time of packets of a CSMA/CD system.....	133
Figure 6.4. Mean queue length versus total packet generating rate of a optimal CSMA/CD system	136
Figure 6.5. Mean packet response time (in the unit of packet transmission time) versus total packet generating rate of a optimal CSMA/CD system.	137
Figure 6.6. Packet blocking probability versus total packet generating rate of a optimal CSMA/CD system	137

Figure 6.7.	Channel throughput versus total packet generating rate of a optimal CSMA/CD system	138
Figure 6.8.	Channel throughput versus transmission permission probability in a CSMA system	139
Figure 6.9.	Packet response time versus transmission permission probability in a CSMA system	139
Figure 6.10.	Channel throughput and group throughputs	143
Figure 6.11.	Mean packet response time of users in Group 1 and 2	143
Figure 6.12.	Mean queue length of users in Group 1 and 2	144
Figure 6.13.	Packet blocking probability of users in Group 1 and 2.....	144
CHAPTER 7.		147
Figure 7.1.	State flow graph of the tagged user under (P2).....	151
Figure 7.2.	Channel throughput.....	161
Figure 7.3.	Mean message response time.....	161
Figure 7.4.	Channel throughput.....	162
Figure 7.5.	Mean message response time.....	162
Figure 7.6.	Channel throughput.....	164
Figure 7.7.	Mean message response time.....	164
Figure 7.8.	Mean queue length.....	165
Figure 7.9.	Message blocking probability.....	165
Figure 7.10.	Mean queue length.....	167
Figure 7.11.	Message blocking probability.....	168
CHAPTER 8.		170
Figure 8.1.	State flow graph of tagged data user in PRMA systems	175
Figure 8.2.	Voice terminal activity model	178
Figure 8.3.	State flow graph of voice user when talkspurt is not dropped	179
Figure 8.4.	Voice packet dropping probabilities given different initial conditions	192
Figure 8.5.	Some important equilibrium probabilities	192
Figure 8.6.	Voice packet dropping probability	195
Figure 8.7.	Data packet blocking probability	195
Figure 8.8.	Mean Data packet response time	197
Figure 8.9.	Mean talkspurt access delay.....	197
Figure 8.10.	Total throughput in the channel	198
Figure 8.11.	Total throughput of data terminals	199
Figure 8.12.	Mean queueing length of data terminals.....	199

CHAPTER 9.	201
APPENDIX A.	215
Figure A.1. State flow graph of a two-state Markov chain	216
Figure A.2. State flow graph for calculating the required PGF	217
Figure A.3. State flow graph for calculating $\phi_{11}(z)$	217
Figure A.4. State flow graph of two-state semi-Markov chains	219
APPENDIX B.	220
Figure B.1. State transition rate diagram for Geo/Geo/1/K	222

List of Tables

CHAPTER 1.	1
CHAPTER 2.	16
CHAPTER 3.	29
Table 3.1. Mean queue length of buffered S-ALOHA system with $N = 3$ and $p = 0.1$	61
Table 3.2. Mean queue length of buffered S-ALOHA system with $N = 10$	61
CHAPTER 4.	63
Table 4.1. Stable Region of transmission probability p	86
CHAPTER 5.	95
CHAPTER 6.	122
Table 6.1. The probability distribution p_k of user buffer length in a CSMA/CD system with $N = 100$, $\lambda = 0.00035$, $T = 20$, $D = 1$, $L = 8$ and $p = 0.01$	133
Table 6.2. The probability distribution p_k of user buffer length in a CSMA/CD system with $N = 100$, $\lambda = 0.00035$, $T = 20$, $D = 1$, $L = 8$ and $p = 0.0012$	134
Table 6.3. The probability distribution p_k of user buffer length in a CSMA/CD system with $N = 100$, $\lambda = 0.000175$, $T = 40$, $D = 1$, $L = 8$ and $p = 0.01$	134
Table 6.4. The probability distribution p_k of user buffer length in a CSMA system with $N = 100$, $\lambda = 0.00035$, $T = 20$, $L = 8$ and $p = 0.004$	140
Table 6.5. The probability distribution p_k of user buffer length in a CSMA/CD system with $N = 100$, $\lambda = 0.00035$, $T = 20$, $L = 8$ and $p = 0.002$	141
CHAPTER 7.	147
CHAPTER 8.	170
Table 8.1. PRMA system parameters.	190
CHAPTER 9.	201

APPENDIX A.	215
APPENDIX B.	220

List of Abbreviations

AMPS	Advanced Mobile Phone Service
ARQ	Automatic Repeat Request
AWGN	Additive White Gaussian Noise
BASTA	Bernoulli Arrivals See Time Averages
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
CDMA	Code Division Multiple Access
CDPD	Cellular Digital Packet Data
CPODA	Contention-based Priority-oriented Demand Assignment
CSMA	Carrier Sense Multiple Access
CSMA/CD	Carrier Sense Multiple Access with Collision Detection
DFT	Delay (Defer) First Transmission
DQDB	Distributed Queue Dual Bus
DSMA/CD	Digital Sense Multiple Access with Collision Detection
EPA	Equilibrium Point Analysis
FCFS	First Come First Serve
FDDI	Fiber Distributed Data Interface
FDMA	Frequency Division Multiple Access
FIFO	First In First Out
IFT	Immediate First Transmission
ISMA	Inhibit Sense Multiple Access
LAN	Local Area Network
LCFS	Last Come First Serve
LHS	Left-hand Side
MA	Multiple access
MAC	Medium Access Control
MAI	Multiple Access Interference
MAN	Metropolitan Area Network
NFT	New Packet First Transmission

PASTA	Poisson Arrivals See Time Averages
PER	Packet Error Rate
PGF	Probability Generating Function
PRMA	Packet Reservation Multiple Access
R-ALOHA	Reservation ALOHA
RHS	Right-hand Side
R-ISMA	Reserved Idle Signal Multiple Access
S-ALOHA	Slotted ALOHA
SNR	Signal Noise Ratio
SRMA	Split Channel Reservation Multiple Access
SRUC	Split Reservation upon Collision
TDMA	Time Division Multiple Access
TSMA/PCD	Tone Sense Multiple Access with Partial Collision Detection
TU	Tagged User
TUA	Tagged User Approach
VMST	Virtual Message Service Time
VPST	Virtual Packet Service Time
WAN	Wide Area Network

List of Symbols

$A(z)$	The PGF of the number of packets arriving at the buffer of the tagged user during a packet service time
$A_i(z)$	The PGF of the number of packets arriving at the buffer of TU i during a packet service time
$B(z)$	The PGF of virtual packet or message service time
$B_i(z)$	The PGF of virtual packet or message service time of TU i
$C(z)$	The PGF of the time slots needed for a user at S_0 to enter S_r
$C_r(z)$	The transfer function from CON to REV
$C_s(z)$	The transfer function from CON to SIL
D	The permitted maximum access delay for voice packets in PRMA, or channel propagation delay in other systems
$D(z)$	The PGF of the number of time slots needed to detect a collision
$D_i(z)$	The PGF of the number of time slots needed to detect a collision for users in Group i
E_b	The bit energy
$G(z)$	The PGF of the number of packets contained in a message
$G_s(z)$	The PGF of the number of slots which a user at S_r needs before releasing the reserved slot in R-ALOHA systems
$G_v(z)$	The PGF of the number of slots which a user spends at REV in each cycle
$H(z)$	The PGF of packet length in slots of users in homogeneous systems
$H_i(z)$	The PGF of packet length in slots of users in Group i
L	The buffer capacity of users
L_b	The number of bits contained in each packet
L_i	The buffer capacity of users in Group i
M	The number of time slots in a frame
M_d	The number of data users in PRMA systems
M_v	The number of voice users in PRMA systems
N	The number of users in homogeneous systems

N_i	The number of users in Group i
N_c	The number of chips per bit in CDMA ALOHA
$O_{ch}(z)$	The channel output process
$O_i(z)$	The output process of users in Group i
$P(z)$	The PGF of the number of packets present in user buffers at any time
$P_{be}(k)$	The BER of any one of the users when there are k users transmitting packets simultaneously
$P_i(z)$	The PGF of the number of packets present in user buffers of Group i at any time
$Q_s(k)$	The packet successful probability of the tagged user given k simultaneous transmission
$R(z)$	The PGF of packet or message system response time
$\bar{R}_d$	The mean system response time of data packets
$R_i(z)$	The PGF of packet or message system response time of users in Group i
$\bar{R}_m$	The mean system response time of an accepted message
$\bar{R}_p$	The mean system response time of a packet in an accepted message
$S_c(z)$	The PGF of the time slots that a voice user spends at SIL in a cycle
S_{load}	The mean packet acceptance rate of a user
S_{out}	The mean packet departure rate of a user
T	The mean packet transmission time of users
T_{CON}	The mean time slots that a busy user spends at mode CON in a cycle
T_{IDL}	The mean length of the idle period in a cycle
T_{REV}	The mean time slots that a user spends at mode REV in a cycle
T_{SIL}	The mean time slots of voice users spending at SIL in each cycle
T_{WAI}	The mean time slots that a user spends at mode WAI in a cycle
T_i	The mean packet transmission time of users in Group i
$\bar{T}_v$	The average channel access delay of undropped talkspurts
$W(z)$	The PGF of packet or message waiting time
$W_i(z)$	The PGF of packet or message waiting time of users in Group i
$\bar{W}_m$	The mean waiting time of an accepted message
b	The mean VPST of users

b_i	The mean VPST of users in Group i
$b^{(n)}$	The n th order moment of VPST of users
$b_i^{(n)}$	The n th order moment of VPST of users in Group i
c_k'	The probability that a talkspurt consists of k packets and is dropped
c_k''	The probability that an active voice user reserves a slot after an access delay of k slot
$\bar{d}_{ch}$	The mean channel interdeparture time
$\bar{d}_i$	The mean interdeparture time of users in Group i
g_i	The probability that a message contains i packets
l	The number of packets present in user buffers at any time
l_i	The number of packets present in user buffers of Group i at any time
l_v	The total number of packets in a talkspurt
m	The number of user groups in heterogeneous systems
n_{block}	The mean number of packets contained in a blocked message
n_{drop}	The number of dropped packets in a talkspurt
p	The probability that a user with a deferred packet would start transmission or sense the channel in a slot in ALOHA or other protocols, respectively
p_B	The packet blocking probability of users
p_I	The probability that a slot is sensed idle by the tagged user given that it senses the channel in that slot
p_a	The probability that a slot is available
p_{ad}	The probability that a slot is available to data users in PRMA systems
p_{av}	The probability that a slot is available to voice users in PRMA systems
p_b	The probability that a user is busy
$p_{c \bar{R}}$	The probability that a user is at mode CON given that it is not at mode REV
p_{con}	The probability that a voice user is contending for a slot at any time
p_{ctu}	The probability that a collision would occur and involve the tagged user given that there is a transmission in a slot and the tagged user is busy
p_d	The permission probability of data users
p_{drop}	The voice packet dropping probability in PRMA systems
$p_{i,B}$	The packet blocking probability of users in Group i

$p_{i,l}$	The probability that a slot is sensed idle by TU i given that it senses the channel in that slot
$p_{idle \bar{R}}$	The probability that a user is at mode IDL given that it is not at mode REV
$p_{i,b}$	The probability that a user in Group i is busy
$p_{i,ctu}$	The probability that a collision would occur and involve TU i given that there is a transmission in a slot and TU i is busy
$p_{i,k}$	The probability that there are k packets or messages in a user buffer of Group i at any time
$p_{i,sen}$	The permission probability that a user in Group i with a deferred packet would sense the channel in a slot
$p_{i,ctx}$	The probability that the tagged user in Group i would initiate a transmission in a slot given that it is busy
$p_{i,s tx}$	The probability that a transmission is successful given that TU i is busy and there is packet transmission in a slot
$p_{i,su}$	The probability of a successful transmission by TU i given that it transmits in a slot
$p_{i,tx}$	The probability that a user in Group i would issue a transmission in an idle slot
p_k	The probability that there are k packets or messages in a user buffer at any time
p_{mblock}	The message blocking probability
p_{pblock}	The packet blocking probability
p_r	The probability that a user is at mode REV
p_{rev}	The probability that a slot is reserved
p_s	The probability of a successful transmission by the tagged user given that it transmits in a slot
p_{sen}	The probability that a user with a deferred packet would sense the channel in a slot
p_{sd}	The probability of a successful transmission by the tagged data user given that it transmits in a slot
p_{su}	The probability of a successful transmission by the tagged user given that it transmits in a slot
p_{sv}	The probability of a successful transmission by the tagged voice user given that it transmits in a slot
p_{tx}	The probability that a user would issue a transmission in an idle slot
p_v	The permission probability of voice packets in PRMA systems

	The probability that a voice user is busy
q	The probability that a user with new packet would start transmission or sense the channel in a slot in ALOHA or other protocols, respectively
q_i	The probability that a user in Group i with new packet would start transmission or sense the channel in a slot in ALOHA or other protocols, respectively
$q_{i,k}$	The probability that the transmission is contributed by all users in Group k given that there is a successful transmission in a slot and TU i is busy
$q_{i,u}$	The probability that the transmission is contributed by TU i given that there is a successful transmission in a slot and TU i is busy
$q_{i,\bar{u}}$	The probability that the transmission is contributed by all users in Group i except TU i given that there is a successful transmission in a slot and TU i is busy
r	The packet or message system response time
r_i	The packet or message system response time of users in Group i
t_d	The voice user channel access delay in slots
w	The packet or message waiting time
w_i	The packet or message waiting time of users in Group i
$\bar{w}_{mi}$	The mean waiting time of an accepted message containing i packets
β	$= p_{av}p_{sv}$
γ	The probability that a voice packet is the last packet of a talkspurt
Θ	The channel throughput
Θ_d	The channel throughput when systems operate at desired point
Θ_s	The channel throughput when systems operate at saturated point
θ	The throughput of any user
θ_d	The throughput of a data user in PRMA systems
θ_v	The throughput of a voice user in PRMA systems
θ_i	The throughput of any user in Group i
λ	Bernoulli arrival rate for users in homogeneous systems
λ_d	Bernoulli arrival rate of data packets for data users in PRMA systems
λ_i	Bernoulli arrival rate for users in Group i
λ_v	Bernoulli arrival rate of talkspurts for voice users in PRMA systems
μ	Service rate for packets with geometrically distributed VPST

$\Xi_i(z)$	The PGF of channel access delay of packets in Group i
$\Pi(z)$	The PGF of the number of packets present in user buffers at a packet departure time
$\Pi_i(z)$	The PGF of the number of packets present in user buffers of Group i at a packet departure time
$\pi_{i,j}$	The equilibrium probability that a packet in Group i sees j packets in the buffer at departure time given that TU i is busy
π_j	The equilibrium probability that a packet sees j packets in the buffer at departure time given that the tagged user is busy
$\varphi(z)$	The PGF of the number of time slots in each cycle

1 INTRODUCTION

1.1 Multiple Access Protocols

Multiple access refers to the capability to share a communication resource by a number of users. Multiple access protocols are techniques to allocate the communication resource efficiently to many contending users by controlling their access to the common communication channel. The need for multiple access protocols arises whenever a communication medium is shared by a population of users. In the open systems interconnection (OSI) reference model of the International Organization for Standardization (ISO), multiple access (MA) protocol is implemented in the medium access control (MAC) sublayer of data link layer. The classification of multiple access protocols is not unique [2][3][4]. Here MA protocols are classified into the following categories [2][4]: fixed assignment, random access, demand assignment, CDMA protocols, and adaptive strategies and mixed modes.

1.1.1 Fixed Assignment

Fixed assignments are techniques for allocating a fixed portion of available channel bandwidth to each user, independently of their activity. The most common forms of these techniques are frequency division multiple access (FDMA) and time division multiple access (TDMA). In fixed assignments, the bandwidth allocated to all users in a system does not exceed the system capacity. Fixed assignment is highly efficient when each user has

data for transmission all the time. However, it is very inefficient if the user traffic is bursty. In this case, most of the resource allocated to each user is wasted.

1.1.2 Random Access

In computer communications, most data traffic is characterized as bursty. An inherent nature of bursty traffic is the large peak-to-average ratio in the required data transmission rate. However, the strong law of large number guarantees that the demand from all users at any instant will be approximately equal to the sum of average demands of all users with a very high probability if the user population is sufficiently large.

In random access protocols, a single channel can be accessed by all users randomly. Since the transmission medium is shared by all users, packet collisions may occur if more than one user transmits the data at the same time. Normally, all the packets involved in a collision are destroyed except in the case of channel capture where a packet with the strongest power at base station may capture the channel [19][20]. The price for channel capture is the higher than necessary signal power which would consume more than necessary precious battery power and result in higher co-channel and adjacent channel interference. Channel capture seldom occurs in the wireless system where power control is practiced in each user. The most important problem of random access protocols is how to resolve the contention among users such that the performance for a given system can be optimized.

The typical examples of random access protocols are pure ALOHA [5], S-ALOHA (slotted ALOHA) [6][88], CSMA (carrier sense multiple access) [7][14], CSMA/CD (carrier sense multiple access with collision detection) [9][11]. In addition, there are also many variations of the above protocols such as DSMA/CD (digital sense multiple access with

collision detection), the medium access control protocol of cellular digital packet data (CDPD) system [77], ISMA (inhibit sense multiple access) [21], R-ISMA (reserved idle signal multiple access) [23], and CSMA/CA (carrier sense multiple access with collision avoidance), the medium access control protocol of IEEE 802.11 wireless LAN [25].

1.1.3 Demand Assignment

Demand assignment techniques require that explicit control information regarding the user demand for the communication resource be exchanged. The demand assignments are further divided into two classes: centralized control and distributed control based on whether the assignment is controlled by a central station or the users themselves.

1.1.3.1 Demand Assignment with Centralized Control

In a demand assignment with centralized control, a central station is employed to determine which user has the next channel access right. The typical examples of demand assignment with centralized control include cellular phone systems such as AMPS (advanced mobile phone service), TDMA digital cellular standard IS-54 and CDMA digital cellular standard IS-95 for voice communications [15], polling [16], probing [17] and SRMA (split-channel reservation multiple access) [18] for packet data communications.

1.1.3.2 Demand Assignment with Distributed Control

In a demand assignment with distributed control, each user monitors the demand requests from all the other users and executes an identically distributed algorithm dependent upon the requests to determine which user should have the next channel-access right.

There are two reasons why distributed control demand assignment is desirable. The first is the reliability. A distributed control system does not depend on a central station. This

avoids the single point failure. The second is the improved performance, especially when dealing with systems with long propagation delays such as those using satellite channels.

An important class of protocols in the demand assignment techniques with distributed control is the virtual token passing scheme such as IEEE 802.4 token bus, IEEE 802.5 token ring, FDDI (fiber distributed data interface) and slot ring systems [24][28].

Another important class of protocols in the demand assignment techniques with distributed control is the reservation protocols where the user who wants to transmit packets must make a reservation for channel access before initiating a transmission. Based on how the control information is exchanged among users, explicit and implicit reservation protocols are defined.

In explicit reservation protocols, normally a portion of channel capacity is used as the reservation sub-channel by all users to send reservation requests. The rest of channel capacity is used for data transmission on a reservation basis, free of conflict. Now, a new problem arises regarding how to control the access to reservation sub-channel. Most of the existing reservation protocols adopt either a random access protocol such as S-ALOHA or a fixedly assigned TDMA protocol as the multiple access protocols for reservation sub-channel. The well known examples of the contention-based reservation protocols are CPODA (contention-based priority-oriented demand assignment) [6][29], ALOHA-Reservation [1] and the reservation protocol proposed in [30]. An example of the fixed TDMA reservation schemes is Reservation-TDMA [31]. An explicit reservation protocol which does not use reservation sub-channel is the medium access control protocol of DQDB (distributed queue dual bus), the IEEE 802.6 standard for MAN (metropolitan area networks) [25].

In implicit reservation protocols, no channel bandwidth is allocated for reservation purpose. Two well known examples of implicit reservation protocols are R-ALOHA [32] for satellite communication systems, and PRMA (packet reservation multiple access) [36][37][38], a wireless access protocol for the integration of packetized voice and data over a common communication channel.

1.1.4 CDMA Protocols

CDMA protocols achieve their multiple access property by assigning each user a different code. When a receiver receives multiple wideband signals, it will retrieve the original signal from a particular user using its code while the wideband signals from all the other users, called interfering users, remain wideband and appear as noise.

When the number of interfering users is sufficiently small, the SNR (signal noise ratio) will be significantly large so that the original signal can be extracted free of errors. In this case, the protocol behaves as a contentionless one. However, if the number of interfering users is so large that the original signal may not be extracted, the protocol behaves as a contention protocol.

When behaving as a contentionless protocol, a CDMA protocol can be further included in fixed assignment or demand assignment categories. For instance, CDMA digital cellular system IS-95 can be classified as demand assignment protocol. When behaving as a contention protocol such as CDMA S-ALOHA system [88], CDMA may be considered as a random multiple access protocol.

1.1.5 Adaptive Strategies and Mixed Modes

Each multiple access protocol in the four categories above has its advantages and limitations. No one performs better than any others over entire range of traffic load or demand [2]. Since the traffic load placed on the channel is time varying, adaptive strategy and mixed mode techniques combine the ideas from distinct protocols and are able to adapt to the varying demand, such that near-optimal performance can be achieved at all times. Two typical examples of these protocols are SRUC (split reservation upon collision) [42] and URN [43].

1.2 Performance Measurements

The performance issues for different multiple access protocols might be different. For instance, the most important performance measurements for ALOHA systems are channel throughput and packet delay. However, for cellular phone system, the most important issue is the system capacity, or the number of simultaneous conversations supported in a cell given the number of channels and maximum call blocking probability. The voice packet delay are not the concern since IS-54 is circuit switched system.

In this thesis, the focus is on the multiple access protocols in which contention is required for channel access such as S-ALOHA, CSMA/CD and PRMA, which are called random or contention-based multiple access protocols. The performance of random multiple access protocols is evaluated using many criteria. The most desirable performance characteristics are high channel throughput and low packet response time. However, a number of other attributes are also important, such as the capability to support multimedia traffic, which involves that of different types, priorities and delay constraints. To guarantee the

proper operation of MA schemes with distributed control, the system stability is also very important.

In addition, there are many other criteria depending on the systems and applications. For instance, in interconnected local area networks or multihop networks, a channel output process may constitute part of the input processes to other adjacent networks. In this case, the channel output process is also important in order to analyze the performance of interconnected networks. In multiple access systems, each user has a buffer and actually is a queueing system with finite capacity. Hence, the queue length and packet blocking probability are also important performance characteristics.

1.2.1 Channel Throughput

There is no unique definition for channel throughput. It can be defined as the average number of bits transmitted through the channel per unit time. Since packet can be corrupted by collision or other reasons, only bits of error-free packets are counted when calculating channel throughput.

Normally, throughput is defined as the average number of bits transmitted through the channel per bit time. Hence, throughput is a normalized amount between 0 and 1 in this case. This definition of throughput is employed throughout the thesis. Sometimes, throughput is also called channel utilization or channel efficiency.

1.2.2 Response Time

The packet response time is the time duration from a packet arrival to departure that the packet spends in the system. In addition to response time, there are other parameters which are associated with the delay characteristics of multiple access protocols. They are waiting

time and access delay. The packet waiting time is the time duration from packet arrival to its ready for transmission. The channel access delay is the time duration from a packet ready for transmission to the starting of the successful transmission of the packet. Clearly, response time is the summation of waiting time, access delay and packet transmission time.

Generally, the distribution of response time is desirable since it characterizes the response time uniquely. In the case that the distribution is able to be obtained, the mean response time is expected, which describes the system delay characteristics in some degree.

1.3 Motivations

The objective of this thesis work is to develop a simple and unified technique for the performance analysis of contention based multiple access protocols with buffered users, and expect that the same technique can be extended to the performance analysis of other buffered multiple access protocols.

Generally speaking, in a random multiple access system each user is a queueing system which interacts with other users through a common shared channel. Since the queueing systems of the users are not mutually independent, interfering queueing problem arises. There is a very large number of publications available on performance evaluation of random multiple access protocols. However, only few of them are dedicated to the buffered systems due to the analysis complexity. Moreover, most of the available literatures on buffered systems are not able to use the existing results in the field of queueing theory. Needless to say, the ability to employ the classical queueing theory would further simplify the system analysis.

As discussed in Chapter 2, exact Markov analysis on buffered multiple access systems is intractable normally. The approximate techniques have to be employed. As a matter of

fact, among the available literatures on the performance analysis of buffered random multiple access protocols, no one is dedicated to the exact analysis of buffered systems with more than two users. Even the case of systems with two users cannot be analyzed exactly in general. That is why a special case of two-node system was treated in [48].

The main problem associated with the available approximate approaches on buffered systems is that there is no simple and unified analysis approaches available for buffered systems. EPA (equilibrium point analysis) [1] is a unified approach for performance evaluation of unbuffered random multiple access protocols. Though it can be applicable in some simple buffered systems, it cannot handle more complicated cases.

Another problem is that many approximate approaches still require a two or higher dimensional Markov chain and have to calculate the state transition probability matrix. Special care must be taken for numerical stability. Those approaches cannot be applied in complicated systems because of the high analysis and calculation complexity.

So far, there are no literatures available on the buffered systems with bulk packet arrivals or on heterogeneous buffered systems.

In this thesis research, we want to develop a unified approach for the performance analysis of buffered multiple access systems so that most of the above problems can be treated. The solution is to use queueing theory to the system analysis since most of the above problems will be solved by doing so, as discussed later on. Though it is known that the interfering queueing problem cannot be solved exactly using classical queueing theory [45], it is possible to apply queueing theory in the system analysis approximately. Not mention that approximation is necessary for the analysis of buffered systems.

1.4 Thesis Contributions

In this thesis, a unified approach called tagged user approach (TUA) is proposed for the approximate analysis of buffered random multiple access systems. Though it is well known that each user in the systems can be treated as a queueing system, only quite a few literatures attempted to use queueing theory for the system analysis because of the difficulty to find the packet service time distribution [60][61][63]. In TUA, a state flow graph [76] is employed to find the packet service time distribution of the buffered multiple access systems easily for the first time. In addition, the analyses on channel contention scheme and user queueing behavior are decoupled in TUA. As a result, the analysis complexity of TUA is comparatively low since we can focus our analysis on channel contention scheme. The analysis on user queueing behavior is left to the queueing theory. The amount of numerical computation in TUA is also small since state transition probability matrix of Markov chain analysis is avoided.

The basic idea of TUA is to treat each user as a queueing system which is assumed to always operate with its equilibrium probabilities. Consequently, classical queueing theory can be applied in the system analysis if the packet service time distribution is available. Though the queueing system of each user is not mutually independent of others because of the interfering queueing problem, each user is considered approximately to operate with its own equilibrium probabilities, independent of the behaviors of other individual users. The dependences among users are reflected only by the fact that the equilibrium probabilities of each user are determined by the behaviors of all users in the system.

A widely employed assumption is the channel symmetry, whose definition is given in Definition 3.5 in Chapter 3. The channel symmetry implies that the channel is fair to each user so that the behaviors of all users with the same parameters are statistically equivalent. Hence, the users in a system can be divided into different groups such that users in each group have the same system parameters and statistically identical behaviors. A user is selected from each group randomly and called the tagged user of the corresponding group. The key for system analysis is to understand the behaviors of all the tagged users. If they are known, the behaviors of all users are known, so is the performance of the system.

In order to determine the distribution of packet service time for each tagged user, state flow graph [76] is employed in TUA. It is found that packet service time of each user is a function of the busy probabilities of all users since only busy users will contend for the channel. On the other hand, queueing theory shows that the busy probability of each queueing system is determined by the distribution of its packet service time. The above function relationships are sufficient to determine packet service time distribution and user busy probability for each tagged user, which are then employed to analyze the system performance straightforwardly.

In this thesis, TUA is applied in the performance analysis of several typical MA protocols like S-ALOHA, R-ALOHA, CSMA/CD and PRMA. Specific techniques are developed in order to analyze these protocols. In addition, an example is demonstrated regarding stability and equilibrium point analysis of the systems. The accuracy of TUA is verified using the simulation results contributed by available publications.

In general TUA has the following advantages:

1. Packet service time distribution is easy to be obtained using state flow graph. The variable packet size, channel propagation delay and different transmission principle can be handled.
2. The analyses on channel contention scheme and user queueing behavior are decoupled so that the results in classical queueing theory can be employed for the system analysis. Complicated queueing behaviors of users such as bulk arrival process and arbitrary service discipline can be handled as long as the corresponding queueing theory is available.
3. A wide range of performance measurements can be obtained using TUA. In addition to mean values, the distribution of most measurements can also be obtained.
4. TUA can be applied to non-ideal channel. No matter what is the channel condition, TUA can be used as long as the packet error rate (PER) is known.
5. TUA is able to handle memory problem like propagation delay and message consisting of an arbitrarily distributed number of packets.
6. TUA is able to analyze the stability of buffered random multiple access protocols.
7. TUA can be applied to heterogeneous systems.
8. The numerical algorithms of TUA can be developed easily.
9. Calculation complexity of TUA is low since state transition probability matrix is not required for the system analysis.
10. There is no numerical instability problem with TUA.
11. The accuracy of TUA is good.

1.5 Organization of the Thesis

The remained part of the thesis is organized as follows. Firstly, a review of the available approaches on buffered random multiple access protocols is carried out in Chapter 2. Then, the principle of TUA is presented in Chapter 3 by applying TUA in a very general model of slotted random multiple access protocols which has S-ALOHA, CSMA and CSMA/CD as its special cases. This model is a heterogeneous system where users may have any packet arrival processes and service disciplines. In this chapter, the emphasis is on the principle and methodology of TUA, details are not given. In Chapter 4 to Chapter 8, we apply TUA in several typical protocols, which may or may not be considered as special cases of the

model in Chapter 3. The detailed analysis is presented. Numerical algorithm is developed and the numerical results are presented.

In Chapter 4, a buffered S-ALOHA without channel propagation delay is treated. The system stability is studied when buffer capacity is infinite. A system is said to be unstable if the buffer size and system response time grow without limit. There is no stability problem associated with the systems with finite buffer capacity since the buffer size is limited. However, the system equilibrium points are also studied. The analysis would assist us to select transmission control scheme properly so that the desired system performance can be reached. The analytical results are verified by simulations.

To demonstrate how to apply TUA in systems with channel propagation delay, a S-ALOHA system operating in satellite communications environment is studied in Chapter 5. Channel propagation delay would introduce memory which would complicate the system analysis. However, this problem is handled well by TUA. In the system model of Chapter 4, it is assumed that a user transmits a new or blocked packet in a slot with same probability. More complicated transmission principle is employed in satellite S-ALOHA where new packets and blocked packets would be transmitted using different probabilities in each slot. The analytical results are verified by simulations.

In order to show how to apply TUA in systems in a wireless channel, a CDMA S-ALOHA system is also studied in Chapter 5. It shows that the analysis on multiple access protocols is independent of the analysis on channel packet error rate (PER). If the channel PER is given, the performance analysis is straightforward. To illustrate the advantage of TUA that the analysis on channel contention scheme is independent of the analysis on user

queueing behavior, the multi-packet message arrival is permitted in this CDMA S-ALOHA system. That is, each user has a bulk arrival process. By developing the corresponding queueing theory, the system is analyzed using TUA.

In Chapter 6, we apply TUA in the analysis of buffered CSMA/CD systems. Several examples are given. Firstly, we study the influence of permission probability on system performance given user population and packet arrival rate. As shown by numerical results, too small or large permission probability would degrade the system performance. Secondly, an optimal CSMA/CD system in term of channel throughput is studied. In this case, the permission probability is not predetermined, but adapted to the system status. Thirdly, a CSMA system, a special case of CSMA/CD system is studied. Finally, we analyze a heterogeneous CSMA/CD systems consisting of two groups of users. So far, there is no publications available on buffered heterogeneous CSMA/CD systems.

In Chapter 7, a buffered R-ALOHA system is studied using TUA, where long channel propagation delay is allowed, and the message size may have arbitrary distribution in the integral number of slots. So far, the only available analysis on R-ALOHA is EPA. Though a R-ALOHA system without propagation delay was studied in [33], the approach there focused on the statistic behavior of the whole system, and did not address the system stability and the behavior of individual users. EPA has been applied in R-ALOHA systems where each user can hold at most one message, and message size is geometrically distributed. However, it is very hard, if not impossible, to extend EPA analysis to systems with buffered users and arbitrarily message size distribution because of the complexity for modeling R-ALOHA system in EPA.

In Chapter 8, a PRMA system is studied using TUA. PRMA system is a heterogeneous system consisting of voice users and data users which have different demands. A main problem associated with the available approaches is that the voice packet dropping scheme is not modeled properly. This problem is solved using TUA, for the first time. In addition, the available approaches cannot be applied in PRMA systems where data users are allowed to reserve a slot if they have multi-packet message ready for transmission such as file transfer application. Though this system is not analyzed using TUA in this chapter, the capability of TUA for handling this system is discussed.

In addition, a brief introduction of state flow graph is presented in Appendix A, and the queueing analysis of Geo/Geo/1, Geo/Geo/1/K, Geo/G/1 and Geo/G/1/K is given in Appendix B for convenience.

1.6 Summary

In this chapter, firstly a brief introduction on multiple access protocols is presented. The interested performance measurements for contention based multiple access protocols are identified. Then the motivation for this thesis work is given. The main contribution of this thesis is that TUA, a simple and unified approach, is proposed for the performance analysis of contention based multiple access protocols with buffered users. The basic principle of TUA is introduced and the main advantages are listed. Finally, the organization of the thesis is given.

2 LITERATURE REVIEW

2.1 Introduction

There are a lot of publications dedicated to the performance analysis of contention based multiple access protocols. Regardless of which protocols or performance measurements are focused on, three classes of techniques are widely employed. They are S-G analysis, Markov analysis, and equilibrium point analysis (EPA) [1].

The S-G analysis or Poisson analysis can be used to analyze both unslotted and slotted systems [5][8][9][10]. In the S-G analysis of slotted multiple access protocols, it is assumed that the traffic source consist of an infinite number of users that collectively form an independent Poisson source with an aggregate mean packet generating rate of S packets per slot, and the combined packet transmission process consisting of new transmission and retransmission can be approximated as a Poisson process at a rate of G packets per slot. It is S-G approach which firstly predicted that the maximum channel throughput in slotted ALOHA system is $1/e \approx 0.368$. A detailed introduction on the S-G analysis is given in [10]. The S-G analysis has been applied to the analysis of many multiple access protocols because of its simplicity. The emphasis of S-G analysis is on the behavior of the overall system. However, an obvious drawback of the S-G analysis is that it cannot address individual users, and it cannot handle the system stability [1].

The second approach, Markov analysis, is used for the first time to analyze the performance of a S-ALOHA system in [12]. This technique formulates a Markovian model of the system and obtains the steady state occupancy probability distribution by calculating its state transition probabilities. It can provide an exact analysis of the dynamic behavior and stability of the system. Because of this, it is widely used for the performance analysis of all kinds of multiple access protocols [7][12][14][44][46]. In principle, Markov analysis can be applied to the analysis of any multiple access protocols. However, this technique is usually used to analyze homogeneous systems with unbuffered users. This is because in this case the system state is a scalar and the size of the state space is just the total number of users in the system. In buffered homogeneous system, the state space of Markov analysis is multi-dimensional, and the size of the state space is an exponential order of the user buffer capacity. The great size and multidimensional nature of state space make the Markov analysis intractable, not to mention the greater complexity on the Markov analysis of buffered heterogeneous systems.

The third technique, EPA is a fluid-type approximate analysis which can be applied to the steady state analysis of the system. In EPA, it is assumed that the system always operates at its equilibrium points such that the number of users at any mode be fixed at any time [1]. The number of users in any mode does not change with time since it is assumed that the number of users flowing into any mode be always equal to the number of users flowing out. EPA does not need the state transition probabilities, the reason behind its simplicity. Using EPA, the stability analysis of the system is straightforward. If a system has only one equilibrium point, the system is globally stable. If a system has more than one

equilibrium point, the system is locally stable. Since the calculation of state transition probabilities is not required, EPA can be applied to many multiple access systems which cannot be handled by Markov analysis due to the large size and multi-dimension of state space, such as packet reservation multiple access (PRMA) systems [39][40][41][47]. EPA is an effective technique for approximate analysis of multiple access protocols [1][45]. A disadvantage of EPA is that system models cannot move from state to state. In addition, the system model is very complicated for systems with large buffer capacity, large channel propagation delay or bulk packet arrival processes so that the analysis may be intractable. Because the system state information in EPA is not carried by equilibrium probabilities, EPA can not use the results in queueing theory. Hence, EPA cannot be applied to buffered multiple access systems with complicated queueing behavior efficiently.

In addition to these three general techniques, there are many other specific techniques which may be applied to some specific multiple access protocols [47][57]-[63][65]-[69]. Some of them may combine the ideas from the three techniques above. For instance, an analytical method which combines both Markov analysis and EPA is proposed in [47] for the performance analysis of joint voice and data PRMA over random packet error channels.

In the following, we focus on the performance analysis of slotted contention based multiple access protocols with buffered users since this is the main objective of this thesis work.

2.2 Analysis on Buffered Random Access Protocols

Generally speaking, each user in a slotted multiple access system is a discrete time queueing system though the behavior of each user is affected by others via the channel shared among them. The proper model for multiple access systems is the one which treats

each user as a queueing system. The problem is that these queueing systems are not mutually independent. The behavior of one user is determined by the behaviors of all users in the system. This is called interfering queueing problem. If the packet service time distribution can be obtained somehow, the classical queueing theory can be applied to the system analysis. Unfortunately, it is intractable to derive the packet service time distribution exactly even for buffered S-ALOHA, the simplest random multiple access protocols. Hence, classical queueing theory cannot be used for exact analysis of random multiple access protocols.

As a matter of fact, the only possible method which can provide an exact system analysis on multiple access protocol is Markov analysis. Unfortunately, Markov analysis on buffered systems is intractable because of large size and high dimension of the state space of the Markovian model. For instance, to describe the behavior of a homogeneous MA system with finite population of N users and finite buffer capacity of L packets exactly, a N -dimensional Markov chain with state vector $\mathbf{b} = (b_1, b_2, \dots, b_N)$ is required, where $0 \leq b_i \leq L$, $i = 1, 2, \dots, N$, is the number of packets in the queue of the i th station. Consequently, the model is too complicated to be studied using Markov analysis.

It becomes clear that approximate techniques are required for the analysis of the random multiple access systems with buffered users in order to keep the size of the state space comparatively small and to make the theoretical analysis tractable. The basic idea for the approximate analysis is to simplify the system model by exploring the redundant system information and removing it. Needless to say, a method which uses existing results in classical queueing theory would further simplify the system analysis.

However, even the approximate analysis of the buffered multiple access systems is tough. There is a large number of publications dedicated to the performance analysis of all kinds of contention based multiple access protocols, but only few of them focused on the system with buffered users due to the analysis complexity.

Among the analysis methods above, S-G analysis clearly cannot be applied to buffered systems since it does not need any information about individual users. EPA can be applied to the buffered system with simply queueing behavior such as packet arrivals following Bernoulli process and multi-packet messages with geometric distribution, but it is very difficult, if not impossible, to extend EPA to systems with more complicated queueing behavior such as those with general packet arrival process or priority service discipline since it is difficult to model these complicated systems using EPA.

2.3 Review of Related Research

As mentioned before, the approximate technique is necessary for the performance analysis of buffered contention based protocols. Though there is a large number of publications on performance analysis of multiple access protocols, only few of them are dedicated to buffered users so far. The purpose of approximate techniques is to reduce the size and dimension of the state space so that the analysis is mathematically tractable. Normally, the size and dimension of the state space are dependent on the user population, buffer capacity of each user, and the packet size (measured in slots), depending on specific analysis approaches.

One approach for the approximate analysis is to describe the system using a state vector $\mathbf{n} = (n_1, n_2, \dots, n_L)$, where $n_i, i = 1, 2, \dots, L$, is the number of stations having i packets

in their buffers. Obviously, $n_0 = N - \sum_{i=1}^L n_i$, and the total number of possible states in the state space is $\binom{N+L}{L}$ [67]. When $L < N$, this approach results in a considerable reduction in the size and dimension of state space.

A system consisting of two interfering queues with infinite buffer capacity each is considered in [48]. The exact analytical results are obtained using Markov analysis for some nonsymmetric special cases. The exact packet average waiting time and queue length distribution are also obtained for ALOHA-type systems with two symmetric queues. The analytical complexity again implies that Markov analysis for even a small user population is intractable. In fact, there have been no publications dedicated to the exact analysis of buffered systems with more than two users. An approximate approach is also proposed in [48] for a network with user population N .

Another buffered two-node packet radio network is studied in [49]. Instead of using an approximate method to analyze an “exact” model of S-ALOHA system, an exact analytical method for an “approximate” model of S-ALOHA system is developed. In the model, some constraints which control the packet arrival processes are placed on the network parameters so that a simple, closed-form solution for the system equilibrium probabilities can be obtained.

Because of the difficulty to make an exact analysis of mean values for ALOHA-type systems consisting of more than two users, some upper and lower bounds for mean queue lengths of both symmetric and nonsymmetric ALOHA-type systems with finite number of users are proposed in [51].

In addition, a buffered two-node CSMA/CD system is studied in [50] and the mean packet queueing delay of the system is derived explicitly.

A homogeneous S-ALOHA system with infinite buffer capacity is studied in [52], where new packets and retransmission packets are treated differently. Each user is supposed to operate with its own equilibrium probabilities, independent of others. Then a two dimensional Markov chain (x, N) is used to describe the state of each station, where x is a Boolean variable indicating the status of transmitter, and N representing the number of packets in a user's TRANSMIT buffer. A controversial point arises on how to calculate the state transition probabilities [53][54]. An approach similar to [52] is presented in [55] to study the performance of both symmetric and nonsymmetric S-ALOHA systems with finite user buffer capacity. Additionally, the stability condition for the system with infinite buffer is given.

A two-hop centralized S-ALOHA system is investigated in [71]. In the case of star configuration with buffered repeaters, if each repeater is treated as a buffered user and each population of terminals associated with a repeater is treated as the source to the corresponding repeater, the two-hop centralized system is equivalent to a special buffered S-ALOHA system where the arrival process of each user is dependent on its own transmission status.

In [56], an analysis approach on random multiple access systems with finite buffer capacity is proposed. A two dimensional vector $v = (n, k)$, where n and k are the number of busy terminals and the total of outstanding packets in the system, respectively, is used to represent the system state. The symmetrical property of the channel is exploited in order to reduce the size of the state space to $(N + 1)[2 + (L - 1)N]/2$ [67]. Markov analysis is

applied to the simplified model. The presented approach is applicable in optimal ALOHA and Urn schemes.

Another idea for simplifying the model of a finite buffer S-ALOHA system without losing track of the interdependence between user queue sizes is proposed in [57], which uses two Markov chains, called System Markov Chain and User Markov Chain, to describe the system. Each user is associated with a User Markov Chain whose state is denoted by (i, j) , where i is used to describe the user status (idle, active, or blocked), and j to indicate the number of packets in the buffer. The state of System Markov Chain is denoted by (n, m) , where n and m indicate the total number of blocked and active users in the system respectively. The total number of idle users in the system is therefore given by $N - n - m$. The interaction among users is partially taken into account through the inherent coupling between the two Markov chains. Though the transitions among system states and among user states are not Markovian [59], they were considered Markovian approximately with transition probabilities equal to the average transition probabilities. The resulting equations can be solved numerically for performance estimation. A similar idea is also employed in [59], which formulates a model separating the system status-states from the node queue-size states. However, the state space of system status-states is expanded in order to track the status of every terminal separately. As the result, the model can “remember” the status of each user in going from slot to slot, and can be applied to nonsymmetric systems. The same idea is extended to the analysis of a buffered CSMA/CD system in [58].

In [67][68], the state vector for a homogeneous CSMA/CD with fixed packet size is chosen as $\mathbf{b} = (i, j)$, where i is the number of busy users if tagged user is excluded and j

is the number of packets in the buffer of the tagged user. Again, The symmetrical property of the channel is exploited in order to reduce the size of the state space. With the assumption that all users have the same statistical behavior, the selected state vector is sufficient to describe the system behavior. Finally, a Markov analysis is applied to the simplified system model. In this case, the size of state space is reduced to $N \cdot (L + 1)$.

The basic idea behind above analyses except those on two-node systems is to describe the overall system using two Markov chains, namely System Markov Chain and User Markov Chain. The former is used to describe the number of busy users in the system since only busy users may contend for the channel, the latter the number of packets in an arbitrary user, called the tagged user, since all users have statistically identical behavior because of channel symmetry. The system analysis is accomplished by analyzing the two coupled Markov chains. Since the busy or idle status of the tagged user is implied by its User Markov Chain, it is not necessary to include the tagged user in the System Markov Chain. As a result, the size of state space can be further reduced [67][68].

A quite different approach for the analysis of buffered multiple access protocols is proposed in [69][70] where the statistical behavior of the system is obtained using the characteristics of the restricted occupancy urn model. It defines the system state as the total number of packets in the system. The number of packets in each user buffer is determined then using the restricted occupancy urn model. The number of possible states of the system is further reduced to $N \cdot L + 1$. In addition, the obtained $N \cdot L + 1$ equations are linear. compared with the fact that a set of nonlinear equations would be obtained using almost any of the other approaches, this has significant advantage. However, even for this size of state

space, the analysis becomes much difficult for a system with $N = 100$ and $L = 20$. It should be noticed that this system model is multiple dimensional in characteristics even though the state space is one dimensional. A system is said to be a n dimensional system in characteristics if a n dimensional vector is required to describe the system naturally though one dimensional vector can be used to describe a n dimensional system through properly mathematical mapping.

It should be mentioned that the analysis complexity depends on not only the size of the system state space, but also its dimensions. It is observed that all of the above approaches employ Markov analysis on simplified system model whose state space has two or more dimensions. The computation of transition probability matrix is required in order to calculate the numerical results. Problems regarding numerical stability and convergence may arise. Consequently, these approaches may be not able to handle systems with large population or user buffer capacity.

Tasaka [45] studied the dynamic behavior of a homogeneous CSMA/CD system using EPA. Though the size and dimension of the state space are quite large, the amount of computation is quite small since no transition probability matrix is involved.

It is interesting to notice that although each user is a queueing system, almost all of the analysis approaches do not use the results existing in the field of queueing theory. We find in [45] a strong statement that the problem of interfering queues cannot be solved using classic queueing theory. Needless to say, this statement refers to the exact analysis only. For approximate analysis, the classical queueing theory can be used if the packet service time distribution can be obtained approximately.

Several attempts have been made for applying classical queueing theory to buffered systems [60][61]. Protocols such as URN and ALOHA are studied in [61]. In [60], CSMA systems are analyzed by treating each user as an M/G/1 queue. The packet service time distribution is obtained and then applied to the user queueing analysis. However, the analysis complexity prohibits them from being widely used in the analysis of buffered multiple access protocols. Hence, what restricts the application of classical queueing theory in the analysis of buffered random multiple access protocols is the difficulty to derive the packet service time distribution.

It is also noticed that almost all available approaches do not decouple the analysis on channel contention and the analysis on user queueing behavior. As the result, they cannot be extended to systems with more complicated queueing behavior.

In addition to the above analysis approaches for slotted multiple access protocols, there are some approaches for unslotted systems. In [63], an approximate approach is proposed for a continuous time CSMA/CD system with infinite buffer capacity. Each user is treated as a M/G/1 queueing system. The idea is to assume the existence of the packet service time distribution and attempt to find it so that the corresponding M/G/1 queueing system can be solved using classical queueing theory. Since immediate first transmission (IFT) model [44][71][72] is used, an arriving packet which initiating a busy period has to be distinguished from the other packets. As the result, the service time distributions for both kinds of packets have to be obtained in order to apply the results of [64] to this particular M/G/1 queueing system.

A continuous time CSMA/CD network with a buffered gateway is analyzed in [66]. This system consists of N homogeneous terminals with buffer capacity of one packet and a gateway with buffer capacity L packets. A two dimensional vector $I = [i_t, i_g]$ is used to describe the system states, where i_t denotes the number of backlogged terminals and i_g denotes the number of packets in the gateway buffer. The exact performance analysis is obtained using Markov analysis. In this system, only gateway is a queueing system with buffer capacity more than one packet, and the other terminals are unbuffered users.

2.4 Summary

Because of the analysis complexity of buffered random access protocols, approximation is necessary in order to develop the analytical methods properly. So far, there is no unified method for the performance analysis of buffered random access protocols. Many different assumptions, approaches and techniques are required to obtain the different performance measurements such as throughput, waiting time and interdeparture time.

Though the available analytical methods are quite different from one to another in establishing the analytical model and employing the mathematical tools, almost all of them have a state space of 2-dimension or more, and cannot handle the systems if the user population or buffer capacity is fairly large. In addition, like EPA, these approaches cannot apply the results in classical theory to the system analysis except [60], [61] and [63] even though each user is a queueing system in fact. This matter itself implies the complexity of those approaches in some degree.

Though the classical queueing theory cannot be applied to the exact system analysis of buffered random access protocols because the queues of the users interfere with each other

through a shared broadcast channel [73][74], it is possible to apply classical queueing theory to the approximate analysis [60][63]. An obvious advantage by using queueing theory is the analytical simplicity. The focus of the analysis can be put on the channel contention scheme, and the analysis on complicated user queueing behavior can be handled by well developed queueing theory. The remaining problem is how to find the packet service time distribution for the tagged user.

3 TAGGED USER ANALYSIS

3.1 Introduction

In this chapter, a new analytical approach, called tagged user approach (TUA), for multiple access system is proposed¹. It is an analytical method which can be applied to the analysis on many slotted random multiple access protocols with emphasis on the buffered systems. The idea of TUA is originated from a set of literatures typified by [60], [61] and [75]. It is backed up by two major branches of theory, i.e. dynamic probabilistic system and queueing theory. The attempts have been made in [60] and [61] to analyze the buffered MA systems using the classical queueing theory. However, their approaches are not widely employed since their methodologies for finding the packet service time distribution is very complicated. In [75], the state flow graph is employed to find the packet service time distribution in order to analyze an unbuffered TSMA/PCD system. Unfortunately, this idea has not been extended to the analysis of buffered MA systems.

Before we present the principle of TUA, we give the following definitions.

Definition 3.1 Group: A group is a set of users with the same system parameters so that we cannot distinguish one from another, i.e. the same packet arrival process, the same buffer capacity, and the same transmission permission probability, etc.

1. This work can also be found in [90].

Definition 3.2 Homogeneous Systems: A homogeneous system is a system consisting of only one group of users.

Definition 3.3 Heterogeneous Systems: A heterogeneous system is a system consisting of at least two groups of users.

Definition 3.4 Statistical Equivalence: Two systems are said to be statistically equivalent if they have the same distribution for any interested performance measurements.

Definition 3.5 Channel Symmetry: A channel is said to be symmetrical if it is fair to each user in the system, that is, each user in the same group will have statistically equivalent behavior.

To present TUA, a heterogeneous multiple access system with buffered users is considered which consists of several groups of users. The users with same system parameters are classified into one group. Because of channel symmetry assumption, the behavior of each user in the same group is statistically equivalent. However, the behaviors of users in different groups may be not statistically equivalent. A tagged user is selected randomly from each group. The system analysis is completed by studying the behaviors of all tagged users in the system.

In TUA, different service discipline can be handled, as discussed later one. Here we assume that each user in a system be a FCFS or FIFO (first-come-first-serve or first-in-first-out) single-server queueing system.

The basic idea of TUA is to treat each user in the system as an independent queueing system always operating with its own equilibrium probabilities though this is an interfering queue problem. The interaction among users is reflected by the fact that the equilibrium probabilities of each user are dependent on the behaviors of all users in the system.

Each user is considered as an independent queueing system consisting of a queue and a virtual server. A packet is said to be in virtual server if the packet is ready for transmission (service) or being transmitted (served). In FCFS service discipline, the head-of-line packet is the one in the virtual server.

Definition 3.6 Virtual Packet Service Time (VPST): The VPST is the time duration which a packet spends in the buffer from its ready for transmission to its departure.

VPST actually is the time duration which a packet spends in the virtual server. It is also called packet service time alternatively in this thesis. For any random access systems, the arrival process and service discipline are already given in advance. If the VPST distribution is known, classical queueing theory can be applied to each user. However, because of interactions among all users, VPST is not equal to the packet transmission time, but the packet transmission time plus a random channel access delay, which is determined by the behaviors of all users in the system (see Section 1.2.2). Hence, the key of TUA is to find the distribution of VPST for the tagged user in each group.

Definition 3.7 Busy or Idle Status: A user is said to be busy if its buffer is not empty. Otherwise, the system is idle.

In order to determine the distribution of VPST for each tagged user, state flow graph [76] is employed. Based on the analysis on channel contention scheme, it is found that VPST of each user is a function of the busy probabilities of all users and other system parameters. On the other hand, queueing theory shows that the busy probability of each queueing system is also a function of its VPST. The function relationships above are sufficient to determine the VPST and busy probability of users in each group. The performance analysis is straightforward after the VPST and user busy probabilities are obtained.

In multiple access protocols, the behavior of each user is not mutually independent. However, by assuming that each user always operates with its equilibrium probabilities, each user can be modeled as an independent queueing system always operating at its own equilibrium points. As the result, the analysis is decoupled into the analysis on queueing behavior of individual users and the analysis on channel contention scheme.

Similar to EPA and Markov analysis, there might be zero, one or two reasonable solutions for infinite buffer MA systems in TUA analysis, corresponding to unstable, globally stable and locally stable system, respectively. Hence, TUA can be applied to system stability analysis.

Using TUA, the following performance indices can be obtained: channel throughput, the distribution of packet waiting time or system response time, the distribution of queue length, packet blocking probability (for users with finite buffer capacity), distribution of packet service time, distribution of the interdeparture time in each user, and the mean interdeparture time of the channel. It should be mentioned that when we talk about the distribution we mean the probability generating function (PGF) since there is a one-to-one correspondence between them.

3.2 Model of A General System

Considering a general heterogeneous CSMA/CD system which consists of m groups of users. Each group is numbered between 1 and m . A channel is shared by all users through a slotted non-persistent CSMA/CD protocol. The users within each group are assumed to have same system parameters and exhibit the same statistical behavior. The number of users in Group i is denoted by N_i . Each user in Group i has a buffer capacity of L_i packets,

which might be infinite. Since each user is a discrete time queueing system, we need to specify a rule for the users with finite buffer capacity concerning the acceptance of a packet that arrives in the last slot of a service time when the system is full [79]. In this system, the late arrival model with delayed replacement rule is used.

There are two widely used transmission principles in random multiple access protocols, called immediate first transmission (IFT) and defer first transmission (DFT) [44][71][72]. With IFT, a packet arriving in an empty queue will be transmitted immediately if the current slot is idle. Otherwise it will be transmitted with a given permission probability in an idle slot. With DFT, all packets will be transmitted with same permission probabilities in each idle slot. Hence, the packet issuing a busy period will be treated differently from other packets with IFT, and all packets are treated equally with DFT. Here we define another transmission principle, called NFT (new packet first transmission), where a new packet and deferred packet will be transmitted with different permission probabilities. A packet is said to be a new packet when it becomes ready for service in current slot. It is a deferred packet otherwise. Note that DFT is a special case of NFT where new packets and deferred packets are assigned with same permission probability. In fact, a system using NFT would have better performance than that using DFT. An example will be given in Chapter 5.

The system operates as follows. Time is divided into time slots of length normalized to 1. The packet arrival processes of all users are mutually independent. The packet arrival process of users in Group i is Bernoulli such that in each slot a packet would arrive with probability λ_i , and would not arrive with probability $1 - \lambda_i$. All packet transmissions start and end at the slot boundary. The packet transmission time for users in Group i is a random

number of time slots with PGF of $H_i(z)$. The system uses NFT transmission principle. A busy user in Group i would sense the channel in each slot with probability q_i for a new packet and $p_{i, sen}$ for a deferred packet. If the channel is sensed to be idle, the user starts a transmission immediately. If the channel is sensed to be busy, the user backs off and senses the channel again in next slot with probability $p_{i, sen}$. The channel is noise free. The only reason resulting in an unsuccessful transmission is collision, which occurs when more than one user in the system starts transmission in same slots. All packets involving in a collision are destroyed. When a collision is detected by a transmitting user, the user aborts its transmission immediately. The number of time slots needed to detect a collision for users in Group i is a random variable with PGF of $D_i(z)$.

Note that $D_i(z)$ is very general so that this protocol would include S-ALOHA and slotted non-persistent CSMA as its special cases. In S-ALOHA, all packets have transmission duration of one slot. Hence, S-ALOHA is a special case of the system when $H_i(z) = D_i(z) = z$. CSMA is a special case of this protocol with $D_i(z) = H_i(z)$ since a station would transmit a whole packet no matter if the transmission is successful or not. In CSMA/CD, if all stations are assumed to be able to detect the collision in D slots, we have $D_i(z) = z^D$, where D is determined by the channel propagation delay. Normally $D = 1$ in Ethernet. However, in DSMA/CD, the MAC protocol of CDPD [77][97], where data is transmitted in Reed-Solomon encoded blocks, we have $D = 14$.

3.3 Queueing Analysis

Assume that the channel be symmetric, and the statistical behavior of all users in the same group be identical. A user is arbitrarily selected from each group as its tagged user. The analysis is then focused on the behaviors of all tagged users.

3.3.1 User Queueing Model

Each station is modeled as a FCFS single-server queueing system consisting of a queue and a virtual server. Clearly, this queueing model can be analyzed using classical queueing theory if the VPST distribution is known. Unfortunately, in random multiple access systems, VPST consists of access delay and transmission time. Though packet transmission time is known, the access delay is a random variable determined by system status. Hence, the distribution of VPST in each user is not predetermined, and can only be determined by the system parameters such as user population, permission probabilities, etc.

In general, each user can be modeled as a Geo/G/1/K queueing system. However, if the VPST is found to be geometrically distributed, each user may be modeled as a Geo/Geo/1/K system, which is much easier to be solved than Geo/G/1/K.

3.3.2 Distribution of Virtual Packet Service Time

The analysis is carried out using the state flow graph of tagged users [75][76]. The state flow graph of the tagged user of Group i , called tagged user i (TU i), is shown in Figure 3.1.

When TU i is busy, it can be in one of the four states, i.e. packet ready state S_0^i , transmission state S_{tx}^i , backoff state S_{bk}^i and departure state S_e^i . Whenever a packet enters the virtual server, the station switches to S_0^i . Since NFT transmission principle is employed, a

Figure 3.1. State flow graph of tagged user in Group i

station at S_0^i will transit to S_{tx}^i immediately with probability $q_i p_{i,l}$, where $p_{i,l}$ is the probability that a channel is sensed to be idle in a slot by TU i given that it senses the channel in that slot. Otherwise, the packet is deferred for one more slot and the station switches to S_{bk}^i . For a station in state S_{bk}^i , it will start transmission and switch to S_{tx}^i with probability $p_{i,ctx}$ at the beginning of each slot, or stay in S_{bk}^i for one more slot with probability $1 - p_{i,ctx}$, where $p_{i,ctx}$ is the probability that TU i at S_{bk}^i would start a transmission in a slot. When a station enters S_{tx}^i , it starts transmission immediately. Since the channel is shared by all stations in the system, a transmission cannot be guaranteed to be successful. Denoting by $p_{i,su}$ the probability of a successful transmission by tagged user i given that it transmits in a slot, and by $H_i(z)$ the PGF of the packet transmission time (in slots) in

Group i , the state transfer function from S_{tx}^i to S_e^i is given by $p_{i,su}H_i(z)$ as shown in Figure 3.1. If a transmission is not successful, which occurs with probability $1 - p_{i,su}$, the station will abort the transmission and backoff to S_{bk}^i after detecting the transmission failure. S_e^i is called departure state since a station transits to this state from S_{tx}^i at the moment of a packet departure. It should be mentioned that S_0^i and S_e^i are dummy states in which a user never spends any time.

The states defined here are used to describe the user behavior at channel contention. The sole purpose for defining them is to assist us find the PGF of VPST. They should not be confused with the user states defined in queueing analysis where the state is used to describe the number of packets in the buffer.

By the definition of virtual packet service time, it is just the time duration from a user's entering S_0^i to its exiting S_e^i . Using Mason's formula [81][82][83] on Figure 3.1, the PGF of the VPST of TU i , denoted by $B_i(z)$, is given by

$$B_i(z) = \frac{[q_i p_{i,l} - (q_i p_{i,l} - p_{i,tx})z] p_{i,su} H_i(z)}{1 - (1 - p_{i,tx})z - (1 - p_{i,su}) p_{i,tx} D_i(z)}, \quad i = 1, 2, \dots, m, \quad (3.1)$$

which implies that the queueing system maintained in each user has a generally distributed service time. Hence, the queueing system in each station is a discrete time FCFS queueing system with Bernoulli arrivals, general service time distribution and finite buffer capacity, or a $Geo/G/1/K$ system in Kendall's notation.

The transmission probability $p_{i,tx}$ is related to the random multiple access systems. For instance, it is a given constant in slotted ALOHA system, but a variable determined by system parameters and status in other systems such as CSMA/CD and PRMA. In CSMA and CSMA/CD, it is given by

$$p_{i,tx} = p_{i,sen}p_{i,I}, \quad (3.2)$$

where $p_{i,sen}$ is the permission probability that a backoffed station in Group i is permitted to sense the channel in any slot.

To simplify the analysis, we assume that each user in the system require D slots to detect the collision and abort its transmission. That is, $D_i(z) = z^D$, $i = 1, 2, \dots, m$. This assumption holds for CSMA/CD systems, but only for CSMA systems in which the packet transmission time of all users is equal to D slots.

So far, we find that $B_i(z)$ is a function of two unknowns $p_{i,I}$ and $p_{i,su}$ by (3.1) and (3.2). Intuitively, $B_i(z)$ is determined by the busy or idle status of all users in the system. In the following, we establish the necessary function relationships such that $B_i(z)$ is a function of only the idle probabilities of all tagged users.

3.3.2.1 Probability of Channel Sensed To Be Idle

In order to determine $p_{i,I}$, the behavior of channel has to be analyzed. Employing the classical scheme for the analysis of CSMA/CD, the time duration in the channel is thought to consist of a sequence of cycles [14][44]. Each cycle consists of an idle duration followed by a busy duration which corresponds to either a successful transmission or a collision.

Since TU i has packets available for transmission at this moment, it is busy with probability 1. All the other users in the system are busy with probability $p_{j,b}$, $j = 1, 2, \dots, m$, since they operate with their equilibrium probabilities by assumption. Denoting by I_i the idle duration length in a cycle observed by TU i given that it is busy, its mean is given by

$$E(I_i) = \frac{1}{1 - (1 - p_{i, sen})(1 - p_{i, tx})^{N_i - 1} \prod_{j \neq i} (1 - p_{j, tx})^{N_j}} - 1, \quad (3.3)$$

where $p_{j, tx}$ is the probability that a user in Group j would issue a transmission in an idle slot.

Clearly, only busy user is able to issue a transmission. When DFT rule is employed, we do not distinguish a new packet from a deferred packet. Equality $q_j = p_{j, sen}$ always holds. In this case, we simply have

$$p_{j, tx} = p_{j, b} p_{j, sen}. \quad (3.4)$$

However, $p_{j, tx}$ is much difficult to be determined with NFT rule. Since new packet and deferred packet have different permission probabilities, we must identify which kind of packets that a user would start to transmit in an idle slot. There are two possibilities: (1) A user in Group j would serve a just arriving packet with probability $p_{j,0} \lambda_j$ since it is idle with probability $p_{j,0}$ in any slot according to the assumption that a user always operate with its equilibrium probabilities. This packet is new and will be served with permission probability q_j . (2) A user in Group j is busy with probability $p_{j,b}$ in any slot by assump-

tion. Since it would serve a just arriving packet with probability $p_{j,0}\lambda_j$, it would serve a packet waiting in buffer with probability $p_{j,b} - p_{j,0}\lambda_j$. However, a waiting packet may be a new packet or deferred packet. Since a packet is new only in the first slot of its service time, the probability is generally considered to be ignorable if the channel access delay is sufficiently long. For instance, if the mean channel access delay is $\bar{d}_c$ slots, a waiting packet will be treated as a new packet for transmission with probability $1/\bar{d}_c$, which can be ignored safely in most of systems. Hence, we suppose that a user in Group j would serve a deferred packet with probability $p_{j,b} - p_{j,0}\lambda_j$. Consequently, we have

$$p_{j,tx} = p_{j,0}\lambda_j q_j + (p_{j,b} - p_{j,0}\lambda_j)p_{j,sen} = p_{j,b}p_{j,sen} + p_{j,0}\lambda_j(q_j - p_{j,sen}), \quad (3.5)$$

which includes the result of DFT rule as its special case as expected.

Denoting by $p_{i,k}$, $k = 1, 2, \dots, L_i$, the probability that a user in Group i has k packets in its buffer at any time, then we have

$$p_{i,b} + p_{i,0} = 1, \quad i = 1, 2, \dots, m. \quad (3.6)$$

Given that there is packet transmission in a slot, the probability $p_{i,s|tx}$ that the transmission is successful is given by

$$p_{i,s|tx} = q_{i,u} + q_{i,\bar{u}} + \sum_{j \neq i} q_{i,j}, \quad (3.7)$$

where

$$\begin{aligned}
q_{i,u} &= \frac{p_{i, sen}(1-p_{i, tx})^{N_i-1} \prod_{j \neq i} (1-p_{j, tx})^{N_j}}{1 - (1-p_{i, sen})(1-p_{i, tx})^{N_i-1} \prod_{j \neq i} (1-p_{j, tx})^{N_j}} \\
&= [E(I_i) + 1] p_{i, sen}(1-p_{i, tx})^{N_i-1} \prod_{j \neq i} (1-p_{j, tx})^{N_j}
\end{aligned} \tag{3.8}$$

is the probability that the successful transmission is contributed by the busy TU i ,

$$q_{i,\bar{u}} = [E(I_i) + 1](N_i - 1)p_{i, tx}(1-p_{i, sen})(1-p_{i, tx})^{N_i-2} \prod_{j \neq i} (1-p_{j, tx})^{N_j} \tag{3.9}$$

is the probability that the successful transmission is from the other users in Group i ; and

$$\begin{aligned}
q_{i,j} &= [E(I_i) + 1] N_j p_{j, tx} (1-p_{j, tx})^{N_j-1} \\
&\quad (1-p_{i, sen})(1-p_{i, tx})^{N_i-1} \prod_{k \neq j, k \neq i} (1-p_{k, tx})^{N_k}
\end{aligned} \tag{3.10}$$

is the probability that the successful transmission is from the users in Group j , $j \neq i$.

Given that there is a transmission in a slot, the probability that a collision would occur and involve TU i , denoted by $p_{i,ctu}$, is given by

$$p_{i,ctu} = \frac{p_{i, sen} \left[1 - (1-p_{i, tx})^{N_i-1} \prod_{j \neq i} (1-p_{j, tx})^{N_j} \right]}{1 - (1-p_{i, sen})(1-p_{i, tx})^{N_i-1} \prod_{j \neq i} (1-p_{j, tx})^{N_j}}. \tag{3.11}$$

While at S_{bk}^i , in each cycle TU i would sense the channel idle in each idle slot. In addition, it would sense the channel idle in the first slot of a busy duration which corresponds to either a successful transmission or a collision. Hence, in each cycle the total number of possible slots in which the backoffed TU i would sense the channel idle is $E(I_i) + 1$.

Obviously, in CSMA/CD protocol a user never attempts to sense the channel during its own transmission time except in the first slot of the transmission in which it has to sense the channel in order to initiate a transmission. Hence, in each cycle the total number slots in which the tagged user is able to sense the channel is given by

$$E(I_i) + 1 + (T_i - 1)q_{i,\bar{u}} + \sum_{j \neq i} (T_j - 1)q_{i,j} + (D - 1)(1 - p_{i,s|tx} - p_{i,ctu}),$$

where T_i is the mean packet transmission time in Group i , which is given by

$$T_j = H'_j(1), \quad j = 1, 2, \dots, m. \quad (3.12)$$

Therefore, the $p_{i,I}$ is finally obtained as

$$p_{i,I} = \frac{E(I_i) + 1}{E(I_i) + 1 + (T_i - 1)q_{i,\bar{u}} + \sum_{j \neq i} (T_j - 1)q_{i,j} + (D - 1)(1 - p_{i,s|tx} - p_{i,ctu})}. \quad (3.13)$$

For a general CSMA system in which the packet size is not equal, we may find $p_{i,I}$ using the same idea above. However, it is much more difficult because the length of the busy duration for each collision is determined by the maximum packet length of all packets involved in a collision. Consequently, each collision scenario has to be considered such that the analysis is intractable.

As to $p_{i,su}$, it is just the probability that none of the other busy users starts a transmission at the same slot given that TU i starts a transmission in that slot. Hence, we have

$$p_{i,su} = (1 - p_{i,tx})^{N_i - 1} \prod_{j \neq i} (1 - p_{j,tx})^{N_j}, \quad i, j = 1, 2, \dots, m. \quad (3.14)$$

Equations (3.2)-(3.14) show that all the unknown parameters in (3.1) can be expressed as a function of $p_{i,0}$, $i = 1, 2, \dots, m$. Hence, it is sufficient to conclude that $B_i(z)$ can be expressed as a function of the busy or idle probabilities of all stations in the system. That is, generally we have

$$B_i(z) = f_i(p_{1,0}, p_{2,0}, \dots, p_{m,0}), \quad i = 1, 2, \dots, m, \quad (3.15)$$

where $f_i(\cdot)$ denotes a function relationship. In fact, (3.15) is true for any random multiple access protocols.

It should be noticed that the analytical errors are introduced from two sources. One is associated with the inherent assumption in TUA that users always operate with their equilibrium probabilities. The other is associated with the technique for finding $p_{j,l}$ and $p_{j,su}$ for $j = 1, 2, \dots, m$. The latter is dependent on the specific systems.

3.3.3 Queueing Analysis of Tagged Users

With (3.15) in hand, the queueing system for the tagged user in any groups can be solved using classical queueing theory. For the model given in Section 3.2, each user is treated as a Geo/G/1/ or Geo/G/1/K queueing system, depending on whether its buffer capacity is infinite or finite. The queueing analysis follows Section B.5 and Section B.6 entirely, which is presented here for completeness.

3.3.3.1 Finite Buffer Capacity

Denoting by $a_{i,k}$ the probability that k packets arrive in the buffer of TU i during a packet service time. The PGF $A_i(z)$ of the number of packets arriving at the buffer of TU i during a packet service time is given by

$$A_i(z) = B_i(1 - \lambda_i + \lambda_i z). \quad (3.16)$$

Denoting by $\pi_{i,j}$, $j = 0, 1, \dots, L_i - 1$ the equilibrium probability that a packet in Group i sees j packets in the buffer at departure time. Since a departure packet takes one space before its leaving, it can see at most $L_i - 1$ packets in the queue. Hence, the following equations are obtained [79]:

$$\pi_{i,k} = \pi_{i,0} a_{i,k} + \sum_{j=1}^{k+1} \pi_{i,j} a_{i,k-j+1}, \quad 0 \leq k \leq L_i - 2, \quad (3.17)$$

$$\sum_{k=0}^{L_i-1} \pi_{i,k} = 1. \quad (3.18)$$

An efficient algorithm for computing $\pi_{i,k}$, $0 \leq k \leq L_i - 1$, can be found in [79][80] by defining

$$\pi'_{i,k} = (\pi_{i,k} / \pi_{i,0}), \quad k = 0, 1, \dots, L_i - 1. \quad (3.19)$$

Then it is found using (3.17)-(3.19) that $\pi'_{i,k}$, $0 \leq k \leq L_i - 1$ can be recursively calculated as follows:

$$\begin{aligned} \pi'_{i,0} &= 1, \\ \pi'_{i,k+1} &= \frac{1}{a_{i,0}} \left(\pi'_{i,k} - \sum_{j=1}^k \pi'_{i,j} a_{i,k-j+1} - a_{i,k} \right), \quad 0 \leq k \leq L_i - 2. \end{aligned} \quad (3.20)$$

As the result, $\pi_{i,0}$ is found using (3.18) and (3.19) as

$$\pi_{i,0} = \left(\sum_{k=0}^{L_i-1} \pi'_{i,k} \right)^{-1} \quad (3.21)$$

$$\pi_{i,k} = \pi_{i,0} \pi'_{i,k}, \quad k = 1, 2, \dots, L_i - 1.$$

Let $p_{i,k}$, $0 \leq k \leq L_i$, be the probability that there are k packets present in TU i at an arbitrary time (slot boundary). Using the results of Geo/G/1/K system [79] again, we have

$$p_{i,k} = \frac{\pi_{i,k}}{\pi_{i,0} + \rho_i}, \quad 0 \leq k \leq L_i - 1 \quad (3.22)$$

$$p_{i,L_i} = 1 - \frac{1}{\pi_{i,0} + \rho_i} = p_{i,B},$$

where

$$\rho_i = \lambda_i b_i, \quad (3.23)$$

$$b_i = B'_i(1).$$

In (3.23), b_i is the mean VPST of Group i . Equation $p_{i,B} = p_{i,L_i}$ in (3.22) is obvious since the packet blocking probability $p_{i,B}$ of users in Group i is just the probability that an arriving packet in TU i finds the queue full and gets dropped consequently, which happens to be p_{i,L_i} .

3.3.3.2 Infinite Buffer Capacity

If the users in Group i have a very large buffer capacity such that almost all arriving packets are accepted into the buffer, each user in Group i can be modeled as a Geo/G/1 queueing system.

Denoting by $\pi_{i,k}$ and $p_{i,k}$, $0 \leq k \leq \infty$ the probability that a leaving packet sees k packets in queue and the probability that there are k packets present in the queueing system at any time in Group i , respectively. Then it is found [79] that the corresponding PGFs of them, denoted by $\Pi_i(z)$ and $P_i(z)$ respectively, are given by

$$P_i(z) = \Pi_i(z) = \frac{(1 - \rho_i)(1 - z)B_i(1 - \lambda_i + \lambda_i z)}{B_i(1 - \lambda_i + \lambda_i z) - z}, \quad (3.24)$$

where ρ_i is still given by (3.23). It is noticed that $P_i(z) = \Pi_i(z)$ holds only for infinite buffer system. In this case,

$$p_{i,0} = 1 - \rho_i. \quad (3.25)$$

Section 3.3.3.1 and 3.3.3.2 show that $p_{i,0}$ is solely determined by $B_i(z)$. As a matter of fact, no matter which queueing model is used, it is always true that a station's idle probability is a function of the distribution of service time given the arrival processes and queue service discipline. That is, for users in Group i , generally we have

$$p_{i,0} = h_i(B_i(z)), \quad i = 1, 2, \dots, m, \quad (3.26)$$

where $h_i(\cdot)$ denotes a function relationship. Equations (3.15) and (3.26) are sufficient to determine $B_i(z)$ and $p_{i,0}$ for $i = 1, 2, \dots, m$. As the result, all the performance indices about each user can be obtained. After that, it is straightforward to determine the performance of the channel.

It is found that the procedure for obtaining (3.15) is independent of the arrival processes and service principles of users. Hence, TUA can be applied to systems with arbitrary arrival processes and service principles without introducing extra difficulty since the difficulty is

already handled by the classical queueing theory. Based on the arrival processes and buffer capacity, different queueing models can be employed. For instance, if the system has bulk arrival following Bernoulli process and infinite buffer capacity, $Geo^{[x]}/G/1$ is the proper model for each user. In addition, TUA can also be applied to the systems where the packet service principle in users may not be FCFS such as LCFS (last come first serve) system, random service system, or priority queueing system as long as the corresponding queueing analyses are available.

In this chapter, the channel is considered to be perfect and only collision would result in transmission failure. However, TUA can be extended to the analysis of noise channel such as a random packet error channel or a wireless channel with shadowing, fading and channel capture. Reviewing Figure 3.1 and the analysis in Section 3.2 and 3.3, it is found that the only parameter dependent on channel environment is $p_{i, su}$, $i = 1, 2, \dots, m$. By modifying (3.14) properly based on the channel characteristics, TUA can be applied to the analysis of the system with noise channel. An example will be given in Section 5.1.

3.4 Performance Evaluation

In this section, we evaluate the system performance using the results obtained above. The basic idea in TUA is to understand the behavior of each tagged user. The performance analysis is accomplished if the behaviors of all users are known.

Since users in a group may have finite or infinite buffer capacity, both cases are dealt with in this section. The discussions on these two cases are distinguished only if it is necessary. Otherwise, the obtained results are valid in both of them.

3.4.1 Throughputs

Throughput is defined as the number of successfully transmitted information bits in each bit time. In this section, channel throughput and user throughput are given. The channel throughput is a measurement of the channel utilization. The user throughput is just the throughput of each user. We emphasize again that the statistical behavior of each user in the same group is identical and equal to the behavior of the tagged user of its group.

Since all the packets accepted into the queue would be transmitted successfully sooner or later, the user throughput is equal to the average information bits accepted by the queue in one bit time. Hence, the throughput θ_i of the tagged user in Group i is given by

$$\theta_i = \lambda_i T_i (1 - p_{i,B}), \quad (3.27)$$

where T_i is given in (3.12), and $p_{i,B}$ is the packet blocking probability of users in Group i . Eventually $p_{i,B}$ is the probability that the user buffer in Group i is seen full by arriving packets. If users in Group i have infinite buffer capacity, we have $p_{i,B} = 0$ since all arriving packets can be accepted into the buffer. Otherwise, from (3.22) it is given by

$$p_{i,B} = p_{i,L_i}, \quad (3.28)$$

which is rewritten here since packet blocking probability itself is an important performance parameter. The channel throughput Θ is just the summation of the throughputs from all users. Hence, we have

$$\Theta = \sum_{i=1}^m N_i \theta_i. \quad (3.29)$$

3.4.2 Queue Length of Users

If the users in Group i have finite buffer capacity, the probability distribution of queue length is given in (3.22). The mean queueing length of users in Group i , denoted by l_i , is therefore given by

$$E(l_i) = \sum_{k=0}^{L_i} k p_{i,k}. \quad (3.30)$$

If the users in Group i have infinite buffer capacity, the PGF of queue length is given in (3.24), and the mean queue length is given by

$$E(l_i) = P_i'(1) = \rho_i + \frac{\lambda_i^2 b_i^{(2)} - \lambda_i \rho_i}{2(1 - \rho_i)}, \quad (3.31)$$

where ρ_i is given by (3.23), and

$$b_i^{(2)} = E(X_i^2) = B_i''(1) + B_i'(1), \quad (3.32)$$

where X_i is the VPST of packets in Group i .

3.4.3 Waiting Time, Response Time and Access Delay

Clearly, packet waiting time, response time and access delay do not make sense for packets rejected by the system. They are applicable only in the packets accepted into the systems. In the analysis, the time unit for all time amount is slot.

Denoting by w_i and r_i the packet waiting time and response time of users in Group i , respectively, if users in Group i have finite buffer capacity, the PGF $W_i(z)$ of the packet waiting time is given by using (B.27)

$$W_i(z) = \frac{\pi_{i,0}(1-z)}{1-\lambda_i-z+\lambda_i B_i(z)} \left(1 - \left[\frac{\lambda_i B_i(z)}{z-1+\lambda_i} \right]^{L_i} \right) + [B_i(z)]^{L_i-1} \sum_{j=0}^{L_i-1} \pi_{i,j} \left(\frac{\lambda_i}{z-1+\lambda_i} \right)^{L_i-j}. \quad (3.33)$$

Therefore, the mean waiting time of packets in Group i is given by

$$E(w_i) = W_i'(1) = \frac{E(l_i)}{\lambda_i(1-p_{i,B})} - b_i. \quad (3.34)$$

Since the packet response time is equal to waiting time plus VPST, the PGF $R_i(z)$ of packet response time in Group i is given by

$$R_i(z) = B_i(z)W_i(z). \quad (3.35)$$

The mean packet response time is given by

$$E(r_i) = R_i'(1) = \frac{E(l_i)}{\lambda_i(1-p_{i,B})}, \quad (3.36)$$

which can be obtained directly using Little's law also.

If users in Group i have infinite buffer capacity, the results are much simple. The PGF of mean waiting time and response time are given by

$$W_i(z) = \frac{(1-\rho_i)(1-z)}{1-\lambda_i-z+B_i(z)} \quad (3.37)$$

and

$$R_i(z) = \frac{(1-\rho_i)(1-z)B_i(z)}{1-\lambda_i-z+B_i(z)}, \quad (3.38)$$

respectively.

By definition, channel access delay is equal to VPST minus packet transmission time. Therefore, the PGF $\Xi_i(z)$ of channel access delay of packets in Group i , no matter the user buffer capacity, is given by using (3.1)

$$\Xi_i(z) = \frac{P_{i, stx}P_{i, su}}{1 - (1 - p_{i, stx})z - (1 - p_{i, su})P_{i, stx}D_i(z)}. \quad (3.39)$$

The mean channel access delay is therefore equal to $b_i - T_i$.

3.4.4 Interdeparture Time

Denoting by $\bar{d}_i$ the mean interdeparture time of users in Group i . For a group whose user has infinite buffer capacity, we have $\bar{d}_i = 1/\lambda_i$. For a group whose user has finite buffer capacity, we have $\bar{d}_i = 1/[\lambda_i(1 - p_{i, B})]$, where $p_{i, B}$ is the packet blocking probability in Group i .

The packet service rate of each user in Group i is therefore given by $1/\bar{d}_i$ packets per slot. As the result, the total packet service rate of the channel is $\sum_{i=1}^m N_i/\bar{d}_i$ packets per slot.

Hence, the mean channel interdeparture time, denoted by $\bar{d}_{ch}$, is given by

$$\bar{d}_{ch} = 1 / \left(\sum_{i=1}^m \frac{N_i}{\bar{d}_i} \right) = 1 / \left(\sum_{i=1}^m N_i \lambda_i (1 - p_{i, B}) \right). \quad (3.40)$$

3.4.4.1 Output Process of Users

Here we are interested in the PGF of interdeparture time of users in Group i . Let t_0 and t_1 be two contiguous departure points and $n_i(t)$ be the number of packets in the queue at

time t . The interdeparture time $\tilde{t}$ is therefore given by $t_1 - t_0$. Denoting by $O_{i, idle}(z)$ the output process given $n(t_0) = 0$. Then, the output process $O_i(z)$ of users in Group i is given by

$$\begin{aligned} O_i(z) &= E(z^{\tilde{t}} | n_i(t_0) = 0) \cdot \pi_{i,0} + E(z^{\tilde{t}} | n_i(t_0) \neq 0) \cdot (1 - \pi_{i,0}) \\ &= \pi_{i,0} O_{i, idle}(z) + (1 - \pi_{i,0}) B_i(z) . \end{aligned} \quad (3.41)$$

As to $O_{i, idle}(z)$, since the interdeparture time given $n_i(t_0) = 0$ is equal to idle time, which is geometric with mean $1/\lambda_i$ plus VPST, it is given by

$$O_{i, idle}(z) = \frac{\lambda_i z}{1 - (1 - \lambda_i)z} B_i(z) . \quad (3.42)$$

Substituting (3.42) into (3.41) and cancelling the term $O_{i, idle}(z)$, we have

$$O_i(z) = B_i(z) + \frac{\pi_{i,0}(z-1)B_i(z)}{1 - (1 - \lambda_i)z} . \quad (3.43)$$

3.4.4.2 Output Process of The Channel

The output process of the channel is much more interesting than that of users. The analysis of the distribution of the interdeparture time in the channel is important in relation to the interconnected local area networks or multihop networks. For example, a channel output process may constitute part of the input processes to other adjacent networks. The channel output process is therefore important in order to analyze the performance of interconnected networks.

One very difficult problem is how to obtain the distribution of the channel interdeparture time. By our knowledge, there are no publications available so far on the analysis of

the distribution of channel interdeparture time of buffered multiple access protocols. Section 3.4.1 inspires that only equilibrium probabilities of all users are not sufficient. The state evolution of all users are also required in order to determine the channel output process $O_{ch}(z)$. Unfortunately, this information is not available in TUA. It can be obtained only by Markov analysis, which cannot be applied to the analysis of buffered random multiple access systems as mentioned earlier.

An approximate analysis of the channel output process is proposed in this section using the state flow graph in Figure 3.2, where each user is supposed to operate with its equilibrium probabilities at any time. To simplify the analysis, it is assumed that $D_i(z) = z^D$ for $i = 1, 2, \dots, m$. The packet departure times are selected as the imbedded points for the analysis. In Figure 3.2, g_0 and g_e are the n th and $(n+1)$ th packet departure points from the channel, respectively, and g_{idle} is the status when channel is idle. The channel interdeparture time is just the transition time from g_0 to g_e .

Figure 3.2. State flow graph of the channel

After the n th packet departure, the channel is idle and enters g_{idle} . If there is only one user which senses the channel at this slot, this user would initiate a successful transmission, and the channel status will enter g_e immediately. Therefore, we have

$$f_1(z) = \sum_{i=1}^m N_i p_{i, su|id} H_i(z), \quad (3.44)$$

where

$$p_{i, su|id} = p_{i, b} p_{i, sen} (1 - p_{i, b} p_{i, sen})^{N_i - 1} \prod_{j \neq i} (1 - p_{j, b} p_{j, sen})^{N_j}. \quad (3.45)$$

is the probability that a user in Group i would start a successful transmission at a slot given that the channel is idle.

When no user senses the channel at a slot, the channel remains idle for one more slot and stays in g_{idle} . If more than two users sense the channel at an idle slot, a collision occurs and the channel keeps busy for D slots before going back to idle again. Hence, we have

$$f_2(z) = \prod_{i=1}^m (1 - p_{i, b} p_{i, sen})^{N_i} z + \left[1 - \prod_{i=1}^m (1 - p_{i, b} p_{i, sen})^{N_i} - \sum_{i=1}^m N_i p_{i, su|id} \right] z^D. \quad (3.46)$$

Therefore, the channel output process $O_{ch}(z)$ is given by

$$O_{ch}(z) = \frac{f_1(z)}{1 - f_2(z)}. \quad (3.47)$$

Theoretically, the mean channel interdeparture time in (3.40) is not equal to that obtained using (3.47) because different approach is used. However, the difference between them is very small as indicated by numerical results. Moreover, the two results are almost equal when users are almost always busy. That is, (3.47) is a very good approximation with heavy

traffic assumption. This is true since in this case we do not need to consider the state evolution of all users since they are always in busy status.

An assumption implied by Figure 3.1 is that the slot immediately following the end of a successful transmission can be sensed to be idle by each user. Immediate acknowledgment holds for systems where acknowledgment delay is negligible such as PRMA. However, in some CSMA/CD systems this slot may be sensed to be busy by all users due to propagation delay. These systems can be modeled using the general model in Section 3.2 by considering the slot after a successful transmission as the last slot of a transmission. That is, each packet length is increased by one slot. When calculating the system throughput, this slot is removed since there is no data transmission at all at this slot. For instance, if the users in Group i has fixed packet size τ_i slots, we have $H_i(z) = z^{\tau_i + 1}$. Then the system is analyzed as above. However, if the user throughput obtained using the model in Figure 3.1 is θ_i . The real throughput of the tagged user is given by $\theta_i \tau_i / (\tau_i + 1)$.

3.5 Accuracy of TUA

TUA is an approximate approach developed for the analysis of buffered random access protocols. To understand the accuracy of TUA is important in order to apply it to the system analysis. There are two general ways to evaluate the accuracy of an approach. One is to compare it with an existing exact approach. The other is to compare it with simulation. In this section, we focus on the former. The latter would be considered in Chapter 4 to Chapter 8 when TUA is applied to several typical random multiple access protocols. A problem associated with the former is that Markov analysis is the only available exact

approach, which unfortunately cannot be applied to buffered random multiple access protocols due to the multi-dimension and huge size of its state space. However, Markov analysis can be applied to unbuffered systems, which are a special case of buffered systems with buffer capacity of one unit. Hence, firstly we compare the results of TUA and Markov analysis on an unbuffered slotted ALOHA system.

In this section, the results of TUA are compared with that of Markov analysis on a homogeneous unbuffered S-ALOHA system with IFT transmission principle. Obviously, in an unbuffered system, IFT does not distinguish from NFT with $q_j = 1$ for $j = 1, 2, \dots, m$. The system model and Markov analysis can be found in [1][12]. The results for three different system configurations are given in Figure 3.3, Figure 3.4 and

Figure 3.3. Channel throughput and packet response time of a S-ALOHA with $N = 100$ and $\lambda = 0.0042$

Figure 3.5. In all cases, the packet generating rate of the system is $N\lambda = 0.42$ packets per

Figure 3.4. Channel throughput and packet response time of a S-ALOHA with $N = 60$ and $\lambda = 0.007$

Figure 3.5. Channel throughput and packet response time of a S-ALOHA with $N = 20$ and $\lambda = 0.021$

slot. It is shown that the accuracy of TUA is very high for sufficiently small or large permission probability p in each case. It should be noticed that both sufficiently small and large permission probability would lead to high user busy probability. The former is due to infre-

quently transmission attempts, and the latter due to frequently collisions. In fact, when user busy probability is 1, equation (3.1) gives the exact PGF of VPST since the assumption that users always operate at equilibrium point is satisfied, and (3.5) does not introduce any approximate errors. In this case, TUA gives the exact performance analysis. However, in the region of p where throughput has a downward jump and packet delay has a upward jump, the error of TUA is relatively large. In addition, it shows that the larger the user population N , the higher the accuracy of TUA.

It is observed that there is a sharp corner for the TUA analytical solution in Figure 3.3. This is because the curve on the left hand side of the downward jump corresponds to a system operating at its globally stable equilibrium point, and that on the right hand side to a system at its saturated equilibrium point. The detail is discussed in Section 4.5.2.

The larger user population N leads to higher accuracy of TUA might be explained as follows. Take two users from the system and name them A and B . Let

$$\Delta = Pr(A \text{ busy} | B \text{ busy}) - Pr(A \text{ busy} | B \text{ idle}). \quad (3.48)$$

Clearly, if each user always operate with its equilibrium probabilities, then we should have $\Delta = 0$. The fact is that B will contend for the channel if it is busy. Hence, the behavior of user B would interfere with the operation of user A . Among the total interferences applied on A , only $1/(N-1)$ is contributed by B . The larger the N , the weaker the portion of interference contributed by B on A . If N is sufficiently large, $|\Delta|$ is expected to be very close to zero. In this case, B would always see A operating with its equilibrium probabilities. Now, consider B as the tagged user and A as any one of other users. To the tagged user, the other

users are more and more like operating with its equilibrium probabilities with the increasing of N . That is why the analytical error decreases with the increment of user population N .

In both Figure 3.5 and Figure 3.6, the system has the same user population $N = 20$ but

Figure 3.6. Channel throughput and packet response time of a S-ALOHA with $N = 20$

different packet arrival rates. It is found that the accuracy of TUA increases with λ . Since p_b would increase with λ , the above observation implies that TUA has satisfied accuracy for sufficiently large user busy probability as discussed above.

It should be pointed out that the region of retransmission probability p leading to large analytical error of TUA is in fact not the expected operating region of the system operation. To reach large throughput and small packet delay, the expected operating region should be around 0.038, 0.06 and 0.15 for the system configurations in Figure 3.3, Figure 3.4 and Figure 3.5, respectively. The proper margin is required so that the operating region is not

close to the point where throughput downward jump occurs. It is found that even for mid large N , the analysis of TUA is quite accurate for the expected region of p .

Though the conclusions above are drawn for a unbuffered S-ALOHA system, they are expected to be reasonable to buffered S-ALOHA systems since the PGF of VPST is determined by the number of busy users, no matter the buffer capacity and the number of packets in each busy user. TUA analysis implies that the number of busy users at any time is governed by binomial distribution. This assumption is justified for systems with sufficiently large user busy probability. Since the purpose to introduce buffer is to reduce the packet rejection rate, buffered systems are expected to have larger user busy probability than unbuffered systems when other system parameters are fixed. As discussed before, larger user busy probability would lead to higher accuracy. Hence, more satisfied accuracy is expected on the TUA analysis of buffered systems. In the remaining chapters, the accuracy of TUA analysis on buffered systems will be verified using simulation.

The results of TUA analysis on an infinite buffer S-ALOHA system are compared with that in [51], which gives the lower and upper bounds of the mean buffer size. The bounds in [51] are tight for small N , but not tight for large N , unfortunately. The results are given in Table 3.1 and Table 3.2, where λ is the packet arrival rate per slot per user. The symbol p denotes packet retransmission probability. EQ is the mean queue length obtained using TUA. EQL and EQH are the lower and upper bounds of the mean queue length obtained in [51], respectively.

As shown in Table 3.1, for small user population $N = 3$, EQ is less than the lower bound EQL for $\lambda \geq 0.02$. However, their values are pretty close even in that case. In the

Table 3.1. Mean queue length of buffered S-ALOHA system with $N = 3$ and $p = 0.1$

λ	EQ	EQL	EQH
0.01	0.1125	0.1119	0.1133
0.02	0.2583	0.2590	0.2622
0.03	0.4568	0.4618	0.4676
0.04	0.7465	0.7610	0.7707
0.05	1.2151	1.2500	1.2661
0.06	2.1181	2.2000	2.2285
0.07	4.6388	4.8681	4.9318
0.08	57.2874	60.800	61.600

Table 3.2. Mean queue length of buffered S-ALOHA system with $N = 10$

λ	$p = 0.05$			$p = 0.15$		
	EQ	EQL	EQH	EQ	EQL	EQH
0.005	0.1165	0.1132	0.1221	0.0360	0.0349	0.1124
0.01	0.2809	0.2662	0.3111	0.0787	0.0733	0.2684
0.02	0.9921	0.9391	1.4411	0.1975	0.1673	0.9154
0.025	2.1088	2.0671	3.4789	0.2879	0.2133	1.9854
0.03	10.329	10.714	18.002	0.4251	0.2943	4.3214
0.0315	1287.8	∞	∞	0.4837	0.3262	8.2754

case of $N = 10$, EQ is located between the lower bound EQL and upper bound EQH except when $\lambda \geq 0.03$ and $p = 0.05$. Hence, with the increasing of user population N , the analysis of TUA becomes more accurate. This agrees well with our previous analysis. The results here also imply that TUA tends to overestimate the system performance.

3.6 Summary

In this chapter, a new approach, called tagged user analysis (TUA), is proposed for the performance analysis of a class of contention based multiple access protocols approxi-

mately. Each user in TUA is modeled as an independent queueing system which operates with its own equilibrium probabilities. The interactions among users through sharing the common channel are reflected by the fact that the equilibrium probabilities of one user are dependent on the behaviors of all users. Classical queueing theory is employed for the system analysis so that TUA can be applied to the analysis of complicated random access protocols such as heterogeneous system. TUA can handle the system in which user buffer may have arbitrary capacity (in packets), arbitrary packet service principle, and arbitrary arrival processes. In addition to the mean values, the distribution of almost all interested performance indices are obtained. Though TUA is presented to analyze S-ALOHA, CSMA, and CSMA/CD here, it can be applied to other random multiple access systems such as R-ALOHA and PRMA, as demonstrated in Chapter 7 and Chapter 8. TUA is a unified analysis approach for the queueing analysis of random multiple access protocols.

4 S-ALOHA SYSTEMS

4.1 Introduction

Slotted ALOHA (S-ALOHA) is widely used in wireless communications not only by itself as a multiple access protocol but also as a component in many reservation protocols such as R-ALOHA [1][33] and PRMA [36]. In this chapter, we study the performance of S-ALOHA systems using TUA [91]. The stability analysis is also presented. Although there are literatures on the stability analysis of unbuffered ALOHA systems [12][13], and queueing analysis of buffered ALOHA systems [52][67], there are no publications available for the stability analysis of buffered S-ALOHA systems. The region of retransmission probability for stabilizing and optimizing the system performance is also given. In addition, a complete equilibrium point analysis is also presented here for S-ALOHA systems with finite buffer capacity to identify the range of permission probability so that the optimal system performance is expected.

4.2 System Model

The system used here is a homogeneous slotted ALOHA system with finite user population of N users and finite buffer capacity of L packets. When L is infinite, we obtain the model with infinite buffer capacity. All packets are equal in size and need one slot of transmission time. For a discrete time system with a finite queue capacity, we need to specify a

rule concerning the acceptance of a packet that arrives in the last slot of a service time (where a successful transmission occurs) when the system is full [79]. In this system, the late arrival model with delayed replacement rule is used. Moreover, we assume the DFT transmission principle for the system.

The system operates as follows:

- 1) Time is divided into slots of length normalized to 1 packet transmission time. All packet transmissions start and end at the slot boundary.
- 2) A busy user transmits a packet in each slot with probability p .
- 3) The channel is noise free. The only reason resulting in unsuccessful transmission is collision, which occurs when more than one user starts transmission at the same slot. No packets can survive the collision.
- 4) All packets are of the same length. The transmission time of each packet is one slot. A station would know the transmission status (success or failure) of a packet immediately after finishing the transmission. If the transmission is successful, the packet is departed immediately. Otherwise, the packet is retransmitted in next slot with probability p .
- 5) Packets arrive in the queue in each station following a Bernoulli process such that a message arrives in a slot with probability λ and no messages arrive in a slot with probability $1 - \lambda$.
- 6) The packet service in user buffer is based on first come first served (FCFS). The arriving packets finding the queue full get dropped and do not return.

4.3 System Analysis

With the assumption that the channel is symmetric, we focus on the behavior of the tagged user and the impact of other users on it through sharing the channel. Since the system is homogeneous, the group identification number is dropped.

The analysis is carried out using the state flow graph of the tagged user as shown in Figure 4.1. State S_0 is the packet-ready state to which a busy user enters when it has new

Figure 4.1. State flow graph of tagged user for a S-ALOHA system

packet ready for service. Since a new packet and a retransmitting packet are not distinguished in the system, the tagged user would enter S_{bk} from S_0 immediately. When a user starts transmission in a slot with probability p , it enters the transmission state S_{tx} . If the transmission fails at the end of a slot, the tagged user returns to the backoff state S_{bk} . Otherwise, it enters the departure state S_e and departs the transmitted packet. p_s is the probability of successful transmission given a packet is transmitted. In fact, both S_0 and S_e are dummy states in which a user does not spend any time.

From Figure 4.1, the packet service time is just the time duration from a user entering S_0 to its exiting S_e . Using Mason's formula [81], the PGF $B(z)$ of packet service time is given by

$$B(z) = \frac{pp_s z}{1 - (1 - pp_s)z}. \quad (4.1)$$

That is, the packet service time is geometrically distributed with mean $1/(pp_s)$. Hence, each user in the system is just a Geo/Geo/1 (for infinite queue capacity) or Geo/Geo/1/K (for finite buffer capacity) system. It becomes clear now why so many publications choose DFT instead of IFT model for the system analysis.

In fact, the conclusion above can be made by the following arguments. For DFT model, a busy user would transmit a packet with probability p in each slot. The transmission would success with probability pp_s , or fail with probability $1 - pp_s$. Therefore, the service time is geometrically distributed with mean $1/(pp_s)$. The advantage of state flow graph is that the PGF of packet service time for more complicated system can be obtained where intuition fails to work.

As to p_s , it is just the probability that no other users transmit at the same slot. Therefore, we have

$$p_s = (1 - p_b p)^{N-1}, \quad (4.2)$$

where p_b is the user busy probability.

Now we have obtain the PGF $B(z)$ of packet service time by studying the contention scheme. Though p_s is unknown, we know that each user is a Geo/Geo/1/K system. The remained work is to apply the queueing theory of Geo/Geo/1/K to the system analysis. In the following, we will study the S-ALOHA systems with finite buffer capacity first, and the systems with infinite buffer capacity are treated as the special case of the systems with finite buffer capacity.

4.3.1 S-ALOHA with Finite Buffer Capacity

Denoting by p_k , $k = 1, 2, \dots, L$, the probability that there are k packets in tagged user at any time boundary, clearly we have

$$p_b + p_0 = 1. \quad (4.3)$$

Following the analysis on Geo/Geo/1/K system in Appendix B.3, we have

$$p_0 = \left(\frac{\mu}{\mu - \lambda} - \frac{\lambda^2}{\mu(\mu - \lambda)} \left[\frac{\lambda(1 - \mu)}{\mu(1 - \lambda)} \right]^{L-1} \right)^{-1} \quad (4.4)$$

and

$$\begin{aligned} p_i &= \left[\frac{\lambda(1 - \mu)}{\mu(1 - \lambda)} \right]^{i-1} \frac{\lambda}{\mu(1 - \lambda)} p_0, \quad i = 1, 2, \dots, L-1 \\ p_L &= \frac{\lambda}{\mu} \left[\frac{\lambda(1 - \mu)}{\mu(1 - \lambda)} \right]^{L-1} p_0. \end{aligned} \quad (4.5)$$

where the packet service rate is given by

$$\mu = p p_s = p[1 - (1 - p_0)p]^{N-1}. \quad (4.6)$$

The two equations (4.4) and (4.6) are sufficient to solve the two unknown variables μ and p_0 . After p_0 is obtained, all p_i , $i = 1, 2, \dots, L$, can be obtained using (4.5). Therefore,

the Geo/Geo/1/K queueing system is in fact solved. Unfortunately, (4.4) and (4.6) are non-linear equations. It is very difficult, if not impossible, to solve them analytically. A numerical algorithm has to be developed, which fortunately can be found without difficulty.

Before we present the numerical algorithm, let us establish the relations between p_i , $i = 1, 2, \dots, L$, and the interested performance indices of the systems. Recalling that the PGF of the packet service time is given by (4.1) and (4.6) as

$$B(z) = \frac{\mu z}{1 - (1 - \mu)z}. \quad (4.7)$$

4.3.1.1 Blocking Probability and Service Grade

The packet blocking probability p_B is the probability that an arriving packet finds the queue full and therefore gets lost. Since the queue is full with probability p_L , we have

$$p_B = p_L. \quad (4.8)$$

4.3.1.2 Channel Throughput

Channel throughput is defined as the number of successfully transmitted packets in each time slot. The channel throughput is a measurement of the channel utilization. We emphasize again here that the statistic behavior of each user in the system is identical and equal to the behavior of the tagged user.

The user throughput is equal to the average number of packets accepted by the queue in each slot. It is given by

$$\theta = \lambda(1 - p_B) = \lambda(1 - p_L). \quad (4.9)$$

Since the channel is shared by N users with the same statistical behavior, the channel throughput is therefore obtained as

$$\Theta = N\theta = N\lambda(1 - p_L). \quad (4.10)$$

Another way to obtain θ is given by

$$\theta = \mu p_b, \quad (4.11)$$

which is explained as the mean packet service rate timed by the percentage time when the tagged user is busy.

4.3.1.3 Queue Length

The probability distribution function of the queue length l or the number of packets in the tagged user and therefore each of the other users, is already given in (4.4) and (4.5). The mean queue length in the tagged user is given by

$$E(l) = \sum_{k=0}^L k p_k. \quad (4.12)$$

4.3.1.4 Waiting Time and Response Time

Using (B.7), the PGF $W(z)$ of the packet waiting time w is given by

$$W(z) = \frac{P_0}{1 - p_L} + \sum_{i=1}^{L-1} \frac{P_i}{1 - p_L} [\mu B^{i-1}(z) + (1 - \mu) B^i(z)]. \quad (4.13)$$

Substituting (4.5) and (4.7) into (4.13), we have

$$W(z) = \frac{P_0}{1 - p_L} \left[1 + \frac{\lambda}{1 - \lambda} \cdot \frac{1}{1 - (1 - \mu)z} \cdot \frac{1 - [xB(z)]^{L-1}}{1 - xB(z)} \right], \quad (4.14)$$

where

$$x = \frac{\lambda(1-\mu)}{\mu(1-\lambda)}. \quad (4.15)$$

As to the mean waiting time, it can be obtained using (4.14) by

$$E(w) = W'(1). \quad (4.16)$$

However, it seems to be quite complicated to compute (4.16). The alternate choice is to use Little's law to compute the mean system response time, which minus the mean service time $1/\mu$ gives the mean waiting time. That is

$$E(w) = \frac{E(l_q)}{\lambda(1-p_B)} - \frac{1}{\mu}. \quad (4.17)$$

Substituting (4.14) into (B.9), the PGF of packet response time or system time is given by

$$R(z) = \frac{p_0 B(z)}{1-p_L} \left[1 + \frac{\lambda}{1-\lambda} \cdot \frac{1}{1-(1-\mu)z} \cdot \frac{1-[xB(z)]^{L-1}}{1-xB(z)} \right]. \quad (4.18)$$

Using Little's law, the average response time is given by

$$E(r) = \frac{E(l_q)}{\lambda(1-p_B)}. \quad (4.19)$$

4.3.1.5 Output Process of The Channel

The channel output process can be developed using the results in Section 3.4.4.2. Here we have a homogeneous system with

$$f_1(z) = N p p_b p_s z, \quad (4.20)$$

and

$$f_2(z) = (1 - N p p_b p_s) z. \quad (4.21)$$

Therefore, the channel output process is given by

$$O_{ch}(z) = \frac{f_1(z)}{1 - f_2(z)} = \frac{Npp_b p_s z}{1 - (1 - Npp_b p_s)z}. \quad (4.22)$$

That is, the channel output process is a Bernoulli process.

4.3.2 S-ALOHA with Infinite Buffer Capacity

As discussed before, each user in a S-ALOHA system with infinite buffer capacity can be considered as a Geo/Geo/1 system, which may be treated as a special case of Geo/Geo/1/K system where $K = \infty$. However, in order to stabilize a Geo/Geo/1 queue, we must have $\mu > \lambda$, which obviously is also the sufficient and necessary conditions for the stability of S-ALOHA system.

Definition 4.1 Stability of Queueing System: A queueing system is unstable if and only if the buffer size increases toward ∞ with nonzero probability. Otherwise, it is stable.

Clearly, if a user is unstable in the MA system, all users in the group which the user belongs to are unstable since it is assumed in TUA that all users in a group have statistically identical behaviors.

Definition 4.2 Stability of MA Systems: A multiple access system is said to be unstable if a group of users is unstable.

Using the results in Section 4.3.1 or Section B.4, we have

$$p_0 = 1 - \rho, \quad (4.23)$$

and

$$p_i = \left[\frac{\lambda(1-\mu)}{\mu(1-\lambda)} \right]^{i-1} \frac{\lambda}{\mu(1-\lambda)} p_0, \quad i = 1, 2, \dots, \infty, \quad (4.24)$$

where μ is given in (4.6) and

$$\rho = \frac{\lambda}{\mu}. \quad (4.25)$$

Noting that μ is given in (4.6) in both finite buffer and infinite buffer cases. This is true since a user would contend the channel with probability p as long as it is busy, no matter what its buffer capacity is.

Solving (4.6) and (4.23), the values of p_0 and μ are obtained. Hence, all the other interested performance measurements can be obtained using Section B.4.

It is shown in (B.18) that the output process of users depends only on λ . This is true only if the system is stable or $\rho < 1$. However, the system stability is determined by all the system parameters. If the system is not stable, (B.18) does not hold.

Equation (4.22) holds also for infinite buffer system. In this case, we have $pp_b p_s = \mu \rho = \lambda$. Therefore, the output process is given by

$$O_{ch}(z) = \frac{N\lambda z}{1 - (1 - N\lambda)z}. \quad (4.26)$$

In continuous time system, it is well known that the total packet arrival to a system from one or more sources follows a Poisson process if each of the source is Poisson. Equation (4.26) implies the similar property for discrete time system. Though what is developed in Section 3.4.4.2 is an approximate method, it is believed that (4.26) is accurate, intuitively.

4.3.3 Stability Analysis of S-ALOHA with Infinite Buffer Capacity

By the definition above, stability analysis makes sense only for infinite buffer systems. In this section, we study the stability of the S-ALOHA system with infinite buffer capacity, and the influence of all parameters on the system performance.

Substituting (4.6) and (4.25) into (4.23), we have

$$p_0 = 1 - \frac{\lambda}{p[1 - (1 - p_0)p]^{N-1}}, \quad (4.27)$$

which implies that there are total N solutions for p_0 . However, only the solutions with $p_0 \in [0, 1]$ make sense. Let

$$y = p(1 - p_0), \quad (4.28)$$

we must also have $y \in [0, 1]$.

Using (4.28), (4.27) can be rewritten as

$$\lambda = y(1 - y)^{N-1}. \quad (4.29)$$

Taking $d\lambda/dy = 0$, it is found that λ reaches its maximums only at points $y = 1$ and $y = 1/N$. By (4.28) point $y = 1$ implies $p_0 = 0$ and $p = 1$ which lead to $\lambda = 0$ by (4.29) and mean that all users are always busy, but no packet would be transmitted successfully in any slot. Clearly, this is the worst case for a communication system. Point $y = 1/N$ is located between 0 and 1 and therefore is the only reasonable maximum. Hence, the maximum user throughput is given by

$$\theta_{max} = \lambda_{max} = \frac{1}{N} \left(1 - \frac{1}{N}\right)^{N-1}. \quad (4.30)$$

As the result, the maximum channel throughput or channel capacity Θ_{max} is obtained as:

$$\begin{aligned} \Theta_{max} &= N\theta_{max} = \left(1 - \frac{1}{N}\right)^{N-1}, \\ \lim_{N \rightarrow \infty} \Theta_{max} &= e^{-1} \cong 0.3679. \end{aligned} \quad (4.31)$$

The results of (4.31) are well known. When system operates at point $y = 1/N$, using (4.28) we have

$$p_0 = 1 - \frac{1}{Np}. \quad (4.32)$$

In order to keep $p_0 \in [0, 1]$, we must have $p \in [1/N, 1]$. It seems that the larger the permission probability p the better since large p leads to large p_0 and short waiting time according to (4.32). However, it is not true. As discussed below, a system operating at point $p = 1$ is unstable.

Based on Definition 4.1 and Definition 4.2, there is one-to-one correspondence between user stability and S-ALOHA system stability. Hence, we focus on the stability of the tagged user from now on. Throughput curve and load curve are employed to carry on the stability analysis. The throughput curve describes the packet output rate which is given by

$$S_{out} = p_b \mu = p(1 - p_0)[1 - p(1 - p_0)]^{N-1}, \quad (4.33)$$

since the packet service rate is μ if the tagged user is busy. The load line describes the packet arrival rate which is given by

$$S_{load} = \lambda. \quad (4.34)$$

For infinite buffer system, it is stable if and only if

$$S_{out} = S_{load}. \quad (4.35)$$

Noting (4.35) gives the relation in (4.29) again.

The stability analysis here is carried out using the idea similar in [12]. The fluid approximation for the trajectory of the pair (p_0, S_{out}) is employed. Consider a case where p is fixed while $\lambda = \lambda(t)$ is time-varying as shown in Figure 4.2. This corresponds to a system

Figure 4.2. Time-varying packet arrival rate $\lambda(t)$

which has arrival rate λ_1 normally but experiences input disturbance at time t_1 . The disturbance is over finally at time t_3 and the input goes back to normal. The solid curve without arrow in Figure 4.3 is obtained using (4.33), which describes the relation between the user throughput and its idle probability p_0 . This curve is called throughput line here. The set of dotted lines parallel to each other in Figure 4.3 is obtained using (4.34), called load lines. Each of them corresponds to a packet arrival rate of the tagged user. The throughput line divides Figure 4.3 into two regions. The one above the throughput line is called Region 1, and the other Region 2. In Region 1, idle probability p_0 would decrease since packet input rate is greater than output rate for given p_0 . Therefore, the fluid flow direction

Figure 4.3. Fluid Approximation Trajectories for $N = 100$ and $p = 0.025$

for each load line in Region 1 is from right to left, as indicated by the arrows. Similarly, the direction in Region 2 is from left to right.

The operating points or system equilibrium points are those where the load line and throughput line intersect. A system may have zero, one or two equilibrium points for given N and p , depending on the arrival rate λ . An infinite buffer system with zero, one or two equilibrium points is said to be unstable, globally stable or locally stable, respectively.

When a system has two equilibrium points such as A_1 and A_2 , only the one corresponding to the larger p_0 such as A_1 is locally stable. The other such as A_2 is unstable. Two possible cases are shown corresponding to different values λ_3 of the input pulse in Figure 4.3. The solid line (Case 1) represents a trajectory which returns to the original state A_1 on the equilibrium contour despite the input pulse. The dashed line (Case 2) represents

a less fortunate situation in which the decrease in the packet arrival rate at time t_2 is not sufficient to bring the trajectory back into Region 2. As the result, the system becomes unstable. When a system has only one equilibrium point such as A_3 , the system is always stable since the throughput is greater than arrival rate at $p_0 = 0$. No matter how large the queue size is built during the period of input disturbance, the queue size decreases and goes back to normal finally after the disturbance is over. A_3 is the globally stable equilibrium point of the system.

As indicated in (4.33), the throughput line is entirely dependent on transmission probability p with N given. The throughput lines for p equal to 1, 0.2, 0.05, 0.02, 0.0136, 0.01, 0.0071, and 0.003 are given in Figure 4.4. It is obvious that there is more than one value

Figure 4.4. Throughput lines for different transmission probabilities given $N = 100$

which p may take to stabilize the system if $\lambda < \lambda_{max}$. The system with $\lambda > \lambda_{max}$ is always unstable.

If $\lambda = \lambda_{max}$, the system is unstable for $p < 0.01 = 1/N$. The system has two equilibrium points overlapped for $p > 0.01 = 1/N$, one stable and the other unstable. The system is also unstable at $p = 0.01 = 1/N$ since we have $p_0 = 0$ or $\rho = 1$. Therefore, the system is always unstable if $\lambda = \lambda_{max}$.

Now, we focus on the system with $\lambda < \lambda_{max}$. We proceed to find the region in which p may take values in order to stabilize and optimize the system performance.

It is found that the performance of system with $p < 1/N$ is worse than that with $p = 1/N$ since for the same throughput the former has smaller idle probability or longer response time. In addition, the latter may be stable for some load for which the former is unstable. Therefore, for a good system design, we must have $p \geq 1/N$.

Now, we try to find the upper bound of p . It is well known that too large p would result in frequent collisions and therefore degrade the performance of the system. Examples are the throughput lines for $p \geq 0.2$ in Figure 4.4. For any load, there are two equilibrium points. When the system is operating at the locally stable point, the performance is very wonderful because of the large idle probability and small response time. Unfortunately, the system would leave this point and move to the unstable point due to load disturbance and finally the system collapses. It has been proved that any systems with two equilibrium points are unstable. For now on, we focus only on the system with a globally stable point.

The method to determine the bounds of p is best described through an example using Figure 4.5 given by (4.28) and (4.33). Note user throughput decreases monotonically with

Figure 4.5. User throughput versus $y = p(1 - p_0)$

y for $y > 0.01 = 1/N$. It shows that there are two values of y called y_1 and y_2 corresponding to any λ . These two values divide the y axis into three regions. Using (4.28), we have

$$p_0 = 1 - \frac{y}{p}, \quad (4.36)$$

which implies that we must have $y \leq p$ in order that p_0 makes sense. Let $p_{01} = 1 - y_1/p$ and $p_{02} = 1 - y_2/p$. If $p \leq y_1 < y_2$, then $p_{02} < p_{01} \leq 0$. That is, the system has no equilibrium point. The system is therefore unstable. If $y_1 < p < y_2$, then we have $p_{02} < 0 < p_{01}$. Therefore, the system has only one equilibrium point given by $p_{01} = 1 - y_1/p$ which is

globally stable. If $y_1 < y_2 \leq p$, we have $0 \leq p_{02} < p_{01}$. This is corresponding to a system with two equilibrium points, one unstable and the other locally stable.

Hence, in order that the system is stable, we must have

$$y_1 < p < y_2. \quad (4.37)$$

Note the maximum point of the curve in Figure 4.5 corresponds to $y = 1/N$, and we always have $y_1 < 1/N < y_2$ if $\lambda < \lambda_{max}$. It is already shown that p is lower bounded by $1/N$ for best performance. Combine the two results together, we conclude that in order that the system is stable and has good performance, we must have

$$1/N \leq p < y_2. \quad (4.38)$$

For instance, it is found that $y_1 = 0.00705$ and $y_2 = 0.01366$ for a system with $\lambda = 3.5 \times 10^{-3}$ and $N = 100$. In this case, we must have $0.01 \leq p < 0.01366$ for best system performance.

How to select a permission probability p is determined by the performance requirements. For instance, both $p = 0.01$ and $p = 0.0136$ satisfy (4.38) given $\lambda = 3.4 \times 10^{-3}$ in Figure 4.4. By (4.36), the larger p leads to the larger p_0 and therefore the smaller response time. However, the former corresponds to a system with better stability margin since the system has higher throughput at $p_0 = 0$, at which the system operates if there is a sufficiently long positive load disturbance such that the system is always busy during a period. When the disturbance disappears, the system with $p = 0.01$ would move back to

its equilibrium point faster than that with $p = 0.0136$. Hence, there is a trade-off in selecting p for the consideration of system stability margin and packet response time.

4.3.4 Equilibrium Point of S-ALOHA with Finite Buffer Capacity

There is no stability problem for a queueing system with finite queue capacity. However, if the retransmission probability p is not selected properly, the system would have very poor performance. In this section, we analyze the equilibrium points of the finite buffer system.

The throughput line is again given by (4.33), but the load line is given by

$$S_{load} = \lambda(1 - p_B). \quad (4.39)$$

The fluid approximation trajectories for S-ALOHA with finite buffer capacity for two different cases are given in Figure 4.6 and Figure 4.7, where solid lines are the throughput

Figure 4.6. Fluid approximation trajectories of finite buffer system for $p = 0.008 \leq 1/N$, $L = 5$ and $N = 100$

lines and dashed lines are the load lines.

In Figure 4.6 a throughput line for $p \leq 1/N$ is given. It shows that the system has only one globally stable equilibrium point for any λ , whose value can be read from the corresponding load line at $p_0 = 1$.

The throughput line in Figure 4.7 corresponds to a system with $p > 1/N$. In this case, the system may have one or three equilibrium points. In the case that system has only one equilibrium point, the system is globally stable. If the λ is comparatively small such as in lines 5, 6 and 7, the packet waiting time would be short because of large p_0 . If the λ is

Figure 4.7. Fluid approximation trajectories for $p > 1/N$

comparatively large such as in lines 1, the packet waiting time would be large. When the system has three equilibrium points such as line 3, the two equilibrium points corresponding to smallest p_0 and largest p_0 respectively are locally stable. The former is called the saturated point, and the latter the desired point. The remaining point between them is an unstable point since the system always tends to move away from this point. Lines 2 and 4 are special cases of three equilibrium points, where one of the locally stable point overlaps with the unstable point. In line 2, the system behaves as it has only one globally stable point at D since point C is unstable. Similarly, line 4 behaves as the system has only one globally stable point at B.

Hence, for given p in Figure 4.7, the λ corresponding to line 4 is the upper bound for choosing λ such that the system operates in a globally stable point. For given λ corre-

sponding to line 4, on the other hand, the p corresponding to the throughput line in Figure 4.7 is the upper bound for choosing p such that the system operates at a globally stable point. Again, the lower bound of p is $1/N$ as in Section 4.3.3. For instance, in order that a system with $N = 100$ and $\lambda = 0.0034$ operates at globally stable point, we must have $0.01 \leq p < 0.016$. Similar to the infinite buffer capacity case, there is a trade-off between stability margin and response time for selecting p .

4.4 Numerical Algorithm

In this section, numerical algorithms are developed for both infinite and finite buffer S-ALOHA systems. As discussed in Section 4.3, p_0 is the key for obtaining all the other performance parameters. In this section, we present iterative algorithms for both finite and infinite buffer S-ALOHA systems.

4.4.1 Finite Buffer S-ALOHA

- 1 set $n = 0$ and select a very small positive number, say $\delta = 1.0e - 6$; set the initial value of p_0 between zero and one, say $p_0^{(n)} = 0.1$;
- 2 calculate $\mu^{(n)}$ using (4.6);
- 3 calculate $p_0^{(n+1)}$ by substituting μ^n into (4.4);
- 4 if $|p_0^{(n+1)} - p_0^{(n)}| / p_0^{(n)} \geq \delta$, let $n = n + 1$ and goto 2;
- 5 let $p_0 = p_0^{(n+1)}$ be the solution.

Numerical results show that this iterative algorithm can find the only equilibrium point if the system is globally stable. However, if the system is bi-stable, the obtained results may converge to one of the two locally stable equilibrium points based on the initial condition of p_0 . If the initial condition is set to $p_0^{(0)} = 0$, which implies a busy system, the results

always converge to the saturated point. If the initial condition is $p_0^{(0)} = 1$, the results always converge to the desired point. This property is observed also in simulation.

4.4.2 Infinite Buffered S-ALOHA

The algorithm for finding p_0 in infinite buffer case is same as the one for finite buffer system except that (4.25) instead of (4.4) is employed in Step 3. Similar to finite buffer system, the obtained p_0 corresponds to the only equilibrium point if the system is globally stable, and the locally stable equilibrium point if the system has two equilibrium points.

4.5 Numerical Results

In this section, we apply TUA to analyze S-ALOHA systems. Firstly, we analyze a system with infinite buffer capacity. Then we will deal with a system with finite buffer capacity.

4.5.1 System with Infinite Buffer Capacity

In this section, we focus on the performance of a S-ALOHA system with user population $N = 100$ and different packet arrival rate. For a system with infinite buffer capacity, the channel throughput is not necessary to be given since it is equal to total packet arrival rate in the system. Here we are interested in the mean packet response time, which is given in Figure 4.8. It is noticed that all the lines in the diagram do not cover the axis corresponding to very small p values. This is because the system is not stable in this region as discussed above. The region of p such that the system has only one globally stable equilibrium

Figure 4.8. Mean packet response time versus transmission probability

point is given by (γ_1, γ_2) in Table 4.1. Note the mean queue length of the system is not given since it is proportional to the mean packet response time by Little's law.

Table 4.1. Stable Region of transmission probability p

λ	γ_1	$1/N$	γ_2
2.00E-3	2.5838E-3	1.0E-2	2.5312E-2
3.00E-3	4.8589E-3	1.0E-2	1.7856E-2
3.50E-3	7.0537E-3	1.0E-2	1.3662E-2
3.60E-3	7.8732E-3	1.0E-2	1.2475E-2
3.69E-3	9.3874E-3	1.0E-2	1.0638E-2

Though the performance looks like better for larger p from Figure 4.8, it is not true since p with value exceeding the upper bound of its global stability region would lead to a system with two equilibrium points, which in fact is unstable. That is, we should always keep in mind Table 4.1 when evaluating the system performance using Figure 4.8. In

Table 4.1, (y_1, y_2) given in (4.37) is the region of p for system stability. However, for best performance, the region of p should be given by $(1/N, y_2)$.

4.5.2 System with Finite Buffer Capacity

We consider a system with $N = 100$ and $\lambda = 0.0035$. The equilibrium points of the system with respect to different transmission probability p and the buffer capacity of 30,

Figure 4.9. User idle probabilities at equilibrium points versus transmission probability

15, and 5 packets are illustrated in Figure 4.9. It shows that the number of system equilibrium points changes from one to three when p increases to around 0.015, 0.0164 and 0.022, respectively. For instance, if we draw a vertical line at $p = 0.02$, it will intersect with the two solid lines at three points. It means that the system has three equilibrium points at

$p = 0.02$ for $L = 30$. When the system has three equilibrium points, those on the top and at bottom are locally stable, and the point in the middle is unstable. It is noticed that the two locally stable points never overlap though the unstable point and the desired point would overlap for some transmission probability. This explains why the lines in Figure 4.9- Figure 4.12 belonging to the same system do not intersect with each other.

The throughput of the systems corresponding to the three cases is given in Figure 4.10,

Figure 4.10. Channel throughput versus transmission probability

where the throughput for each case has one or two values, corresponding to the number of stable equilibrium points. For instance, the system with $L = 30$ has throughputs 0.35 and 0.28 at $p = 0.02$, corresponding to the two local stable equilibrium points $p_0 = 0.65$ and $p_0 = 10^{-4}$, respectively, as shown in Figure 4.9.

The mean packet response time and buffer length are shown in Figure 4.11 and

Figure 4.11. Mean response time versus transmission probability

Figure 4.12. When system operates at the saturated point, the buffer becomes full quickly with increasing p , and packet response time increases without limitation.

4.5.3 Accuracy of TUA on The Analysis of S-ALOHA

So far, a complete analysis on S-ALOHA systems using TUA is presented and the numerical results are given. A problem is how to evaluate the accuracy of TUA. Unfortunately, there is no theoretical analysis available regarding the accuracy of TUA. So far, the only way is to verify the accuracy of TUA by simulation.

In this section we consider a system model in Section 2.4.2 of [1] with the following parameters: $N = 100$, $\lambda = 0.0035$ and buffer capacity of 1, 3 and 50 packets. The numerical results are given in Figure 4.13 and Figure 4.14, where the lines and markers corre-

Figure 4.12. Mean buffer length versus transmission probability

Figure 4.13. Channel throughput versus permission probability

Figure 4.14. Mean packet response time versus transmission probability

spend to the results obtained using TUA and simulation contributed by [1] respectively. When the system is bi-stable, the dotted lines correspond to the desired points. It shows that the simulation and analysis match very well if the system is globally stable. However, there exists a large error between simulation and analysis if the system is bi-stable. This is because a bi-stable system would oscillate between the two locally stable points because of traffic load disturbance. If for some p the system would stay at the saturated point with probability α , and the channel throughputs corresponding to the saturated point and the desired point are Θ_s and Θ_d respectively, the practical system throughput should be given by

$$\Theta = \alpha\Theta_s + (1 - \alpha)\Theta_d. \quad (4.40)$$

Unfortunately, we cannot calculate α using TUA. Hence, TUA cannot give an accurate performance evaluation for bi-stable system. Fortunately, usually a system is expected to operate at its globally stable point. In this case, TUA gives satisfied performance evaluation. Note we always have $\Theta_s < \Theta < \Theta_d$ because of $\Theta_s < \Theta_d$, which explains why the simulation results in bi-stable system locate between those corresponding to the two equilibrium points.

Finally, it should be mentioned that the analytical results are obtained using the numerical algorithm developed in Section 4.4.1. When the system is globally stable, the results are independent of the initial condition of p_0 . However, the results corresponding to the desired point and saturated point are obtained using initial condition $p_0 = 1$ and $p_0 = 0$ respectively. In Figure 4.13 and Figure 4.14, it is found that two lines corresponding to the two locally stable points intersect at point A (solid line and dotted line) and B (dashed line and dotted line), where the solid line or dashed line corresponds to saturated points, and dotted lines correspond to the desired points. This is not true as discussed in previous sections. Instead, points A and B are the places where the unstable point and saturated overlap. The reason is that when we draw the figures, the results with the transmission probabilities corresponding to one globally stable point and the saturated point are connected by lines automatically for convenience. From now on, all the figures look like that the results corresponding two locally stable points would diverge from some points, like points A and B in Figure 4.13 and Figure 4.14, for convenience. Please keep in mind that the results corresponding to the two locally stable points never overlap.

4.6 Summary

In this chapter, TUA is employed to analyze S-ALOHA systems. In addition to mean values, the distributions of the important performance indices such as packet response time and interdeparture time are obtained. The stability of infinite buffer system is studied, and the region of transmission probability p for system global stability and best performance is identified. In addition, a complete analysis of the equilibrium points is presented, which would assist us to choose the proper transmission probability so that the system always operates at a globally stable equilibrium point. Finally, numerical results are given.

It is shown that the distribution of the packet service time of each user in S-ALOHA system with DFT model is geometric. Therefore, the performance analysis is carried out employing the very simple Geo/Geo/1 or Geo/Geo/1/K queueing theory. However, if the service time distribution is not geometric, more complicated queueing analysis such as Geo/G/1 or Geo/G/1/K has to be employed.

As mentioned in Section 4.1, a general approximate analysis approach for buffered ALOHA or CSMA/CD protocols employs a two (or higher) dimensional state space to describe the system, where one is related to the number of packets in the tagged user and the other to the number of busy user in the system. In TUA, the system state space is one dimensional, where the number of states is just the number of packets in the buffer. Although there is no separate state variable to describe the number of busy users in the system, the information is considered to be carried by user idle probability p_0 . This implies that the number of busy users is treated to be binomially distributed. This assumption is jus-

tified for large user population and high busy probability due to the very good accuracy of TUA analysis.

The simulation indicates that the accuracy of TUA is excellent for S-ALOHA systems as long as the system has only one globally stable equilibrium point. However, if the system is bi-stable, only the lower and upper bounds of system performance can be obtained, where the former corresponds to the saturated point, and the latter to the desired point.

5 CDMA S-ALOHA AND SATELLITE S-ALOHA SYSTEMS

In Chapter 4, a complete analysis on S-ALOHA systems using TUA is presented, where the system is assumed to operate at a channel environment without noise and propagation delay. In addition, it is assumed that packet arrival in each user follow a Bernoulli arrival process, and packet transmission follow DFT scheme. In this chapter, a CDMA S-ALOHA system which permits bulk packet arrival is analyzed using TUA firstly [93]. The corresponding queueing analysis on $Geo^{[x]}/Geo/1/K$ is developed. It shows that channel condition and user queueing behavior do not introduce extra difficulty to TUA as long as the corresponding analysis on bit error probability and queueing behavior are available. A S-ALOHA system operating in satellite communications environment, where long propagation delay exists, is then analyzed. In this case, we assume that the new packet and back-off packet be assigned with different transmission probabilities. It is found that the packet service time is no longer geometrically distributed. Hence, the queueing analysis on $Geo/G/1/K$ system is required.

5.1 CDMA S-ALOHA Systems

Recently, CDMA ALOHA draws much attention for satellite and mobile communications because it combines the advantages of both CDMA and ALOHA. In conventional

ALOHA systems, if more than one packet is transmitted at a same slot, the packet successful probability is quite low though a packet with highest power measured at the base station may capture the channel. However, in CDMA S-ALOHA systems, more than one packet may be received successfully at the base station when they are transmitted simultaneously. The unsuccessful probability increases gradually with the increase of the number of simultaneous packet transmission. A buffered CDMA S-ALOHA system has been analyzed in [88] using linear approximate solution based on restricted occupancy urn models. In this chapter, we will analyze a buffered S-ALOHA system operating in the same CDMA channel environment as in [88] using TUA. However, bulk packet arrivals are permitted here.

We consider a system where data is generated by each user in messages. Each message contains one or more packets. Each user contends for the channel to transmit its packets using CDMA S-ALOHA protocol, one packet at a time. That is, each user represents a bulk arrival system. For a bulk arrival queueing system, we have to specify how an arriving message is treated if the buffer does not have enough available space to accept a whole arriving message. One treatment is that part of message is accepted into the buffer till the buffer is full and the rest of message is rejected. The other is that the whole message is rejected. The former is called partial rejection model and the latter total rejection model [78]. The total rejection model is employed here since it makes more sense for systems with multi-packet messages. A queueing methodology for the analysis of such systems is also developed.

5.1.1 System Model

A single cell homogeneous CDMA S-ALOHA system with N users is considered. Each user contends for the uplink channel to transmit a packet to the base station using CDMA

S-ALOHA. Each user has a buffer capacity of L packets. Time is divided into slots of length normalized to 1, which is also the packet transmission time. Data arrival follows a Bernoulli process such that a message arrives in a slot with probability λ . Each message contains a random number of packets with PGF given by

$$G(z) = \sum_{i=1}^{\infty} g_i z^i, \quad (5.1)$$

that is, the probability that a message contains i packets is g_i . All packet transmissions start and end at the slot boundary. The transmission acknowledgment is immediate. A successful packet departs immediately, and an unsuccessful packet will be retransmitted during the next slot with probability p . When there is a message arrival, the message would be accepted into the buffer if the buffer has enough space to hold the whole message. Otherwise, the whole message is rejected. Fresh and blocked packets will always be (re)transmitted with probability p , this is known as DFT. The late arrival model with delayed replacement rule is used for accepting arriving packets into the buffer.

The same channel environment as in [88] is employed here. That is, the modulation scheme of the system is BPSK (binary phase-shift keying), and signals are band-limited by a root cosine roll-off filter. The transmitter and receiver are given in Figure 2 of [88]. Perfect power control is assumed and unsuccessful transmissions are due to the MAI (multiple access interference) and the AWGN (additive white Gaussian noise). As the result, when there are k users transmitting packets simultaneously, the BER (bit error rate) of any one of the users is given by [88]

$$P_{be}(k) = Q\left(\left[\frac{k-1}{2N_c}\left(1 - \frac{\alpha}{4}\right) + \frac{N_0}{2E_b}\right]^{\frac{1}{2}}\right), \quad (5.2)$$

where N_c is the number of chips per bit, E_b is the bit energy, $N_0/2$ is two-sided spectral density of AWGN, α is the roll-off factor, and

$$Q(x) = \int_x^\infty \exp(-u^2/2) du. \quad (5.3)$$

5.1.2 Contention Scheme Analysis

Comparing the system model with the model in Chapter 4, it is found that the state flow graph of the tagged user in Figure 4.1 is applicable in the current model except that packet arrival process is different. Hence, the packet service time distribution is given by

$$B(z) = \frac{pp_s z}{1 - (1 - pp_s)z}, \quad (5.4)$$

which implies that each user in this CDMA S-ALOHA system can be modeled by a queueing system with geometrically distributed service time, Bernoulli bulk arrival process with total rejection model, finite buffer capacity and single server FIFO system. This queueing system is denoted by $Geo^{[x]}/Geo/1/K$ following a total rejection model [78].

In Chapter 4, packet collision is assumed to always lead to transmission failure. However, this assumption is not applicable in CDMA environment. To consider the influence of other users on the tagged user, we need to know the busy probability p_b of all the other users, which due to channel symmetry is equal to the busy probability of the tagged user.

When the tagged user transmits a packet in a slot, each of other $N - 1$ users would transmit a packet in the same slot with probability pp_b . Hence, p_s is given by

$$p_s = \sum_{i=0}^{N-1} \binom{N-1}{i} (pp_b)^i (1 - pp_b)^{N-1-i} Q_s(i+1), \quad (5.5)$$

where $Q_s(k)$ is the packet successful probability of the tagged user given k simultaneous transmission. It is given by

$$Q_s(k) = [1 - P_{be}(k)]^{L_b}, \quad (5.6)$$

where L_b is the number of bits contained in each packet and $P_{be}(k)$ is given by (5.2).

It should be mentioned that Eqn. (5.6) may not be a proper formula for wireless channel where error control coding is employed. However, TUA is applicable in non-ideal channel as long as $Q_s(k)$ is known as indicated by (5.5). It does not care how $Q_s(k)$ is obtained.

5.1.3 Queueing Analysis

We consider the $Geo^{[x]}/Geo/1/K$ system with arrival rate λ and service rate μ . The PGF of the number of packets contained in a message is given by (5.1). Let p_i , $i = 0, 1, \dots, L$, be the probability that there are i packets in a user buffer at steady state. Clearly we have

$$p_b = 1 - p_0. \quad (5.7)$$

By writing the global balance equations for the stationary probability distribution and keeping in mind that a message containing x packets can be accepted into the user buffer only if there are no more than $L - x$ packets in the buffer upon this message arrival, we have

$$\lambda p_0 \sum_{i=1}^L g_i = \mu \left(1 - \lambda \sum_{i=1}^{L-1} g_i \right) p_1, \quad (5.8)$$

$$\begin{aligned} & \left[\mu \left(1 - \lambda \sum_{i=1}^{L-k} g_i \right) + (1-\mu) \lambda \sum_{i=1}^{L-k} g_i + \mu \lambda \sum_{i=1}^{L-k-1} g_{i+1} \right] p_k \\ &= \mu \left(1 - \lambda \sum_{i=1}^{L-k-1} g_i \right) p_{k+1} + \sum_{i=1}^{k-1} [(1-\mu) \lambda g_{k-i} + \mu \lambda g_{k-i+1}] p_i \\ &+ \lambda g_k p_0, \quad 1 \leq k \leq L-2, \end{aligned} \quad (5.9)$$

and

$$\begin{aligned} & [\mu(1 - \lambda g_1) + (1-\mu) \lambda g_1] p_{L-1} \\ &= \left(\mu p_L + \sum_{i=1}^{L-2} [(1-\mu) \lambda g_{L-i-1} + \mu \lambda g_{L-i}] p_i + \lambda g_{L-1} p_0 \right), \end{aligned} \quad (5.10)$$

where (5.8) and (5.10) are obtained by analyzing the probability flows at states 0 and $L-1$, and (5.9) is obtained by analyzing the probability flows at states $1 \leq k \leq L-2$. In each of the above three equations, the LHS and RHS are the probabilities flowing out and into the states 0, 1 to $L-2$, and $L-1$, respectively. A user is said to be at state i if its buffer contains i packets. An equation can also be obtained by analyzing the probability flow at state L , which is a linear combination of equations (5.8)-(5.10) and not shown. Equations (5.8)-(5.10) along with total probability law

$$\sum_{i=0}^L p_i = 1 \quad (5.11)$$

are sufficient to solve the queueing system.

5.1.4 Performance Analysis

The concerned performance parameters are channel throughput, the message and packet blocking probabilities, denoted by p_{mblock} and p_{pblock} respectively, and the mean message and packet response time. The channel throughput is defined as the average number of packets transmitted successfully in each slot. Probabilities p_{mblock} and p_{pblock} are defined as the percentage of arriving messages and packets which are rejected, respectively. The mean message and packet response time are defined as the time duration in slots from their arrival to their successful departure.

In the concerned CDMA S-ALOHA system, by (5.4) each user is a $Geo^{[x]}/Geo/1/K$ queueing system with service rate

$$\mu = p p_s. \quad (5.12)$$

Probability p_{mblock} can be derived in two ways. One approach is to consider each state and observe that only arriving messages with size no greater than $L - i$ can be accepted into the buffer at state i . As the result, we have

$$p_{mblock} = p_L + \sum_{i=0}^{L-1} p_i \left(1 - \sum_{k=1}^{L-i} g_k \right). \quad (5.13)$$

An alternate approach is to consider an arriving message with any possible size and find the blocking probability of the message. As the result, we have

$$p_{mblock} = \sum_{i=1}^L g_i \sum_{k=L-i+1}^L p_k + \sum_{i=L+1}^{\infty} g_i. \quad (5.14)$$

In fact, equations (5.13) and (5.14) can be derived from each other using (5.11) and the total

probability law $\sum_{i=0}^{\infty} g_i = 1$ by (5.1).

We now consider an arriving message with an arbitrary size. It is observed that any message consisting of more than L packets would be rejected with probability 1, and any message consisting of i , $1 \leq i \leq L$, packets would be rejected if it sees more than $L - i$

packets in the buffer upon arrival, which occurs with probability $\sum_{k=L-i+1}^L p_k$. Hence, we

have

$$E(n, A) = \sum_{i=L+1}^{\infty} i g_i + \sum_{i=1}^L i g_i \sum_{k=L-i+1}^L p_k, \quad (5.15)$$

where $A = \{\text{an arriving message is rejected}\}$, and n is the number of packets contained in an arriving message, a random variable. Consequently, the mean number of packets in a blocked message, denoted by n_{block} , is given by

$$\begin{aligned} n_{block} &= E(n|A) \\ &= E(n, A) / Pr(A) \\ &= \frac{1}{p_{mblock}} \left[\sum_{i=L+1}^{\infty} i g_i + \sum_{i=1}^L i g_i \sum_{k=L-i+1}^L p_k \right] \\ &= \frac{1}{p_{mblock}} \left[G'(1) - \sum_{i=1}^L i g_i + \sum_{i=1}^L i g_i \sum_{k=L-i+1}^L p_k \right], \end{aligned} \quad (5.16)$$

where $G'(1)$ by (5.1) is the mean number of packets contained in a message. Furthermore, the average number of rejected packets is $p_{mblock} \cdot n_{block}$ in each message arrival. Finally,

each message arrival implies that there is $G'(1)$ number of packet arrivals in average. The packet blocking probability is therefore given by

$$p_{pblock} = p_{mblock} \cdot n_{block} / G'(1). \quad (5.17)$$

By BASTA (Bernoulli arrivals see time averages) [79], an arriving message would see k packets in the buffer with probability p_k . Since a message containing i packets would be

accepted into the buffer with probability $\sum_{j=0}^{L-i} p_j$, an accepted message containing i packets

would see k , $k \leq L - i$, packets in the buffer upon arrival with probability $p_k \left(\sum_{j=0}^{L-i} p_j \right)$.

The mean waiting time of an accepted message seeing k , $1 \leq k \leq L - i$ packets in the buffer is equal to $[\mu(k-1) + (1-\mu)k] / \mu = k/\mu - 1$ since the packet being served would depart from the queue immediately with probability μ in each slot. Hence, the mean waiting time is the average service time of $k-1$ or k packets waiting in the buffer with probability μ or $1-\mu$, respectively.

Denoting by $\bar{w}_{mi}$ the mean waiting time of messages containing i packets, we have

$$\bar{w}_{mi} = \sum_{k=1}^{L-i} p_k \left(\frac{k}{\mu} - 1 \right) \left(\sum_{j=0}^{L-i} p_j \right). \quad (5.18)$$

Therefore, the mean message waiting time $\bar{W}_m$ is given by

$$\bar{W}_m = \sum_{i=1}^L \bar{w}_{mi} g_i, \quad (5.19)$$

and the mean message response time $\bar{R}_m$ is given by

$$\bar{R}_m = \sum_{i=1}^L \left(\bar{w}_{mi} + \frac{i}{\mu} \right) g_i. \quad (5.20)$$

As to the mean packet response time $\bar{R}_p$, by Little's law it is given by

$$\bar{R}_p = \frac{1}{\lambda G'(1)[1 - p_{pblock}]} \sum_{k=1}^L k p_k. \quad (5.21)$$

Finally, we calculate the channel throughput. The throughput of each user is just the mean number of packets accepted into the buffer in each slot, that is, $\lambda G'(1)[1 - p_{pblock}]$.

Hence, the channel throughput Θ is given by

$$\Theta = N \lambda G'(1)[1 - p_{pblock}]. \quad (5.22)$$

The above equations are sufficient to solve the system and obtain the concerned performance parameters. A numerical algorithm can be developed similarly to Chapter 4 to solve the equations recursively.

5.1.5 Numerical Examples

Two examples are given in this section. In both cases, the following parameters are employed: $N = 65$, $\alpha = 0.5$, $N_c = 63$, $L_b = 500$, $E_b/N_0 = 10$ dB. Buffer capacity L and transmission probability p are variables which are marked in corresponding figures. The offered load here is defined as mean number of arriving packets in a slot, which is given by $N \lambda G'(1)$.

5.1.5.1 Example 1

In this example, we take $G(z) = z$, that is, each arriving message contains only one packet. In this case, each user is modeled by a $Geo/Geo/1/K$ system, a special case of $Geo^{[x]}/Geo/1/K$ system. It is not necessary to distinguish messages from packets.

The buffer capacity is assumed to be $L = 2$. Hence, the system model is same as the one in [88]. The channel throughput, mean packet response time and packet blocking probability are given in Figure 5.1, Figure 5.2 and Figure 5.3, respectively, where the lines are

Figure 5.1. Channel throughput versus offered load of a CDMA S-ALOHA system

analysis results obtained using TUA and the markers are simulation results contributed by [88]. Compared to the simulation results, it is found that the accuracy of TUA is excel-

Figure 5.2. Mean response time versus offered load

Figure 5.3. Packet blocking probability versus offer Load

lent. It is observed that the maximum throughput for sufficiently large offered load is reached by selecting $p = 0.12$.

The purpose of this example is to verify the accuracy of TUA on the analysis of buffered CDMA S-ALOHA system. For discussions on the results, please refer to [88] for details.

5.1.5.2 Example 2

Here we consider a system with $L = 20$ and $G(z) = 0.3z + 0.6z^3 + 0.1z^7$, that is, an arriving message may contain 1, 3 and 7 packets with probability 0.3, 0.6, and 0.1, respectively. Hence, it is necessary to distinguish messages from packets. The results of throughput, response time, and blocking probability are given in Figure 5.4 to Figure 5.8.

Figure 5.4. Channel throughput versus offer Load

As shown in Figure 5.4, the throughput for a fixed p is almost a constant when the offered load is greater than around 7. Similar to Example 1, the maximum throughput is reached at $p = 0.12$ for sufficiently large offered load. In fact, for very heavy offered load, we have $p_b \equiv 1$ and μ is almost a constant for given p by (5.5) and (5.12). Hence, the channel throughput would converge to $N\mu$ no matter a particular arrival process, which, however, would affect the blocking probability and response time of messages and packets, as illustrated in Figure 5.5 to Figure 5.8.

Figure 5.5. Mean packet response time versus offer Load

As shown in Figure 5.5 and Figure 5.6, the mean packet response time is very close to the mean message response time though it is not easy to determine which one should be greater. Since the user buffer capacity is finite, the packet and message response times are

Figure 5.6. Mean message response time versus offer Load

Figure 5.7. Packet blocking probability versus offer Load

Figure 5.8. Message blocking probability vs. offer Load

bounded, unlike infinite buffer capacity case where the response time would increase toward ∞ for sufficiently large offered load.

It is clear that message with larger size has more chance to be rejected upon arrival. Hence, the mean size of a rejected message should be not less than that of an arriving (average) message, that is, $n_{block} \geq G'(1)$. Using (5.17), we have $p_{pblock} \geq p_{mblock}$, which is verified by Figure 5.7 and Figure 5.8.

5.2 Satellite ALOHA System

In this section, a S-ALOHA in satellite communications environment is studied using TUA. An obvious characteristics of satellite channel is its long propagation delay, which would introduce memory problem into the analysis. We also assume that the system employ NFT transmission principle. It should be mentioned that among the available

approaches, only EPA has been used in the analysis of buffered satellite S-ALOHA system successfully. This is because the long propagation delay in satellite communications would introduce extra complexity into the system analysis such that most of the available approaches, which are already complicated enough when dealing with zero propagation delay systems, cannot handle them.

5.2.1 System Model

A S-ALOHA system with N users is considered. Each user has a buffer capacity of L packets. Time is divided into slots of length normalized to 1, which is also the packet transmission time. Data arrival follows Bernoulli process such that a packet arrives in a slot with probability λ . An arriving packet is rejected if the buffer is full. All packet transmissions start and end at the slot boundary. Because of channel propagation delay, the acknowledgment for a packet will be received $D - 1$ slots after it is transmitted. That is, it takes D slots from starting a packet transmission to receiving the corresponding acknowledgment. After acknowledgment is received, a packet is departed immediately if the transmission is successful, or deferred if the transmission fails. That is, stop-and-wait is employed as the automatic repeat request (ARQ) scheme. The system employs NFT transmission principle. New packets and backoff packets will be assigned with transmission probabilities q and p , respectively. The late arrival model with delayed replacement rule is used for accepting arriving packets into the buffer.

5.2.2 Contention Scheme Analysis

The analysis is carried out using the state flow graph of the tagged user as shown in Figure 5.9, where S_0 , S_{bk} , S_{tx} and S_e are packet ready state, backoff state, transmission

Figure 5.9. State flow graph of tagged user for a S-ALOHA system in satellite communications channel

state and departure state, respectively, as defined in Section 4.2. Since a new packet will be transmitted immediately with probability q or deferred one slot with probability $1 - q$, the transfer function from S_0 to S_{tx} is q and from S_0 to S_{bk} is $(1 - q)z$. Since the acknowledgment delay is D slots from the beginning of a packet transmission, the transfer functions from S_{tx} to S_{bk} and S_e are $(1 - p_s)z^D$ and $p_s z^D$, respectively, where p_s is the packet successful probability given it is transmitted. If a transmission is acknowledged as failure in a

slot, the tagged user returns to the backoff state S_{bk} . In fact, both S_0 and S_e are dummy states in which a user does not spend any time.

From Figure 5.9, the packet service time is just the time duration from a user entering S_0 to its exiting S_e . Using Mason's formula [81], the PGF $B(z)$ of packet service time is given by

$$B(z) = \frac{p_s[q + (p - q)z]z^D}{1 - (1 - p)z - p(1 - p_s)z^D}. \quad (5.23)$$

It shows that the packet service time distribution is no longer geometric in general. If and only if we have both $q = p$ and $D = 1$, the service time distribution is geometric as indicated by (5.23), which is in fact the same model as in Chapter 4. Hence, each user in the satellite ALOHA system is modeled by a Geo/G/1 (for infinite queue capacity) or Geo/G/1/K (for finite buffer capacity) system.

In the system model with $q = p$ and $D = 1$, each user can be treated as a Geo/Geo/1 or a Geo/G/1/K system intuitively without going through the state flow graph analysis as in [40], [41] and [47]. However, intuition has its limit. It is hard to get (5.23) intuitively.

Since a user would wait $D - 1$ slots for acknowledgment after a packet transmission, it does not transmit during this period. Hence, a busy user can be in two modes: CON (contending for channel) and WAI (waiting for acknowledgment).

Denoting by p_c the probability that a user is at CON, the transmission successful probability p_s is given by

$$p_s = [1 - (p_c - p_0\lambda)p - p_0\lambda q]^{N-1}, \quad (5.24)$$

where p_0 is user idle probability. This is because only users at CON are able to contend for transmission. Since a user at CON would serve an just arriving packet with probability $p_0\lambda$, it will serve a deferred packet with probability $p_c - p_0\lambda$ approximately, similar to the discussion on (3.5). From (5.23), the mean packet service time is given by

$$b = \frac{p-q}{p} + \frac{D-1}{p_s} + \frac{1}{pp_s}, \quad (5.25)$$

where there are average $(D-1)/p_s$ slots that a user is waiting for acknowledgments. Hence, we have

$$p_c = \frac{p_b}{b} \left(b - \frac{D-1}{p_s} \right). \quad (5.26)$$

where p_b is the probability that a user is busy, which satisfies

$$p_b = 1 - p_0. \quad (5.27)$$

Substituting (5.25) and (5.27) into (5.26), which is then substituted into (5.24), it is found that p_s is an implicit function of p_0 given the parameters p , q , N and D .

5.2.3 Queueing Analysis

Since each user is modeled as a Geo/G/1/K system as indicated by (5.23), the queueing analysis on Geo/G/1/K system in Section B.5 can be introduced here straightforwardly. Eqns. (5.23)-(5.27) and (B.19)-(B.25) are sufficient to solve all the unknown parameters in the system and obtain the interested performance measurements.

5.2.4 Numerical Examples

Firstly we consider a S-ALOHA systems in satellite communications with DFT transmission principle, that is, $q = p$. In addition, the system model has the following parameters: $N = 100$ and $D = 13$. The same model is also employed in [35], whose simulation results would be compared with the analysis results obtained using TUA here.

The channel throughput and mean packet response time versus transmission probability are given in Figure 5.10 and Figure 5.11 given $\lambda = 0.0035$. As expected, the system with

Figure 5.10. Channel throughput versus transmission probability of a satellite S-ALOHA system with $\lambda = 0.0035$

infinite buffer capacity is stable only in a region, as indicated by the solid line between point A and B. This is because the backoff duration before transmission is too long for too small p though transmission successful probability p_s is large, and channel is congested

Figure 5.11. The mean packet response time versus transmission probability of a satellite S-ALOHA system with $\lambda = 0.0035$

due to frequent collisions for too large p . In either case, the packet service rate μ is not greater than packet arrival rate λ . The system is bi-stable beyond point C for $L = 8$.

The channel throughput and mean packet response time for the system with $\lambda = 0.002$ are given in Figure 5.12 and Figure 5.13. Since smaller packet arrival rate is employed here than in the previous system, the system here would have larger region of global stability.

In the two numerical examples, the analysis and simulation results match well. The accuracy of TUA is very satisfied.

In the above two numerical examples for satellite S-ALOHA systems, DFT transmission principle is employed. Here we study a system with NFT transmission principle, called System 2, where a new packet is transmitted with probability $q = 1$. The system parameters are still $N = 100$ and $D = 13$. The results are given in Figure 5.14 to Figure 5.17.

Figure 5.12. Channel throughput versus transmission probability of a satellite S-ALOHA system with $\lambda = 0.002$

Figure 5.13. The mean packet response time versus transmission probability of a satellite S-ALOHA system with $\lambda = 0.002$

Figure 5.14. Channel throughput versus packet transmission probability of System 1 and 2

Figure 5.15. Mean packet response time versus packet transmission probability of System 1 and 2

Figure 5.16. Mean queue length versus packet transmission probability of System 1 and 2

Figure 5.17. User contention probability versus packet transmission probability of System 1 and 2

For comparison, the numerical results of the above system with DFT model, called System 1, are also plotted in the diagrams, where L_1 and L_2 are the buffer capacity of System 1 and 2 respectively. For convenience, only the results corresponding to the saturated point are given in Figure 5.14 to Figure 5.17 when systems are bi-stable. That is, only the worst performance is considered here.

It is shown that the performance of System 2 is always better than that of System 1 based on channel throughput and mean packet response time for given permission probability p . This phenomenon is easy to be explained for sufficiently small p since System 2 has larger offered load than System 1. When p is sufficiently small, the channel throughput would increase with offered load. However, this cannot explain why the performance of System 2 is better even for large p . In fact, it is hard to explain it intuitively in this case. However, Figure 5.17 indicates that the contention probability p_c in System 2 is much less than that in System 1 for a given p . In fact, the offered load in System 2 is less than that in System 1 in this case, and less collisions occur in System 2. When p is sufficiently large, the probability that a new packet is transmitted successfully is very small. That is, the role played by new packet permission probability q is no longer important. As a result, the difference between the system performance in System 1 and 2 become less and less with the increase of p , as shown in Figure 5.14 and Figure 5.15.

5.3 Summary

In this chapter, CDMA S-ALOHA systems with buffered users and bulk arrival process are analyzed using TUA firstly. Because of the unavailability of the queueing theory on

$Geo^{[x]}/Geo/1/K$ system with total rejection model, which each user in fact is, the queueing theory has also been developed. Though the analysis is verified by simulation only in Example 1, it is believed that the accuracy of TUA is satisfied in other cases such as Example 2 since the assumption of TUA that each user always operates with its own equilibrium probabilities has been justified in Example 1. It is shown that the performance analysis using TUA does not incur extra complexity even if the complexity of user queueing behavior increases as long as the corresponding queueing analysis is available. This is because the analyses on channel contention scheme and user queueing behavior are decoupled in TUA. It also shows that channel condition does not introduce extra complexity to the analysis since only packet successful probability p_s is affected. As long as channel PER is known, p_s is obtained using (5.5).

Secondly, a S-ALOHA system in satellite communications is analyzed using TUA. In this case, it is very hard, if not impossible, to know what the packet service time distribution is intuitively. However, state flow graph can aid us to get the distribution. The performance analysis is completed with the help of available queueing analysis on $Geo/G/1/K$ system. The analysis is verified again by simulation. The impact of new packet transmission probability q on the system performance is also studied.

6 BUFFERED CSMA/CD SYSTEM

6.1 Introduction

Carrier sense multiple access with collision detection (CSMA/CD) is a widely used random access protocol in local area networks (LANs) such as Ethernet, the most representative local area network in the world. Although CSMA/CD is not suitable for the application in wireless networks directly, some of the multiple access protocols employed in wireless data networks, such as TSMA/PCD [68] and DSMA/CD, actually borrow the principle of CSMA/CD. Needless to say, the same techniques for the analysis of CSMA/CD protocol can be applied to the analysis of TSMA/PCD and DSMA/CD in a straightforward way.

In this chapter, a homogeneous slotted CSMA/CD system is analyzed using TUA as an example [94]. The numerical results are given.

6.2 System Model

The system employed here is a homogeneous slotted non-persistent CSMA/CD system with finite population of N users having finite buffer capacity of L packets or infinite buffer capacity. Variable packet size is allowed, but all packets require an integer number of time slots for transmission. Denoting by h_k the probability that a packet requires k slots of trans-

mission time, the probability generating function (PGF) $H(z)$ of packet transmission time is given by

$$H(z) = \sum_{k=1}^{\infty} h_k z^k. \quad (6.1)$$

Moreover, DFT transmission principle and late arrival model with delayed replacement rule are employed.

The system operates as follows. Time is divided into slots of length normalized to 1. All packet transmissions start and end at the slot boundary. A busy user senses the channel in each slot with probability p . If the channel is sensed idle, the station starts transmission immediately. If the channel is sensed busy, the station backs off and senses the channel again in next slot with probability p . The channel is noise free. The only reason resulting in unsuccessful transmission is collision, which occurs when more than one user starts transmission at the same slot. No packet can survive the collision. The stations experiencing a collision will detect the collision and abort the transmission within D slots. All packets are of the same length. Packets arrive in the queue of each station following a Bernoulli process such that a message arrives in each slot with probability λ and no messages arrive in a slot with probability $1 - \lambda$. A user would receive the acknowledgment and depart the packet immediately after the packet is successfully transmitted. However, this model can also handle the system where the acknowledgment delay is not ignorable after a minor modification, as described in the following numerical examples. The queue in each user is FCFS. The arriving packets finding the queue full get lost and never return.

6.3 System Analysis Using TUA

Clearly, the CSMA/CD system defined in this chapter is a special case of the general model in Chapter 3. Since this system has only one group of users, we do not need a group identification number. Similar to Figure 3.1, the corresponding state flow graph is obtained and given in Figure 6.1, where S_0 , S_{tx} , S_{bk} and S_e are packet-ready state, transmission state, backoff state and departure state, respectively. p_{stx} is the probability that a busy user starts a transmission in a slot. p_{su} is the probability that a user transmits a packet successfully given that it starts the transmission in a slot.

Figure 6.1. State flow graph of tagged user

6.3.1 Distribution of Virtual Packet Service Time

Similar to (3.1), the distribution of virtual packet service time is given by

$$B(z) = \frac{p_{stx}p_{su}H(z)}{1 - (1 - p_{stx})z - p_{stx}(1 - p_{su})z^D}, \quad (6.2)$$

where

$$p_{stx} = pp_I, \quad (6.3)$$

$$p_{su} = (1 - p_b p)^{N-1}, \quad (6.4)$$

where p_I is the probability that a busy user senses the channel idle in a slot, and p_b is the busy probability of users. A user is said to be busy if its buffer is not empty. Denoting by p_0 the probability that a user buffer is empty, we have

$$p_b + p_0 = 1. \quad (6.5)$$

Using the results in Chapter 3, we have

$$p_I = \frac{E(I) + 1}{E(I) + 1 + (T - 1)q_{\bar{u}} + (D - 1)(1 - q_u - q_{\bar{u}} - p_{ctu})}, \quad (6.6)$$

where

$$T = H'(1), \quad (6.7)$$

and $E(I)$ is the mean length of channel idle duration in each cycle, which is given by

$$E(I) = \frac{1}{1 - (1 - p)(1 - pp_b)^{N-1}} - 1. \quad (6.8)$$

q_u is the probability that a transmission is successful and contributed by the tagged user, and $q_{\bar{u}}$ is the probability that a transmission is successful and contributed by all the other users in the system, given that there is a transmission in a slot. p_{ctu} is the probability that

there is a collision which involves the tagged user given that there is a transmission in a slot. They are given by

$$q_u = [E(I) + 1]p(1 - p_b p)^{N-1} = p p_{su} [E(I) + 1], \quad (6.9)$$

$$q_{\bar{u}} = [E(I) + 1](N - 1)p p_b (1 - p)(1 - p_b p)^{N-2}, \quad (6.10)$$

$$P_{ctu} = \frac{p[1 - (1 - p p_b)^{N-1}]}{1 - (1 - p)(1 - p p_b)^{N-1}}. \quad (6.11)$$

6.3.2 Queueing Analysis of Tagged User

Equations (6.2)-(6.11) show that $B(z)$ is the function of a single unknown parameter p_b . Now, we establish the rest equations by applying classical queueing theory to the queueing system of the tagged user so that the system can be solved. Again, the results are obtained as a special case of the results in Chapter 3. Analysis on the systems with finite buffer capacity and infinite buffer capacity are presented in this section.

6.3.2.1 Finite Buffer Capacity

In this case, the queueing model of tagged user is Geo/G/1/K and the queueing analysis in Section B.5 can be applied to the system analysis. Let α_k be the probability that k packets arrive at the tagged user buffer during a VPST. The PGF $A(z)$ of the number of packets arriving at the buffer during a VPST is given by

$$A(z) = B(1 - \lambda + \lambda z). \quad (6.12)$$

Denoting by π_i , $i = 0, 1, \dots, L - 1$, the probability that a leaving packet sees i packets in queue, then we have

$$\pi_k = \pi_0 a_k + \sum_{j=1}^{k+1} \pi_j a_{k-j+1}, \quad 0 \leq k \leq L-2, \quad (6.13)$$

$$\sum_{i=0}^{L-1} \pi_i = 1. \quad (6.14)$$

Let p_k , $0 \leq k \leq L$, be the probability that there are k packets present in the system at an arbitrary time (slot boundary). After π_k , $0 \leq k \leq L-1$, is obtained using (6.12)-(6.14), p_k , $0 \leq k \leq L$, is given by

$$\begin{aligned} p_k &= \frac{\pi_k}{\pi_0 + \rho}, \quad 0 \leq k \leq L-1 \\ p_L &= p_B = 1 - \frac{1}{\pi_0 + \rho}, \end{aligned} \quad (6.15)$$

where

$$\begin{aligned} \rho &= \lambda b, \\ b &= B'(1), \end{aligned} \quad (6.16)$$

and b is the mean VPST of each user and p_B is packet blocking probability.

Equations (6.2)-(6.16) are sufficient to solve the system.

6.3.2.2 Infinite buffer Capacity

In this case, the proper queueing model of the tagged user is Geo/G/1. Denoting by π_k and p_k , $0 \leq k \leq \infty$, the probability that a leaving packet sees k packets in the queue and the probability that there are k packets present in the queueing system at any time, respectively, it is found [79] that their PGFs, denoted by $\Pi(z)$ and $P(z)$ respectively, are given by

$$P(z) = \Pi(z) = \frac{(1 - \rho)(1 - z)B(1 - \lambda + \lambda z)}{B(1 - \lambda + \lambda z) - z}, \quad (6.17)$$

where ρ is still given by (6.16). It is noticed that $P(z) = \Pi(z)$ holds only for infinite buffer system. In this case,

$$p_0 = 1 - \rho. \quad (6.18)$$

6.4 Performance Evaluation

In this section, the system performance is evaluated using the results obtained in Section 6.3. The performance measurements are focused on channel throughput, user queue length, packet response time, and channel interdeparture time, whose definitions are given in section 1.2 and 3.4. The results are derived following section 3.4.

6.4.1 Finite Buffer System

It is found that the channel throughput Θ is given by

$$\Theta = N\lambda T(1 - p_B), \quad (6.19)$$

where p_B is the packet blocking probability equal to p_L . The PGF of queue length at any time is given by

$$P(z) = \sum_{k=0}^L p_k z^k, \quad (6.20)$$

and the mean queue length is given by

$$E(l) = \sum_{k=0}^L k p_k. \quad (6.21)$$

The PGF and mean of the waiting time are given by

$$W(z) = \frac{\pi_0(1-z)}{1-\lambda-z+\lambda B(z)} \left(1 - \left[\frac{\lambda B(z)}{z-1+\lambda}\right]^L\right) + [B(z)]^{L-1} \sum_{j=0}^{L-1} \pi_j \left(\frac{\lambda}{z-1+\lambda}\right)^{L-j}, \quad (6.22)$$

and

$$E(w) = W'(1) = \frac{E(l)}{\lambda(1-p_B)} - b, \quad (6.23)$$

respectively.

6.4.2 Infinite Buffer System

Since all of the arriving packets have to be transmitted in this case. If the system is stable, the channel throughput is given by

$$\Theta = N\lambda T. \quad (6.24)$$

The PGF of queue length is already given in (6.17). Hence, the mean queue length is given by

$$E(l) = P'(1) = \rho + \frac{\lambda^2 b^{(2)} - \lambda \rho}{2(1-\rho)}, \quad (6.25)$$

where ρ and b are still given by (6.16), $b^{(2)}$ is given by

$$b^{(2)} = B''(1) + B'(1). \quad (6.26)$$

The PGF and mean of the waiting time are given by

$$W(z) = \frac{(1-\rho)(1-z)}{1-\lambda-z+B(z)}, \quad (6.27)$$

and

$$E(w) = W'(1) = \frac{E(l)}{\lambda} - b, \quad (6.28)$$

respectively.

As to the user output process, the formulas for both finite and infinite buffer systems are exactly same, which is given by

$$O(z) = B(z) + \frac{\pi_0(z-1)B(z)}{1-(1-\lambda)z}. \quad (6.29)$$

The difference between the two systems is justified by the fact that the variables in (6.29) may take different values. Similarly, the channel output process can be obtained using (3.44)-(3.47).

6.5 The Algorithm

As mentioned above, we have sufficient equations to determine $B(z)$. However, these are a set of nonlinear equations which is very hard to be solved analytically, if not impossible. Fortunately, iterative algorithms can be developed straightforwardly for both finite and infinite buffer systems following the steps for deriving the equations in section 6.3 and 6.4. The iterative algorithms operate as follows.

- 1 selecting a very small positive number δ , say $\delta = 1.0e-6$, and setting an arbitrary initial value between zero and one for p_0 ;
- 2 calculating $B(z)$ using (6.2)-(6.11);
- 3 using (6.12)-(6.16) and (B.22)-(B.24) to obtain a new estimate $\hat{p}_0$ of the probability p_0 for finite buffer system, or using (6.16) and (6.17) for infinite buffer system;
- 4 if $|\hat{p}_0 - p_0|/p_0 \geq \delta$, let $p_0 = \hat{p}_0$ and goto step 2. Otherwise, p_0 and $B(z)$ are obtained.
- 5 calculate the corresponding performance measurements using the obtained results.

In step 4, we do not check if other variables such as p_k , $0 \leq k \leq L$, converge. In fact, in Geo/G/1/K queueing system, all variables are related to p_0 directly or indirectly. They

are known as soon as p_0 becomes known. Therefore, they converge as fast as p_0 . Hence, it is not necessary to check them out. Numerical analysis verified this conclusion.

6.6 Numerical Results

In this section, four numerical examples are presented. In first two examples, we apply TUA to CSMA/CD system models in [1] and [68], respectively. In example 3, we apply TUA to a CSMA system with fixed packet length, which is a special case of CSMA/CD system. The analytical results are compared with simulation results contributed in [1] and [68]. In example 4, we consider a heterogeneous CSMA/CD system.

6.6.1 Example 1: Homogeneous Systems

A system model in [1] with the following parameters is considered here: $N = 100$, $\lambda = 0.00035$, $T = 20$. The buffer capacity and user permission probability are variables.

In this model, it is assumed that the channel be sensed idle one slot after the end of a packet transmission. That is, the channel propagation delay is one slot. However, zero propagation delay is assumed in the TUA analysis developed in this chapter. In order to apply TUA to this system, the system is treated as a model with $T = 21$ and $D = 2$ but no propagation delay, as discussed in Section 3.4.4.2. Then the model is analyzed using TUA. However, if the obtained throughput for the modified model is x , the system throughput is given by $\Theta = 20x/21$ since each packet of length 21 slots contains only 20 slot information.

The channel throughput is given in Figure 6.2 for different buffer capacities and permission probabilities, where the lines are results obtained using TUA, and markers are sim-

ulation results from [1]. When permission probability is small enough, the system has only

Figure 6.2. Channel throughput of a CSMA/CD system

one equilibrium point, as indicated by the solid lines. However, when the permission probability is sufficiently large, the system has two equilibrium points. One corresponds to higher throughput, as indicated by dashed lines. The other corresponds to channel congestion. In this case, channel throughput is low, as indicated by solid lines.

The packet response time is illustrated in Figure 6.3. The results show that the analysis and simulation fits very well. The analytical results for the systems using EPA are also available in [1]. Comparing the results obtained using TUA with those using EPA, it is found that the accuracy of TUA and EPA is in the same order.

Figure 6.3. The system response time of packets of a CSMA/CD system

The probability distribution of user buffer length for CSMA/CD systems with different configurations is given in Table 6.1, Table 6.2 and Table 6.3, where the source for simula-

Table 6.1. The probability distribution p_k of user buffer length in a CSMA/CD system with $N = 100$, $\lambda = 0.00035$, $T = 20$, $D = 1$, $L = 8$ and $p = 0.01$

k	TUA	SIMULATION	EPA
0	0.8458	0.855	0.859
1	0.1314	0.124	0.122
2	0.1947E-1	0.172E-1	0.165E-1
3	0.2882E-2	0.287E-2	0.224E-2
4	0.4265E-3	0.549E-3	0.303E-3
5	0.6312E-4	0.180E-4	0.410E-4
6	0.9343E-5	0.900E-5	0.555E-5
7	0.1383E-5	0.000	0.751E-6
8	0.2032E-6	0.000	0.101E-6

Table 6.2. The probability distribution p_k of user buffer length in a CSMA/CD system with $N = 100$, $\lambda = 0.00035$, $T = 20$, $D = 1$, $L = 8$ and $p = 0.0012$

k	TUA	SIMULATION	EPA
0	0.1515	0.203	0.150
1	0.1402	0.162	0.139
2	0.1288	0.137	0.128
3	0.1183	0.113	0.118
4	0.1087	0.996E-1	0.109
5	0.9982E-1	0.943E-1	0.100
6	0.9168E-1	0.734E-1	0.924E-1
7	0.8421E-1	0.632E-1	0.851E-1
8	0.7678E-1	0.536E-1	0.778E-1

Table 6.3. The probability distribution p_k of user buffer length in a CSMA/CD system with $N = 100$, $\lambda = 0.000175$, $T = 40$, $D = 1$, $L = 8$ and $p = 0.01$

k	TUA	SIMULATION	EPA
0	0.9168	0.918	0.931
1	0.7682E-1	0.752E-1	0.648E-1
2	0.5909E-2	0.589E-2	0.405E-2
3	0.4528E-3	0.548E-3	0.253E-3
4	0.3470E-4	0.454E-4	0.158E-4
5	0.2659E-5	0.000	0.989E-6
6	0.2037E-6	0.000	0.619E-7
7	0.1561E-7	0.000	0.387E-8
8	0.1187E-8	0.000	0.240E-9

tion results and EPA analytical results is [1]. Note that the system in Table 6.3 has different packet size from Table 6.1 and Table 6.2, but the generated data rate $N\lambda T$ is same. The simulation results in the tables again confirm that the accuracy of TUA analysis is excel-

lent. It should be noticed that there exist larger error in the estimation of probability distribution of buffer length than the estimation of channel throughput or mean packet response time by simulation since longer simulation time is required for the former, particularly if some of the probabilities are very small.

6.6.2 Example 2: Optimal CSMA/CD System

Here we consider a system without propagation delay, that is, $D = 1$. The user population and packet size are fixed and given by $N = 10$ and $T = 40$. The buffer capacity, permission probability, and packet generating rate are variables.

We suppose that during a contention period each user would attempt to sense the channel in a slot with probability $1/n^*$, where n^* is the number of busy users at the beginning of the slot. It is proved [68] that if the above scheme is employed for determining user permission probability, the system performance is optimal in the term of throughput.

In order to apply TUA to optimal CSMA/CD system, the analytical approach developed in this chapter has to be modified. Now, we determine the permission probability for the tagged user. According to the assumptions used in TUA, the tagged user would see each of the other users busy with probability p_b . During the contention period, the tagged user is busy with probability 1. Hence, there are in average $n^* = 1 + (N - 1)p_b$ busy users in the system when the tagged user contends for channel. By the scheme above, the permission probability of the tagged user is given by

$$p = \frac{1}{1 + (N - 1)p_b}. \quad (6.30)$$

Comparing the analysis before, we have one more unknown parameter p but also one more equation (6.30). Hence, the system can be solved. The numerical algorithm developed before can be applied to this case with a little modification.

Figure 6.4. Mean queue length versus total packet generating rate of an optimal CSMA/CD system

The numerical results are given in Figure 6.4, Figure 6.5 and Figure 6.6, where the lines correspond to analytical results using TUA and markers to simulation results from [68]. Though the user population is small, the accuracy of TUA is still very good. In fact, the errors between TUA and simulation are less than what is shown in the diagrams since the simulation itself has errors as shown in Figure 10 of [68]. Seriously speaking, an optimal CSMA/CD system should have constant throughput for sufficiently long packet response time. However, the throughput decreases very slowly with the increasing of packet

Figure 6.5. Mean packet response time (in the unit of packet transmission time) versus total packet generating rate of a optimal CSMA/CD system

Figure 6.6. Packet blocking probability versus total packet generating rate of a optimal CSMA/CD system

response time in Figure 10 of [68]. That is, the simulation model is not optimal. However, the model of TUA here is optimal, as illustrated by Figure 6.7, where the channel through-

Figure 6.7. Channel throughput versus total packet generating rate of a optimal CSMA/CD system

put increases toward a constant for sufficiently large $N\lambda$. This may be part of the reasons that the results of TUA are a little bit better than that of simulation.

6.6.3 Example 3: CSMA System

The CSMA system in Figure 10.5 of [1] is considered here. We have $N = 100$ and $\lambda = 0.00035$. The packet transmission time is 20 slots like in example 1. Since one slot propagation delay is assumed in the system, we have $T = 21$. In CSMA system, a user would know a collision only after a packet is transmitted. Hence, we have $D = T = 21$.

The numerical results are given in Figure 6.8 and Figure 6.9, where the lines and mark-

Figure 6.8. Channel throughput versus transmission permission probability in a CSMA system

Figure 6.9. Packet response time versus transmission permission probability in a CSMA system

ers correspond to results obtained using TUA and simulation from [1], respectively. It is found that the system has only one equilibrium point for sufficiently small or large permission probability p , and two equilibrium points for the permission probability between them. This is because too many slots are wasted in each collision. When p is sufficiently large, frequent collisions result in the channel congestion. In fact, this is also true for CSMA/CD system. However, only few slots are wasted in each collision in CSMA/CD system. In wide range values of p , the system would be bi-stable. The system status would oscillate between these two equilibrium points. Only for very large p the system would become congested with probability 1. For instance, in example 1, the CSMA/CD system would become congested with probability 1 for some sufficiently large p which is not shown in the diagrams.

The analytical results on this model using EPA is available in [1]. The EPA analysis shows that the system with buffer capacity of 1, 8 and 32 packets is bi-stable for p beyond points C, B and A, respectively. This is quite different from the results obtained here using TUA. Based on the analysis above, we believe that the results from TUA are more reasonable.

The probability distribution of user buffer length for CSMA/CD systems with different configurations is given in Table 6.4 and Table 6.5, where the sources for the results of sim-

Table 6.4. The probability distribution p_k of user buffer length in a CSMA system with $N = 100$, $\lambda = 0.00035$, $T = 20$, $L = 8$ and $p = 0.004$

k	TUA	SIMULATION	EPA
0	0.5284	0.536	0.546
1	0.2513	0.239	0.250
2	0.1177	0.115	0.113

Table 6.4. The probability distribution p_k of user buffer length in a CSMA system with $N = 100$, $\lambda = 0.00035$, $T = 20$, $L = 8$ and $p = 0.004$

3	0.5513E-1	0.540E-1	0.507E-1
4	0.2582E-1	0.307E-1	0.228E-1
5	0.1209E-1	0.131E-1	0.103E-1
6	0.5662E-2	0.645E-2	0.463E-2
7	0.2652E-2	0.386E-2	0.209E-2
8	0.1233E-2	0.213E-2	0.933E-3

Table 6.5. The probability distribution p_k of user buffer length in a CSMA/CD system with $N = 100$, $\lambda = 0.00035$, $T = 20$, $L = 8$ and $p = 0.002$

k	TUA	SIMULATION	EPA
0	0.2596	0.325	0.259
1	0.1995	0.228	0.199
2	0.1518	0.157	0.152
3	0.1155	0.108	0.116
4	0.8789E-1	0.728E-1	0.880E-1
5	0.6688E-1	0.455E-1	0.670E-1
6	0.5089E-1	0.288E-1	0.510E-1
7	0.3873E-1	0.205E-1	0.388E-1
8	0.2925E-1	0.144E-1	0.294E-1

ulation and EPA analysis is [1]. Again, the results of TUA analysis and simulation match well.

6.6.4 Example 4: Heterogeneous Systems

In this section we apply TUA to a heterogeneous CSMA/CD system with two groups of users. The user populations in the two groups are $N_1 = 40$ and $N_2 = 20$, and the user

buffer capacities are $L_1 = 30$ and $L_2 = 10$. For each user, the permission probability for a new packets is one, or $q_1 = q_2 = 1$. The permission probability for a deferred packet will be given in the diagrams below. In addition, we have

$$H_1(z) = z^{51} \quad (6.31)$$

and

$$H_2(z) = \frac{0.05z^{42}}{1 - 0.95z}. \quad (6.32)$$

That is, the packet generated in Group 1 has fixed size of 100 slot transmission time, and the packet generated in Group 2 has a size of 40 slot plus a geometrically distributed number of slot with mean of 20 slots. We also assume that a transmitting user need two slots to detect any collision. Hence, we have $D = 2$. In addition, we assume that packets in Group 2 need more prompt delivery than packets in Group 1 such that we always have $p_2 = 5p_1$, where p_1 and p_2 are permission probabilities for deferred packets in Group 1 and 2, respectively. In order to avoid three dimensional graphs, we assume that $\lambda_1 = 1.6\lambda_2$.

The numerical results for three set of permission probabilities are illustrated in Figure 6.10 to Figure 6.13. The total generated load G is defined as the total generated bits in each bit time, which is given by $N_1 T_1 \lambda_1 + N_2 T_2 \lambda_2$, where T_1 and T_2 are the mean length of packets in Group 1 and 2, respectively.

In Figure 6.10, the channel throughput and group throughput are given, where group throughput is the throughput contributed by all users in one group. Here we focus on group

Figure 6.10. Channel throughput and group throughputs

Figure 6.11. Mean packet response time of users in Group 1 and 2

Figure 6.12. Mean queue length of users in Group 1 and 2

Figure 6.13. Packet blocking probability of users in Group 1 and 2

throughput since the channel throughput is just the summation of all group throughputs. When G is sufficiently small, there is no congestion in the channel and almost all generated packets are delivered. Hence, group throughput is dominated by the total generated load in each group. Based on the parameters above, it is easy to find that the generated load in Group 1 is twice of that in Group 2. However, with the increase of G , channel congests gradually. The group throughput is no longer dominated by group generated load, but by the permission probability. Users with higher permission probability have more chance to send a packet out. The throughput in Group 2 is greater than that in Group 1 for sufficiently large G since $p_2 = 5p_1$. It is found that the channel throughput and group throughput are close to constants for sufficiently large G . This is because the users in both group are almost busy with probability 1 in this case. The system with all user busy is operating at its worst case. Further increase of total generated load has no impact on channel congestion. However, packet blocking probability would increase with G as shown in Figure 6.13.

Figure 6.11 indicates that the mean packet response time in Group 1 is always greater than that in Group 2. This is because the buffer capacity and packet arrival rate of users in Group 1 are larger and the permission probability of users in Group 1 is less. The less permission probability results in longer packet service time. The larger packet arrival rate and buffer capacity results in more packets waiting in the buffer. As illustrated in Figure 6.12, the buffer of users in Group 1 is almost full for $G > 1$. However, the buffer of users in Group 2 is full only if $G > 2.5$. The packet blocking probability given in Figure 6.13 confirms our analysis.

If we set the same parameters to users in both group, the numerical results show that the users in both group are statistically identical though different initial conditions are employed for each group. This should be true since the system is indeed a homogeneous system in this case.

6.7 Summary

In this chapter, a detailed analysis on CSMA/CD systems using TUA is demonstrated. Four numerical examples are given. In the first one, TUA is applied to a homogeneous CSMA/CD system with DFT transmission principle where permission probability is pre-determined. In the second one, TUA is applied to a optimal CSMA/CD where the permission probability is adaptive to the system status so that maximum channel throughput can be reached. In the third one, TUA is applied to a homogeneous CSMA system which is a special case of CSMA/CD. In the three examples, simulation results contributed by other authors are used to verify the accuracy of TUA analysis. It is found that the accuracy of TUA is very good. Comparatively large error between simulation and analysis occurs in Example 2 because of small user population and buffer capacity. Finally, a heterogeneous system consisting of two groups of users is studied using TUA in Example 4. NFT transmission principle is employed for all users. This is to demonstrate the power of TUA to handle complicated systems.

7 BUFFERED R-ALOHA SYSTEMS

7.1 Introduction

R-ALOHA is a multiple access protocol originally proposed in [86] to improve the throughput of satellite channels. It is an efficient multiple access protocol in satellite communications and other broadcast media because of its bandwidth efficiency and simplicity in implementation. The basic principles of R-ALOHA are also employed in some other multiple access protocols such as packet reservation multiple access (PRMA) [36]-[41], a wireless access protocol targeted to the third generation wireless communications systems for the integration of packetized voice and data over a common communication channel. In fact, R-ALOHA without channel propagation delay can be considered as a special case of PRMA system with pure voice users where the maximum packet holding time is infinite, vice versa.

In R-ALOHA, the time is divided into slots which are grouped into frames. Each frame consists of a fixed number of time slots. Each slot is long enough for the transmission of one packet. A slot can have either a reserved or available (nonreserved) status. Users contend for access to an available slot on the shared channel using S-ALOHA protocol. If there is no transmission in a slot, this slot is considered to be available for all users in the subsequent frames. If a user is successful in transmitting in an available slot, this slot is reserved

for this user in the subsequent frames until up to a maximum number of packets are transmitted. The slot is then released and becomes available to all users again. How to select the maximum number is a challenge. Normally, large number would make more efficient utilization of the channel, but too large number may result in channel unfairness to all users.

There are two slightly different versions of R-ALOHA, depending on whether an end-of-use flag is included in the header of the last packet of a message:

(P1) end-of-use flag is not included, and

(P2) end-of-use flag is included.

With (P1), a time slot is always wasted whenever a user releases its reserved slot. Hence, (P2) gives rise higher channel throughput than (P1). However, (P1) is easier to implement [33].

The pioneer analysis on R-ALOHA was done in [33] based on the assumption that the channel propagation delay be zero, and a successful transmission would occur in each available slot with a constant probability. The study focused on the statistic behavior of the overall system, and had no answer on the system stability or the behaviors of individual users. The stability and dynamic behavior of R-ALOHA systems are studied in [1][34] using EPA, and the channel propagation delay is considered. However, it is assumed that the number of packets contained in each message be geometrically distributed, and each user buffer can hold up to one message. Because of the modeling complexity of EPA, it seems to be difficult, if not impossible, to apply EPA to R-ALOHA systems where each message consists of a generally distributed number of packets, or each user has a buffer capacity of more than one message.

In this chapter, R-ALOHA systems are analyzed using TUA [95], where each message consists of a generally distributed number of packets. The channel propagation delay in satellite channel is considered. Each user has a buffer capacity of L messages. To maintain the fairness to all users, it is assumed that each message be treated independently. That is, a user would release its reserved slot after finishing the transmission of a message even if it has one or more messages waiting in the buffer. It is noticed that buffered S-ALOHA protocols in satellite communications [35] can be considered as a special case of R-ALOHA with (P2) where each message consists of one packet. Hence, the R-ALOHA system studied here is more general than that in [34][35].

7.2 System Model

The R-ALOHA system consists of N users and each user has a buffer capacity of L messages. Message arrival to the buffer follows a Bernoulli process with arrival rate λ . An arriving message finding buffer full is rejected. Each message consists of a random number of packets with PGF $G(z)$ given by

$$G(z) = \sum_{i=1}^{\infty} g_i z^i, \quad (7.1)$$

that is, the probability that a message contains i packets is g_i . Time is divided into slots of length normalized to 1, which is also the packet transmission time. All packet transmissions start and end at the slot boundary. Each frame consists of M slots. If a message becomes ready for transmission immediately before an available slot, the user would transmit the first packet of the message with probability 1 in this slot. Otherwise, the user would contend

for each available slot with probability p . The channel is assumed to be noise free, and transmission is unsuccessful only in the case of collision. The channel propagation delay is assumed to be $D - 1$ slots, that is, if a user starts a transmission in the beginning of a slot, it takes D slots to receive the acknowledgment. Note we must have $D \leq M$ in order to operate the system properly. In this chapter, only R-ALOHA systems with (P2) is studied. However, the same idea is applicable in the systems with (P1).

7.3 System Analysis

Here we treat message, not packet, as the customer of user queueing system. In order to analyze the R-ALOHA system using TUA, firstly we find the distribution of virtual message service time (VMST) by analyzing the channel contention using state flow graph. Then the obtained VMST distribution is applied to the queueing analysis of the tagged user using available queueing theory.

7.3.1 Message Service Time Distribution

The VMST distribution is obtained using the state flow graph of the tagged user by observing its behavior from a message ready for transmission to its successful departure as illustrated in Figure 7.1, where S_0 is the message-ready state to which a busy user enters when it has a new message ready for transmission. Since a user at S_0 would issue a transmission with probability 1 if the present slot is available, it will enter transmission state S_{tx} with slot available probability p_a , or enter backoff state S_{bk} with probability $1 - p_a$ and spend one slot time there. At state S_{bd} , the head of line message is no longer new. Conse-

Figure 7.1. State flow graph of the tagged user under (P2)

quently, a user would contend for each available slot with probability p at S_{bk} . That is why a user would enter S_{tx} from S_{bk} with probability pp_a , or stay at S_{bk} for one more slot with probability $1 - pp_a$. S_r is the reservation state to which a contending user enters immediately after it captures a slot by transmitting the first packet of a message successfully. A user at S_{tx} would enter S_r after one slot with probability p_s , the probability that the tagged user captures a slot given that it transmits a packet in this slot. Since a user needs $D - 1$ slot to know the status of a transmitted packet, a user at S_{tx} would return to S_{bk} with probability $1 - p_s$, after D slots. Finally, S_e is the departure state where the message is departed from the user. Since the PGF $G_s(z)$ of the number of slots required before releasing the reserved slot is given by

$$G_s(z) = \sum_{i=1}^{\infty} g_i z^{(i-1)M} = G(z^M)/z^M, \quad (7.2)$$

the transfer function from S_r to S_e is given by $G_s(z)z^{D-1}$, where z^{D-1} is due to $D-1$ slots of propagation delay.

Referring to Figure 7.1, the PGF $B(z)$ of VMST is just the time required for a user at S_0 to enter S_e . It is given by

$$B(z) = C(z)G_s(z)z^{D-1}, \quad (7.3)$$

where $C(z)$ is the PGF of the time for a user at S_0 to enter S_r , which is given by

$$C(z) = \frac{p_a p_s z (1 - z + pz)}{1 - (1 - p p_a)z - p p_a (1 - p_s)z^D}. \quad (7.4)$$

As to the unknown parameters p_a and p_s , they will be given later on.

7.3.2 Queueing Analysis

By (7.3) the VMST distribution is generally distributed. Hence, each user can be treated as a Geo/G/1/K queueing system, and the queueing theory in Section B.5 can be employed for the system analysis.

Let a_k be the probability that k messages arrive at the tagged user buffer during a VMST. The PGF $A(z)$ of the number of messages arriving at the buffer during a VMST is given by

$$A(z) = B(1 - \lambda + \lambda z). \quad (7.5)$$

Denoting by π_i , $i = 0, 1, \dots, L-1$, the probability that a leaving message sees i messages in queue, then we have

$$\pi_k = \pi_0 a_k + \sum_{j=1}^{k+1} \pi_j a_{k-j+1}, \quad 0 \leq k \leq L-2. \quad (7.6)$$

$$\sum_{i=0}^{L-1} \pi_i = 1. \quad (7.7)$$

An efficient algorithm for computing π_k , $0 \leq k \leq L-1$ using (7.6) and (7.7) can be found in Section B.5.

Let p_k , $0 \leq k \leq L$, be the probability that there are k messages present in the system at an arbitrary time (slot boundary). After π_k , $0 \leq k \leq L-1$, is obtained, p_k , $0 \leq k \leq L$, is given by

$$\begin{aligned} p_k &= \frac{\pi_k}{\pi_0 + \rho}, \quad 0 \leq k \leq L-1 \\ p_L &= 1 - \frac{1}{\pi_0 + \rho} = p_B, \end{aligned} \quad (7.8)$$

where

$$\begin{aligned} \rho &= \lambda b, \\ b &= B'(1), \end{aligned} \quad (7.9)$$

and b is the mean VMST of each user and p_B is message blocking probability.

In order to determine p_a and p_s , further analysis on the user behavior is necessary. A user can be in one of the four modes: IDL (idle), CON (contending for channel), WAI (waiting for acknowledgment) and REV (reservation). A user is said to be at IDL if its

buffer is empty, CON if it can contend for the channel in each available slots, WAI if it is waiting for the acknowledgment for a transmitted packet, and REV if has a reserved slot.

Substituting (7.2) and (7.4) into (7.3) and using (7.9), we have

$$b = \frac{1}{pp_a p_s} - \frac{1}{p} + 1 + \frac{D-1}{p_s} + M[G'(1) - 1], \quad (7.10)$$

where, it is easy to find that

$$T_1 \equiv M[G'(1) - 1], \quad (7.11)$$

$$T_2 \equiv (D-1)/p_s, \quad (7.12)$$

and

$$T_3 \equiv \frac{1}{pp_a p_s} - \frac{1}{p} + 1 \quad (7.13)$$

are the mean times of a busy user spending at REV, WAI, and CON respectively during each message service time. Note that Mode WAI disappears if the propagation delay is zero as indicated by (7.12).

Now we treat the time in the tagged user as a consequence of cycles. Each cycle consists of an idle period and a busy period. The mean length of the idle period is

$$T_{IDL} = 1/\lambda \quad (7.14)$$

since message arrivals follow a Bernoulli process with rate λ . Since each departure message would leave the buffer empty with probability π_0 , the number of messages served in each busy period is geometrically distributed with the mean $1/\pi_0$. Hence, in each cycle

the mean times T_{REV} , T_{WAI} and T_{CON} of a busy user spending at REV, WAI, and CON are given by

$$T_{REV} = T_1/\pi_0 = M[G'(1) - 1]/\pi_0, \quad (7.15)$$

$$T_{WAI} = T_2/\pi_0 = \frac{D-1}{\pi_0 p_s}, \quad (7.16)$$

and

$$T_{CON} = T_3/\pi_0 = \frac{1}{\pi_0} \left(\frac{1}{p p_a p_s} - \frac{1}{p} + 1 \right), \quad (7.17)$$

respectively. Since the average number of reserved slots in any frame should be equal to the average number of users at REV, and the tagged user at CON does not reserve a slot, we have

$$(N-1)p_r = M p_{rev}, \quad (7.18)$$

where p_r is the probability that a user is at REV mode, which is given by

$$p_r = \frac{T_{REV}}{T_{IDL} + T_{WAI} + T_{CON} + T_{REV}}, \quad (7.19)$$

and p_{rev} is the probability that a slot is reserved. Therefore, the slot available probability is

$$p_a = 1 - p_{rev}. \quad (7.20)$$

Define the following events: $A_1 = \{\text{tagged user is transmitting}\}$, $A_2 = \{\text{the present slot is available}\}$, $S = \{\text{the transmission of the tagged user is successful}\}$, and $R_i = \{i \text{ slots are reserved at present frame}\}$, we have

$$\begin{aligned}
p_s &= \text{Prob}(S|A_1, A_2) \\
&= \sum_{i=0}^{M-1} \text{Prob}(S|A_1, A_2, R_i) \cdot \text{Prob}(R_i|A_1, A_2) \\
&= \sum_{i=0}^{M-1} (1 - p_{c|\bar{R}}p - p_{idle|\bar{R}}\lambda)^{N-1-i} \binom{M-1}{i} p_{rev}^i (1 - p_{rev})^{M-1-i} \\
&= (1 - p_{c|\bar{R}}p - p_{idle|\bar{R}}\lambda)^{N-1} \left[\frac{p_{rev}}{1 - p_{c|\bar{R}}p - p_{idle|\bar{R}}\lambda} + 1 - p_{rev} \right]^{M-1},
\end{aligned} \tag{7.21}$$

where $p_{c|\bar{R}}$ and $p_{idle|\bar{R}}$ are the probabilities that a user is at CON and IDL, respectively, given that it is not at REV. They are given by

$$p_{c|\bar{R}} = \frac{T_{CON}}{T_{IDL} + T_{WAI} + T_{CON}} \tag{7.22}$$

and

$$p_{idle|\bar{R}} = \frac{T_{IDL}}{T_{IDL} + T_{WAI} + T_{CON}}. \tag{7.23}$$

At this point, we have actually obtained sufficient equations to solve all the unknown parameters in the system. It is tedious to check it out by counting all the unknown parameters and the number of equations. In fact, equations (7.1)-(7.19), (7.22) and (7.23) are required to calculate p_{rev} , $p_{c|\bar{R}}$ and $p_{idle|\bar{R}}$, given p_a and p_s . It implies that p_{rev} , $p_{c|\bar{R}}$ and $p_{idle|\bar{R}}$ are the functions of p_a and p_s . However, equations (7.20) and (7.21) implies that p_a and p_s are also functions of p_{rev} , $p_{c|\bar{R}}$ and $p_{idle|\bar{R}}$. Since all the obtained equations are independent, the system can be solved. These are a set of complicated nonlinear

equations. An explicit solution is almost impossible. However, numerical algorithm can be developed to solve them recursively.

7.4 Performance Evaluation

The concerned performance measurements are channel throughput, mean waiting time, mean response time, and message blocking probability which are defined as the average number of packets transmitted successfully in each slot, the mean time that an accepted message has to wait before being served, the mean time that a message spends from arrival to departure, and the percentage of generated messages which are rejected.

The channel throughput is just the average number of packets accepted into the system in each slot, which is given by

$$\Theta = N\lambda G'(1)[1 - p_L]. \quad (7.24)$$

By Little's formula, the mean message response time is given by

$$\bar{R} = \frac{E(l)}{\lambda[1 - p_L]}, \quad (7.25)$$

where l is the queue length of users. Its mean is given by

$$E(l) = \sum_{k=1}^L k p_k. \quad (7.26)$$

Hence, the mean message waiting time is given by

$$\bar{W} = \frac{E(l)}{\lambda[1 - p_L]} - b. \quad (7.27)$$

What have been obtained above are just mean values. However, the distribution of interested performance measurements can also be obtained using TUA. This is done by employ-

ing the results in classical queueing theory directly. For instance, by (B.9) and (B.27) the PGF $R(z)$ of the message response time is given by

$$R(z) = \frac{\pi_0(1-z)B(z)}{1-\lambda-z+\lambda B(z)} \left(1 - \left[\frac{\lambda B(z)}{z-1+\lambda} \right]^L \right) + [B(z)]^L \sum_{j=0}^{L-1} \pi_j \left(\frac{\lambda}{z-1+\lambda} \right)^{L-j} \quad (7.28)$$

7.5 Numerical Algorithm

In this section, an iterative algorithm for solving the obtained equations is developed which operates as follows: (i) selecting a very small positive number, say $\delta = 1.0e-6$ and $n = 0$, and setting an arbitrary initial values between 0 and 1 for p_a and p_s , say $p_a^{(0)} = 0.5$ and $p_s^{(0)} = 0.5$; (ii) using (7.1)-(7.19), (7.22) and (7.23) to calculate p_{rev} , $p_{c|\bar{R}}$ and $p_{idle|\bar{R}}$; (iii) using the obtained p_{rev} , $p_{c|\bar{R}}$ and $p_{idle|\bar{R}}$ to calculate new values of p_a and p_s by (7.20) and (7.21), called p_{anew} and p_{snew} ; (iv) updating p_a and p_s by

$$p_a^{(n+1)} = p_a^{(n)} + \alpha(p_{anew} - p_a^{(n)}) \quad (7.29)$$

and

$$p_s^{(n+1)} = p_s^{(n)} + \alpha(p_{snew} - p_s^{(n)}), \quad (7.30)$$

where α is the step size between 0 and 1; (v) if both conditions $|p_a^{(n+1)} - p_a^{(n)}|/p_a^{(n)} \leq \delta$ and $|p_s^{(n+1)} - p_s^{(n)}|/p_s^{(n)} \leq \delta$ are satisfied, the equations above are considered to be solved, otherwise, repeat steps from (i) to (v).

The step size α is used to control the convergence speed and stability of the algorithm. Larger α leads to faster convergence speed if the algorithm converges. However, the algo-

rithm may oscillate around the equilibrium point but never converge to it if α is too large.

Smaller α normally results in slower convergence speed but better stability.

It should be mentioned that the system has two locally stable equilibrium points when it is bi-stable. Which equilibrium point the algorithm would converge to depends on the initial values of p_a and p_s . As indicated by numerical results, the algorithm converges to the equilibrium point corresponding to high throughput when the initial values of p_a and p_s are large. When the initial values of p_s is very small, the algorithm converges to the equilibrium point corresponding to channel congestion.

7.6 Numerical Examples

Two examples are given in this section. In the first example, a R-ALOHA system is analyzed where the number of packets contained in each message is geometrically distributed, and each user can hold at most one message in the buffer. This system is a special case of the model considered in this chapter. The analytical result is verified by simulation. In the second example, we consider a more complicated R-ALOHA system. In both cases, systems consist of 100 users, each frame consists of 13 slots and the channel propagation delay is 12 slots. That is, the following parameters are used: $N = 100$, $M = 13$, $D = 13$.

7.6.1 Example 1

A R-ALOHA system used in [1][34] is considered here where the number of packets contained in each message is geometrically distributed with mean $1/\gamma$, and each user buffer can hold at most one message. That is, we have $G(z) = \frac{\gamma z}{1 - (1 - \gamma)z}$, and $L = 1$.

The analytical results obtained using TUA and the simulation results in [1][34] are given in Figure 7.2 and Figure 7.3 for $\lambda = 0.0005$ and in Figure 7.4 and Figure 7.5 for $\lambda = 0.001$. As illustrated in these figures, the analysis and simulation fit very well most of the time. The most obvious mismatch occurs in Figure 7.4 and Figure 7.5 for $\gamma = 0.1$. This is because the system is bi-stable in this case for $p > 0.172$. The dotted line corresponds to the equilibrium point where the system has higher throughput, and the dashed line corresponds to the one where the system is congested. When a system is bi-stable, it would oscillates between these two equilibrium points. This implies that the system throughput should locate between those two lines. That is why the simulation results do not follow any one of the two lines, but locate somewhere between them.

The analytical results using EPA on the system above are also available in [1][34]. It is noticed that the accuracy of TUA and EPA is in the same order. In fact, Figure 7.4 and Figure 7.5 shows that TUA is a little bit more accurate. Considering the fact that EPA is the only available analytical approach for R-ALOHA which addresses system stability and the behaviors of individual users, the contribution of TUA for R-ALOHA analysis is obvious.

7.6.2 Example 2

In this example, a R-ALOHA system where the message size (in packets) is generally distributed is considered. Here we have $G(z) = 0.2z^{16} + 0.6z^{11} + 0.2z$ and $L = 5$. The average message size is $G'(1) = 10$ packets. The numerical results are given in Figure 7.6, Figure 7.7, Figure 7.8 and Figure 7.9 for different permission probabilities and message arrival rates. It is noticed that the system has only one equilibrium point for

Figure 7.2. Channel throughput

Figure 7.3. Mean message response time

Figure 7.4. Channel throughput

Figure 7.5. Mean message response time

$\lambda = 0.001$ in the region of $p \leq 0.1$. However, it is bi-stable for $\lambda = 0.0005$ and $\lambda = 0.0008$ in the region of $p \geq 0.083$ and $p \geq 0.044$, respectively, as shown in the diagrams. An interesting phenomenon is that the mean message response time decreases and increases monotonically for permission probability less than and greater than around 0.015, respectively given $\lambda = 0.001$, as shown in Figure 7.7. This is because the channel collision seldom occurs when permission probability is very small. The channel access delay is dominated by backoff time before each contention. The smaller the permission probability, the longer the backoff time and therefore the channel access delay and message response time. When the permission probability is sufficiently large, collision occurs frequently and channel congests. In this case, channel access delay is dominated by collisions. The larger the permission probability, the more frequently the collision occurs, and therefore the longer the channel access delay and message response time. The same reason is applicable in the mean queue length in Figure 7.8 and the message blocking probability in Figure 7.9.

We also studied R-ALOHA systems where message sizes have the same mean values but different distributions. Three cases are considered where

$$G(z) = 0.2z^{16} + 0.6z^{11} + 0.2z, \quad (7.31)$$

$$G(z) = z^{10} \quad (7.32)$$

and

$$G(z) = \frac{0.1z}{1 - 0.9z}. \quad (7.33)$$

That is, the mean message size is 10 packets in each case but message size distributions are general, deterministic and geometric, respectively. It is found that the difference of the

Figure 7.6. Channel throughput

Figure 7.7. Mean message response time

Figure 7.8. Mean queue length

Figure 7.9. Message blocking probability

system performances in the terms of channel throughput, mean message response time, mean queue length and message blocking probability is too small to be noticed. Going through the analysis in section 3, it is found that b_0 , p_a and p_s can be determined by (7.10)-(7.23) if π_0 is given. When $L = 1$ the system is unbuffered and we always have $\pi_0 = 1$. Therefore the system has the same performance as indicated by (7.8), (7.9) and (7.24)-(7.27). The numerical results also show that the probability π_0 for the three cases is almost same when $L \geq 2$, so are b_0 , p_a and p_s . This implies that $C(z)$ is almost same. When the offered load is sufficiently large, the moment of each order derived from $C(z)$ is much larger than that from $G_s(z)$. Hence, $C(z)$ plays the key role in $B(z)$ so that the role of $G(z)$ is ignorable in (7.3). The queueing behavior of users is determined mainly by $C(z)$. That is why the system performance is almost same in the three cases under heavy traffic conditions. If the offered load is very small, $C(z)$ does not dominate the behavior of $B(z)$. However, the buffered system has almost the same performance as unbuffered system when the offered load is sufficiently small. It is known that the unbuffered system has the same performance. Hence, the system performance is almost same under light traffic condition.

Based on the numerical results, we conclude that the influence of the message size distribution on the system performance is very weak in R-ALOHA systems as long as the mean message size is same. This conclusion is very useful since we can always study a system with a generally distributed message size by treating it as the message size is geo-

metrically distributed with the same mean value and using the memoryless characteristics of geometric distribution without losing generality, if only mean value analysis is interested.

The diagrams in Figure 7.10 and Figure 7.11 give the mean queue lengths and message blocking probabilities for the above three message size distribution given $\lambda = 0.0005$ and $L = 5$. Note the channel throughput and the mean message response time can be derived from the mean queue length and message blocking probability. As indicated in the figures, the performance difference is very small in the three cases. The figures also show that the higher the second order moment of the message size, the worse the system performance. This is true intuitively since a Geo/G/1 system has this property, and a finite buffer system behaves just like an infinite buffer system if the message blocking probability is very small.

Figure 7.10. Mean queue length

Figure 7.11. Message blocking probability

7.7 Summary

In this chapter, TUA is used to analyze the R-ALOHA systems with propagation delay, finite buffer capacity and finite user population. The channel throughput, mean waiting time and response time are obtained. The system model is much more general than that employed in EPA analysis [1][34] since generally distributed message size and finite user buffer capacity are permitted. The accuracy of TUA on R-ALOHA is verified by the simulations.

The analysis shows that it is not the distribution but the mean message size which plays the key role in determining the mean value analysis of the performance.

It is noticed that the available analysis on PRMA does not consider the case where a data user is allowed to reserve a slot if it has multiple packets available for transmission. Though it is claimed that the analysis can be done by treating this kind of data users as voice

users in the analysis, it is not true since a data user in this case normally has a buffer capacity of more than one message, and the number of packets contained in each message is generally distributed. Since a data user in PRMA system operates in R-ALOHA protocol if it is allowed to reserve a slot, the idea used in this chapter can be extended to solve this problem.

Though the analysis is focused on the users with finite buffer capacity, it can be extended to the infinite buffer case straightforwardly where each user is treated as a Geo/G/1 queueing system.

8 PACKET RESERVATION MULTIPLE ACCESS (PRMA)

8.1 Introduction

In this chapter, we apply TUA to the performance analysis of Packet reservation multiple access (PRMA) protocol [96]. PRMA is a wireless access protocol for the integration of packetized voice and data over a common communication channel. It has received wide attention because of its bandwidth efficiency and the simplicity of control. The bandwidth efficiency is obtained by granting exclusive use of time slots to voice terminals with reservations, while permitting other voice users and data users to contend for unreserved slots. The simplicity of control is guaranteed since PRMA employs a contention scheme like R-ALOHA.

It is well known that speech signals consist of talkspurts and silence periods. In order to utilize the channel efficiently, only talkspurts are packetized into multiple packets for transmission. A voice terminal releases the channel after the current talkspurt is over, and tries to regain the channel when a new talkspurt starts. For this purpose, a speech activity detector is used at voice terminal to detect the talkspurts. The speech consists of principal spurts and gaps as well as minispurts and minigaps. The former occurs as a result of patterns in talking, pausing, while the latter is due to short silent intervals that punctuate a con-

tinuous speech. Based on the response of the speech activity detector, “fast” and “slow” speech activity detectors are defined [37]. The “slow” speech activity detector can respond only to the principal talkspurts and gaps. However, the “fast” speech activity detector can respond to the principal as well as mini-talkspurts and gaps.

Since 1989, when PRMA was proposed [36], many publications reported performance evaluation of PRMA by simulation [36][37][38] or theoretical analysis [39]–[41], [46][47]. Among them, some focus on the voice-only PRMA system, while others concentrate on PRMA systems integrating voice and data. In the analysis of voice data integration system, it is assumed in [46] that the data users have a buffer capacity of only one packet, while in [38][40][47] infinite buffer capacity of data users is assumed. In either case, these assumptions were made to simplify the system analysis. However, in a practical situation, a data user has a finite buffer capacity and therefore should be modeled as a finite buffer queueing system.

Obviously, PRMA system is a heterogeneous system consisting of voice users and data users. In TUA, each data user is treated as a queueing system with finite buffer capacity. Each voice user is treated as a unbuffered system, or a queueing system with the capacity of one talkspurt. It is assumed that equilibrium probabilities exist for each user and each user always operate with its equilibrium probabilities, which are determined by the behaviors of all users in the system because of common channel sharing, however.

Since both data users and voice users are queueing systems with finite queue capacity, the Markov chain of the states of data user and the imbedded Markov chain of the states of voice user are always ergodic [84]. Therefore, the stability of each subsystem and the exist-

ence of its equilibrium probabilities are always guaranteed. As the result, the stability of the overall system is also guaranteed. TUA can also be applied to PRMA systems where data users have infinite buffer capacity. However, for the system with infinite buffer capacity, the system stability and the existence of equilibrium probabilities are guaranteed only if the utilization factor ρ is bounded by some constant less than 1 [84]. As long as the system is stable and equilibrium probabilities exist, TUA can be used to analyze the system performance. In fact, when the system is unstable and the equilibrium probabilities do not exist, TUA either has no solution (numerical algorithm does not converge) or has an abnormal solution where at least one state probability takes values less than 0 or greater than 1. This has also been observed before when TUA is applied to the analysis of S-ALOHA, CSMA/CD and R-ALOHA. That is, TUA can also be applied to determine the stability of PRMA systems.

8.2 PRMA System Model

PRMA is a random access protocol with potential application in the third generation wireless communications systems. In PRMA, the time is divided into slots which are grouped into frames. Each frame consists of a fixed number M of time slots. The slot status can be either reserved or available. The slot status information is tracked and broadcast to all users by base station over a downlink channel. Voice users and data users having packetized information available for transmission contend for access to an available slot in the shared uplink channel. The channel is assumed to be noise free, and collision is the only reason for unsuccessful transmission. A collision occurs when more than one user simultaneously starts transmission in an available slot. If a voice user issues a successful transmis-

sion in a slot, this slot is reserved and will be used by this voice user in subsequent frames until the talkspurt is over. Following this, the slot is released and becomes available again to all users. Unlike voice users, a data user cannot reserve a slot even if it has more than one packet in its buffer. Each data packet has to experience an independent contention procedure in order to be transmitted. Here we ignore the case that a data user can reserve a slot if it has a multiple packet message for transmission such as in FTP (file transfer protocol) application.

We consider a PRMA system which has M_d data users and M_v voice users. Voice users and data users are assigned with permission probabilities p_v and p_d , respectively. Here we consider a speech model corresponding to a “slow” speech activity detector [37] which responds only to principal talkspurts and gaps. The number of packets within each talkspurt is geometrically distributed with the mean of $1/\gamma$ packets. The number of slots in each silence period is geometrically distributed with the mean of $1/\lambda_v$ slots. Since speech requires prompt packet delivery, packets delayed beyond D slots are considered old and are dropped by the system.

Each data user has a buffer capacity of L packets. Packets arrive in the buffer of a data user following a Bernoulli process: a packet arrives in a slot with probability λ_d and no packet arrives in a slot with probability $1 - \lambda_d$. Each packet arrival occurs at the end of a slot, and each busy data user contends for transmissions at the beginning of a slot. The queue in each data buffer is FCFS. The late arrival model with delayed replacement rule is used. That is, an arriving packet finding the queue full will get dropped even if there is a

packet departing from the queue at the same slot. Moreover, IFT principle is used for voice users since voice packets require prompt delivery, and DFT principle is employed for data users for analytical convenience.

As mentioned earlier, each voice terminal is also a queueing system with a buffer capacity¹ of one talkspurt or message. However, it is not a general queueing system because a talkspurt may be dropped if all the packets belonging to it are delayed for more than D slots

8.3 System Analysis

In this section, we focus our analysis on data users firstly, then voice users and finally the overall system.

8.3.1 Analysis of Data Users

The analysis on the virtual packet service time is carried out using the state flow graph [76] of a tagged data user as shown in Figure 8.1, where a busy data user can be in one of the four states, the packet ready state S_0 , the transmission state S_{tx} , the backoff state S_{bk} and the departure state S_e .

Whenever a packet is ready for service, the station switches to S_0 . Since the DFT model is applicable here, a station entering S_0 will transit to S_{bk} immediately. For a station in state S_{bk} , it will start transmission and switch to S_{tx} with probability $p_d p_{ad}$, where p_{ad} is

1. In fact, a buffer capacity of $\lfloor D/N \rfloor + 1$ packets is sufficient for voice users, where $\lfloor x \rfloor$ is the largest integer not greater than x . However, here we treat a talkspurt as a multiple packet message which is generated at one time without losing generality since we consider a message, not a packet, as a customer in the queueing system.

Figure 8.1. State flow graph of tagged data user in PRMA systems

the probability that a data user finds a slot available in each time slot, and p_d is the permission probability of data user as defined before. Otherwise, it will stay in S_{bk} for one more slot with probability $1 - p_d p_{ad}$. Hence, the transfer function from S_{bk} to itself is $(1 - p_d p_{ad})z$. When a station starts transmission, it enters S_{tx} immediately. Since the channel is shared by all stations in the system, a successful transmission cannot be guaranteed. Denoting by p_{sd} the probability of a successful transmission by a data user given that it transmits in a slot, the state transfer function from S_{tx} to S_e is given by $p_{sd}z$ since the transmission time of each data packet is one slot. If a transmission is not successful, which occurs with probability $1 - p_{sd}$, the station will backoff after detecting the transmission failure and transit to S_{bk} . Here we assume that immediate acknowledgment is possible since PRMA is supposed to be used in local wireless access systems where the round trip

propagation time between terminals and base stations is on the order of a few tens of microseconds outdoors, and less than one microsecond indoors [36]. S_e is called departure state since a station transits to this state from S_{tx} at the moment of a packet departure. If there are other packets waiting in the queue, a new packet will become ready for service after the previous one departs and the station will transit from S_e to S_0 in no time. Otherwise, the station becomes idle.

Note that states defined in Figure 8.1 should not be confused with the states of the queueing system, which is the number of packets in the buffer of the data user. The only purpose of Figure 8.1 is to assist us to find the PGF of packet service time.

By definition, the packet service time is the time duration between a data user entering S_0 to its exiting S_e . By [76], the PGF $B(z)$ of the packet service time is equal to the transfer function from node S_0 to node S_e . Using Mason's formula [82], $B(z)$ is given by

$$B(z) = \frac{p_d p_{ad} p_{sd} z}{1 - (1 - p_d p_{ad} p_{sd}) z}, \quad (8.1)$$

which implies that the packet service time is geometrically distributed with rate

$$\mu = p_d p_{ad} p_{sd}. \quad (8.2)$$

Hence, each data user is actually a Geo/Geo/1/K queueing system in Kendall's notation.

Following the analysis on Geo/Geo/1/K system in Appendix B.3, we have

$$p_0 = \left(\frac{\mu}{\mu - \lambda_d} - \frac{\lambda_d^2}{\mu(\mu - \lambda_d)} \left[\frac{\lambda_d(1 - \mu)}{\mu(1 - \lambda_d)} \right]^{L-1} \right)^{-1} \quad (8.3)$$

and

$$\begin{aligned}
 p_i &= \left[\frac{\lambda_d(1-\mu)}{\mu(1-\lambda_d)} \right]^{i-1} \frac{\lambda_d}{\mu(1-\lambda_d)} p_0, \quad i \in [1, L-1] \\
 p_L &= \frac{\lambda_d}{\mu} \left[\frac{\lambda_d(1-\mu)}{\mu(1-\lambda_d)} \right]^{L-1} p_0.
 \end{aligned} \tag{8.4}$$

8.3.2 Analysis of Voice Users

The analysis on voice users is similar to the analysis of R-ALOHA in Chapter 7. However, the voice user is an unbuffered system and packet dropping may occur.

The voice user can be in one of three modes: silence (SIL), contention (CON), and reservation (REV) as shown in Figure 8.2. A voice terminal is at SIL if it has no packets for transmission. It switches from SIL to CON if a new talkspurt is generated. The mode CON ends in different cases. If a whole talkspurt is dropped, the system switches from CON back to SIL immediately. Otherwise, the system switches from CON to REV immediately after the first packet of a talkspurt is transmitted successfully. The mode REV ends after the last packet of a talkspurt has departed. $S_c(z)$ and $G_v(z)$ are the transfer functions from SIL to CON and from REV to SIL, respectively. An active voice user would switch from CON to REV if it reserves a slot successfully, or it may switch back to SIL from CON directly if all the packets of a talkspurt have access delay more than D slots. In this case, the whole talkspurt is dropped. A talkspurt is said to be dropped if all of its packets are dropped when maximum permitted delay is exceeded. A talkspurt is not dropped if at least one of its packets is transmitted successfully. The transfer functions from CON to REV and CON to SIL are denoted by $C_r(z)$ and $C_s(z)$, respectively.

Figure 8.2. Voice terminal activity model

To determine the transfer functions in Figure 8.2, let us consider the behavior of the tagged voice user given that no talkspurt can be dropped, that is, at least one packet of the talk spurt is transmitted successfully. The analysis is carried out using the state flow graph [76] of an active voice user shown in Figure 8.3, where state V_0 is the start of a talkspurt, V_{bk} means that the user is in backoff state of contention mode, V_{tx} is the state when the user contends for an available slot, and V_c is the end of contention when the tagged user reserves a slot, V_r implies that the user is in REV mode, and V_e is the end of a talkspurt and the start of a silence period.

Similar to the analysis on data user, using Figure 8.3, the PGF $C(z)$ of the time duration of a voice user spending at CON mode is given by

$$C(z) = \frac{(1 - z + p_v z) p_{av} p_{sv} z}{1 - (1 - p_v p_{av} p_{sv}) z}, \quad (8.5)$$

Figure 8.3. State flow graph of voice user when talkspurt is not dropped

where p_{av} is the probability that an active voice user finds a slot to be available, and p_{sv} is the probability that a voice user transmits its packet successfully given that it transmits in an available slot. Similarly the PGF $G_v(z)$ of the time duration at mode REV is given by

$$G_v(z) = \frac{\gamma}{1 - (1 - \gamma)z^M}. \quad (8.6)$$

Denoting by t_d the voice terminal channel access delay in slots, which is just one slot less than the time spent by a voice terminal at CON, from (8.5) we have

$$P(t_d = n) = \begin{cases} \beta, & n = 0 \\ \beta p_v (1 - \beta) (1 - \beta p_v)^{n-1}, & n \neq 0 \end{cases} \quad (8.7)$$

where

$$\beta = p_{av} p_{sv}. \quad (8.8)$$

8.3.2.1 Packet dropping probability

Let n_{drop} and l_v be the number of dropped packets and the total number of packets in a talkspurt, respectively. In a talk spurt, all packets experiencing delay more than D slots must be dropped. Thus, from (8.9) we have for $1 \leq k \leq n-1$

$$\begin{aligned} P(n_{drop} = k | l_v = n) &= \sum_{i=D+(k-1)M+1}^{D+kM} P(t_d = i) \\ &= (1-\beta)(1-\beta p_v)^{D+(k-1)M} [1 - (1-\beta p_v)^M], \end{aligned} \quad (8.9)$$

and for $k = n$

$$\begin{aligned} P(n_{drop} = k | l_v = n) &= \sum_{i=D+(k-1)M+1}^{\infty} P(t_d = i) \\ &= (1-\beta)(1-\beta p_v)^{D+(k-1)M}. \end{aligned} \quad (8.10)$$

Hence, we obtain for $k \geq 1$

$$\begin{aligned} P(n_{drop} = k) &= \sum_{n=k}^{\infty} P(n_{drop} = k | l_v = n) P(l_v = n) \\ &= (1-\beta)(1-\beta p_v)^{D+(k-1)M} (1-\gamma)^{k-1} [1 - (1-\gamma)(1-\beta p_v)^M] \end{aligned} \quad (8.11)$$

and the mean number of packets dropped from a talkspurt is given by

$$E(n_{drop}) = \sum_{k=1}^{\infty} k P(n_{drop} = k) = \frac{(1-\beta)(1-\beta p_v)^D}{1 - (1-\gamma)(1-\beta p_v)^M}. \quad (8.12)$$

Finally, the packet dropping probability p_{drop} is obtained as

$$p_{drop} = \frac{E(n_{drop})}{E(l_v)} = \frac{\gamma(1-\beta)(1-\beta p_v)^D}{1 - (1-\gamma)(1-\beta p_v)^M}, \quad (8.13)$$

where the mean number $E(l_v)$ of packets in a talkspurt is $1/\gamma$ as known before.

Define

$$\begin{aligned}
 c_k' &\equiv P(\text{a talkspurt consists of } k \text{ packets and is dropped}) \\
 &= P(n_{drop} = k | l_v = k) P(l_v = k) \\
 &= (1 - \beta)(1 - \beta p_v)^{D + (k-1)M} (1 - \gamma)^{k-1} \gamma, \quad k \geq 1.
 \end{aligned} \tag{8.14}$$

Since c_k' is also the probability that a voice terminal stays in CON for $D + (k-1)M$ slots before it switches back to SIL directly, we have

$$\begin{aligned}
 C_s(z) &= \sum_{k=1}^{\infty} c_k' z^{D + (k-1)M + 1} \\
 &= \frac{\gamma(1 - \beta)(1 - \beta p_v)^D z^{D+1}}{1 - (1 - \gamma)(1 - \beta p_v)^M z^M}.
 \end{aligned} \tag{8.15}$$

Let c_k'' be the probability that an active voice terminal reserves a slot after an access delay of k slot, we have

$$\begin{aligned}
 c_k'' &= P(t_d = k, \text{a talkspurt is not dropped}) \\
 &= \begin{cases} P(t_d = k, l_v \geq 1), & 0 \leq k \leq D \\ P(t_d = k, l_v \geq i), & D + (i-2)M + 1 \leq k \leq D + (i-1)M, i \geq 2. \end{cases}
 \end{aligned} \tag{8.16}$$

Since t_d and l_v are independent and l_v is geometrically distributed with mean $1/\gamma$, using (8.7) and (8.16) it is easy to obtain

$$c_k'' = \begin{cases} \beta, & k = 0 \\ \beta p_v(1 - \beta)(1 - \beta p_v)^{k-1}, & 1 \leq k \leq D \\ \beta p_v(1 - \beta)(1 - \beta p_v)^{k-1} (1 - \gamma)^{i-1}, & D + (i-2)M + 1 \leq k \leq D + (i-1)M, i \geq 2. \end{cases} \tag{8.17}$$

Since a voice terminal would reserve a slot within one slot after the access delay if the talkspurt is not dropped, we have

$$\begin{aligned}
 C_r(z) &= \sum_{k=0}^{\infty} c_k z^{k+1} \\
 &= \beta z + \frac{(1-\beta)\beta p_v z^2}{1-(1-\beta p_v)z} \left[1 - \frac{\gamma(1-\beta p_v)^D z^D}{1-(1-\gamma)(1-\beta p_v)^M z^M} \right].
 \end{aligned} \tag{8.18}$$

By total probability law, clearly we must have

$$C_r(1) + C_s(1) = 1, \tag{8.19}$$

which can be verified by (8.15) and (8.18). Since the silence duration is geometrically distributed, we also have

$$S_c(z) = \frac{\lambda_v z}{1-(1-\lambda_v)z}. \tag{8.20}$$

It should be mentioned that all prior studies employ a model where a voice terminal at CON would switch back to SIL directly with a given probability (equivalent to γ/M here) at each time slot. Although this assumption is made for analytical convenience, it does not match the packet dropping scheme in a PRMA system. The more proper analysis is the one given here, we believe, for the first time.

8.3.2.2 Talkspurt dropping probability

A talkspurt is said to be dropped if all the packets contained in the talkspurt are dropped because of maximum permitted delay of D slots. A talkspurt is not dropped if at least one of its packets is transmitted successfully. The talkspurt dropping probability q_{drop} is the

probability that a voice terminal switches from CON back to SIL without passing through REV. Using (8.15), we have

$$q_{drop} = C_s(1) = \frac{\gamma(1-\beta)(1-\beta p_v)^D}{1-(1-\gamma)(1-\beta p_v)^N}. \quad (8.21)$$

Compared (8.13) and (8.21), it is found that $q_{drop} = p_{drop}$. From now on, we do not distinguish the packet dropping probability from the talkspurt dropping probability since they are equal. It should be clear that the relation $q_{drop} = p_{drop}$ holds only if the number of packets in a talkspurt is geometrically distributed. If the number of packets of a talkspurt is not geometrically distributed, we have to distinguish one from the other.

8.3.3 Analysis of integrated systems

In PRMA, both data users and voice users share the same channel. Their behaviors are affected by each other. Let p_{rev} be the probability that a slot is reserved by a voice user at any frame, and $p_{c|\bar{R}}$ be the probability that a voice user is in CON given that it is not in REV. Also, we define the following events:

$$\begin{aligned} A_1 &= \{ \text{tagged data user is transmitting} \} \\ A_2 &= \{ \text{the present slot is available} \} \\ S &= \{ \text{the transmission of tagged data user is successful} \} \\ R_i &= \{ i \text{ slots are reserved at present frame} \}. \end{aligned}$$

Since the IFT model is applicable in voice users, a voice user at SIL would contend at an available slot with probability 1 if there is a talkspurt generated in the voice user imme-

diately before this slot, which occurs with probability λ_v . A voice user would be at SIL with probability $1 - p_{c|\bar{R}}$ given that it is not at REV. Hence, we have

$$\begin{aligned}
 p_{sd} &= \text{Prob}(S|A_1, A_2) \\
 &= \sum_{i=0}^{M-1} \text{Prob}(S|A_1, A_2, R_i) \cdot \text{Prob}(R_i|A_1, A_2) \\
 &= \sum_{i=0}^{M-1} (1 - p_b p_d)^{M_d-1} [1 - p_{c|\bar{R}} p_v - (1 - p_{c|\bar{R}}) \lambda_v]^{M_v-i} \binom{M-1}{i} p_{rev}^i (1 - p_{rev})^{M-1-i} \\
 &= (1 - p_b p_d)^{M_d-1} [1 - p_{c|\bar{R}} p_v - (1 - p_{c|\bar{R}}) \lambda_v]^{M_v} \left[\frac{p_{rev}}{1 - p_{c|\bar{R}} p_v - (1 - p_{c|\bar{R}}) \lambda_v} + 1 - p_{rev} \right]^{M-1},
 \end{aligned} \tag{8.22}$$

where the summation ranges from 0 to $M-1$ since at most $M-1$ slots can be reserved given that a slot is available. The data user busy probability p_b is given by

$$p_b = 1 - p_0. \tag{8.23}$$

Similarly, we have¹

$$\begin{aligned}
 p_{sv} &= (1 - p_b p_d)^{M_d} [1 - p_{c|\bar{R}} p_v - (1 - p_{c|\bar{R}}) \lambda_v]^{M_v-1} \\
 &\quad \left[\frac{p_{rev}}{1 - p_{c|\bar{R}} p_v - (1 - p_{c|\bar{R}}) \lambda_v} + 1 - p_{rev} \right]^{M-1}.
 \end{aligned} \tag{8.24}$$

Now, we treat the evolution of time for each voice user as a succession of cycles which are statistically equivalent. Each cycle is a duration from a voice terminal entering SIL to its leaving REV. In each cycle, a voice terminal can switch back and forth between CON and

1. Equations (8.22) and (8.24) are obtained under the assumption of $M_v > M$, which is implied by PRMA systems. Otherwise, it is not necessary to use PRMA since multiplexing TDMA (time division multiple access) would have better performance.

SIL more than once, but it can visit REV only once. From Figure 8.2, the PGF $\phi(z)$ of the cycle length is given by

$$\phi(z) = \frac{S_c(z)C_r(z)G_v(z)}{1 - S_c(z)C_s(z)}. \quad (8.25)$$

Therefore, the mean duration of each cycle is given by

$$\phi'(1) = \frac{S_c'(1)}{C_r(1)} + \frac{C_r'(1) + C_s'(1)}{C_r(1)} + G_v'(1). \quad (8.26)$$

Denoting by T_{SIL} , T_{CON} , and T_{REV} the average time which a voice terminal spends at SIL, CON and REV, respectively, in each cycle, from (8.26), (8.19) and (8.21) we have

$$T_{SIL} = \frac{S_c'(1)}{C_r(1)} = \frac{1}{\lambda_v(1 - q_{drop})}, \quad (8.27)$$

$$T_{REV} = G_v'(1) = M\left(\frac{1}{\gamma} - 1\right), \quad (8.28)$$

and

$$T_{CON} = \frac{C_r'(1) + C_s'(1)}{C_r(1)}. \quad (8.29)$$

Since the average number of reserved slots in any frame should be equal to the average number of voice users in REV mode, we also have

$$M_v p_r = M p_{rev}, \quad (8.30)$$

where p_r is the probability that a voice user is at REV mode, which is given by

$$p_r = \frac{T_{REV}}{T_{SIL} + T_{CON} + T_{REV}}. \quad (8.31)$$

As to $p_{c|\bar{R}}$, we have

$$p_{c|\bar{R}} = \frac{T_{CON}}{T_{SIL} + T_{CON}}. \quad (8.32)$$

It is clear that the probability of slot availability for data and voice users which are contending for channels are different. This is because a voice user in CON cannot be in REV at the same time. For example, in an extreme case where the system has only one voice user and several data users. The voice user will find that the slot is always available whenever it is in CON. However, the data users will find a slot available with a nonzero probability. For a data user, the average number of slots available in each frame is $M_v p_r$. Hence, it will find a slot available with probability

$$p_{ad} = 1 - \frac{M_v p_r}{M} = 1 - p_{rev}. \quad (8.33)$$

However, since a voice user contending for transmission does not reserve a slot, it will find a slot available with probability

$$p_{av} = 1 - \frac{(M_v - 1)p_r}{M}. \quad (8.34)$$

The equations obtained above are sufficient to solve all the unknown parameters. An iterative algorithm will be developed in Section 8.5 for solving the equations. After all the unknown parameters are obtained, they can be applied to the performance evaluation of PRMA system.

8.4 Performance Evaluation

The performance requirements for the two classes of user, voice and data, are different. In this section, the performance indices of interest are derived using the foregoing results.

8.4.1 Data Users

Our interest is in packet blocking probability, mean system response time, mean queue length and throughput of each data user. The packet blocking probability p_B is the probability that an arriving packet finds the buffer full. Hence, we have

$$p_B = p_L. \quad (8.35)$$

The mean queue length is given by

$$E(l) = \sum_{k=0}^L k p_k. \quad (8.36)$$

The system response time is the duration from a packet arrival to its departure. By Little's law, the mean system response time $\bar{R}_d$ is given by

$$\bar{R}_d = \frac{\bar{L}_q}{\lambda_d(1 - p_B)}. \quad (8.37)$$

The average number of packets transmitted in each time slot defines the throughput. The throughput of each data user is given by

$$\theta_d = \lambda_d(1 - p_B). \quad (8.38)$$

Since the queueing system of each data user is solved, other performance measurements, such as the distribution of waiting time, system response time, and packet interdeparture time of each data user can also be obtained, if required, in a straightforward manner.

8.4.2 Voice Users

In this case, we are interested in the packet dropping probability or packet dropping rate, average access delay, and throughput of each user. From Figure 8.2, an active voice terminal spends on the average $C_s'(1)$ slots in CON each time if a whole talkspurt is dropped, otherwise, it will reserve a slot after spending an average time of $C_r'(1)$ slots in CON. The access delay makes sense only if the talkspurt is not dropped. Hence, the average access delay, denoted by $\bar{T}_v$, is given by

$$\bar{T}_v = E(t_d | \text{talkspurt is not dropped}) = C_r'(1) - 1. \quad (8.39)$$

The throughput of each voice user is given by

$$\theta_v = \frac{1/\gamma}{T_{SIL} + T_{CON} + T_{REV}} \quad (8.40)$$

since $1/\gamma$ packets on the average are transmitted in each cycle no matter how many packets of the talkspurt are dropped as long as the whole talkspurt is not dropped. This is due to the memoryless property of the geometrical distribution. The packet dropping probability is already given in (8.13).

8.4.3 Channel Throughput

Another interesting performance index is the channel throughput, which is just the summation of throughputs of all users. It is given by

$$\Theta = M_d \theta_d + M_v \theta_v. \quad (8.41)$$

8.5 Numerical Algorithm

So far, we have sufficient number of nonlinear equations, to determine all the unknowns in the system. For this purpose an iterative algorithm is developed which operates as follows: (i) select a very small positive number, say $\delta = 1.0e-6$ and $n = 0$, and set the initial values for p_0 , p_r and $p_{c|\bar{R}}$ to any value between zero and one, say $p_0^{(n)} = 0.1$, $p_r^{(n)} = 0.3$ and $p_{c|\bar{R}}^{(n)} = 0.3$; (ii) compute p_b and p_{rev} using (8.23) and (8.30); (iii) compute p_{sd} and p_{sv} using (8.22) and (8.24), respectively; (iv) compute p_{ad} and p_{av} using (8.33) and (8.34), and p_{drop} using (8.8) and (8.13); (v) compute T_{SIL} , T_{REV} and T_{CON} using (8.27), (8.28) and (8.29), respectively; (vi) compute μ using (8.2) and update p_0 to obtain $p_0^{(n+1)}$ using (8.3); (vii) update p_r and $p_{c|\bar{R}}$ to obtain $p_r^{(n+1)}$ and $p_{c|\bar{R}}^{(n+1)} = 0.3$ using (8.30) and (8.31); (viii) if both conditions $|p_0^{(n+1)} - p_0^{(n)}|/p_0^{(n)} \leq \delta$ and $|p_{c|\bar{R}}^{(n+1)} - p_{c|\bar{R}}^{(n)}|/p_{c|\bar{R}}^{(n)} \leq \delta$ are satisfied, the system is solved, otherwise, repeat steps from (i) to (viii).

Since it is found that the convergence speed of p_0 and $p_{c|\bar{R}}$ may be different, the convergence of both p_0 and $p_{c|\bar{R}}$ are checked in step (viii). However, it is also found that other parameters converge at least as fast as the slower of p_0 and $p_{c|\bar{R}}$. Hence, it is not necessary to check convergence of every parameter.

8.6 Numerical Results

In this section, we apply TUA to the analysis of a PRMA system with fixed system parameters [37][47] shown in Table 8.1 and variable parameters p_v , M_v and M_d .

Table 8.1. PRMA system parameters

Variable	Nominal	Symbol in Text
Mean talkspurt	1.0 sec	$\gamma = 0.016$
Mean silence duration	1.35 sec	$\lambda_v = 0.000593$
Channel rate	720 kbps/sec	
Speech rate	32 kbps	
Packet size (speech or data)	576 bits	
Slot duration	0.8 msec	
Frame duration	16 msec	
Slots per frame	20	$M = 20$
Data rate	1.2 kbps	$\lambda_d = 0.001667$
Data user buffer capacity	50 packets	$L = 50$
Data permission probability	0.044	$p_d = 0.044$
Maximum voice delay	40 slots	$D = 40$

As discussed in Chapter 4, two equilibrium points exist in homogeneous S-ALOHA systems if they are bi-stable. When a simulation or theoretical analysis starts with all user busy, the system equilibrium point converges to the worse one. Otherwise, it converges to the better one. The same phenomenon is observed here. However, PRMA is a heterogeneous systems. It is found that there are multiple equilibrium points for PRMA system, as shown in Figure 8.4 for $p_v = 0.4$, where p_{ini} is the initial value for $p_{c|\bar{R}}$ using the numerical algorithm developed above. It is found that the system operates at its worst or best equi-

librium point corresponding to the curve ABEGMS or ADGLRS with $p_{ini} = 0.99$ or $p_{ini} = 0.01$, respectively. Compare to $p_{c|\bar{R}}$, the influence of the initial value of p_0 on the system equilibrium point is much weaker since the probability p_d is much less than the probability p_v . In fact, the initial value of p_0 has effect on system equilibrium point only in the case of $p_v = 0.1$. Here we fix the initial value of p_0 at $p_0^0 = 0.01$. When p_{ini} is located between 0.01 and 0.99, The value of p_{drop} is located between the worst and the best values. For instance, for $p_{ini} = 0.1$ and $p_{ini} = 0.23$, the values of p_{drop} follow the curves ACFLRS and ABEJPS, respectively. It is observed that the curve has two jumps. If the first jump occurs at points C or D, the second jump must occur at point L. Otherwise, the second jump may occur at any point between points G and L. It should be noticed that the multiple equilibrium points exist only in the region of $36 \leq M_v = M_d \leq 45$. There is only one equilibrium point for the voice users outside of this region.

Another interested issue is that the voice packet dropping probability experiences two jumps with the increase of M_v and M_d , as shown in Figure 8.4 and Figure 8.6. This phenomenon is also observed in Figure 8.8 and Figure 8.9. To understand this phenomenon, let us consider the curve ABEGMS in Figure 8.4. Given the same system parameters as those associated with the curve ABEGMS in Figure 8.4, some important probabilities are given in Figure 8.5, where p_{vb} is the probability that a voice user is busy, and p_{con} is the probability that a voice user is contending for a slot at any time, which are given by

Figure 8.4. Voice packet dropping probabilities given different initial conditions

Figure 8.5. Some important equilibrium probabilities

$$p_{vb} = \frac{T_{CON} + T_{REV}}{T_{SIL} + T_{CON} + T_{REV}} \quad (8.42)$$

and

$$p_{con} = \frac{T_{CON}}{T_{SIL} + T_{CON} + T_{REV}}. \quad (8.43)$$

At point $M_v = M_d = 36$, it is found that the data user busy probability p_b has a big jump. Since each data user would contend in an available slot with probability $p_b p_d$, the contention probability of each data user increases around 4 times from $M_v = M_d = 36$ to $M_v = M_d = 37$ though only two more users, one voice user and one data user, are added into the system. It is this jump which results in the first jump in curve ABEGMS in Figure 8.4. However, this jump does not affect the channel too much since p_d is much less than p_v . The probability p_{rev} is still very high. With the further increase of users, the second jump occurs when $M_v = M_d = 39$. At this time, it found that the probability p_{con} has a big jump, which results in the contention probability of voice user increasing more than 10 times. It is the second jump which degrades the system destructively. The probability p_{rev} becomes very small. Note that the buffer capacity of data user is 50 packets. If the probability p_{block} is sufficiently large, say 0.1, the probability p_b is almost 1. It is found that the busy probability p_{vb} of voice user is almost a constant. This is because a voice packet has to leave the buffer within D slots, either departed due to a successful transmission or dropped due to the maximum holding time being exceeded. Let us focus on the last packet of a talkspurt. If it is dropped, the busy duration of the voice user is the talkspurt length plus D slots. Otherwise, the busy duration length is less. However, the difference of

busy duration length between these two cases is less than D slots, which is much less than the average talkspurt length of 1 second. In fact, p_{vb} is increasing, though very slowly, with the increase of user population. Though p_{vb} is almost a constant, there is a large difference for the values of p_{con} on the two sides of the second jump. On the left hand side where $M_v = M_d \leq 39$, a busy voice user is almost always at mode REV, but on the right hand side where $M_v = M_d \geq 40$, a busy voice user is almost always at mode CON.

Note that not all of the curves would have two jumps. For instance, there is only one jump in the curve with $p_v = 0.5$ in Figure 8.6. This curve is obtained with $p_{ini} = 0.99$. When p_{ini} is sufficiently small, two jumps are observed for $p_v = 0.5$. The explanation is that the first jump is sufficient to degrade the system destructively. As to the curves with $p_v \leq 0.3$, the second jump may occur beyond the regions.

In the following, all numerical results for different p_v are obtained with the initial values of $p_{c|\bar{R}}$ and p_0 set to 0.99 and 0.01, respectively. In this case, the analysis starts with all voice and data users are contending for channel, and the system would be found to operate in its worst equilibrium point. The worst case is considered in order to obtain a conservative estimation on the system performance.

The voice packet dropping probabilities for different set of values of p_v , M_v and M_d are given in Figure 8.6. It shows that at an acceptable voice packet dropping probability of 1%, 37 simultaneous conversations can be carried over a single channel with p_v between

Figure 8.6. Voice packet dropping probability

Figure 8.7. Data packet blocking probability

0.2 and 0.3. A smaller p_v would result in long access delay as shown in Figure 8.9, and therefore higher packet dropping probability. When $p_v = 0.1$, at most 33 voice users can be served at the same time. Although large permission probability, say $p_v = 0.5$ would result in very small packet dropping probability for voice user population less than 33, the packet dropping probability increases suddenly toward 1 when the voice user population exceeds 34. Therefore, the ideal value for p_v is located between 0.2 and 0.3.

The performance indices of data terminals are given in Figure 8.7 and Figure 8.11. It is shown that the packet blocking probability is less than 10^{-4} for $M_v = M_d \leq 37$ given $p_d = 0.044$ and $p_v \in [0.2, 0.3]$. In this case, almost all arriving data packets are sent out successfully. As shown in Figure 8.7 and Figure 8.11, although $p_v \leq 0.2$ would degrade the performance of voice users, it benefits the data users since p_d is fixed. This fits well with intuitive reasoning.

The delays for data users and voice users are given in Figure 8.8 and Figure 8.9, respectively. The mean system response time of data packets can increase without limit since a data packet in the buffer has to be transmitted no matter how long it has to wait. However, it is found that the mean talkspurt access delay is bounded by around 1240 slots. This is because a whole talkspurt is dropped if all packets in this talkspurt experience access delay more than D slots. An average talkspurt has $1/\gamma = 62.5$ packets. If the talkspurt is not dropped, at least one of its packets must be transmitted successfully. That is, its last packet must be transmitted out in $40 + (62.5 - 1) \times N = 1270$ slots from the time that first packet

Figure 8.8. Mean Data packet response time

Figure 8.9. Mean talkspurt access delay

Figure 8.10. Total throughput in the channel

was generated. Therefore, we expect that the mean talkspurt access delay is bounded by some value around 1270 slots. Again, Figure 8.8 confirms that p_v between 0.2 and 0.3 is a good range for data users since the corresponding packet response time is less than 200 slots.

The total channel throughput contributed by all voice and data users is given in Figure 8.10, and the throughput contributed by all of the data users is given in Figure 8.11. Finally the mean queue length of data user is given in Figure 8.12. The jump in Figure 8.12 is resulted in by the same reason as in Figure 8.6. It is clear that the channel collision probability increases with the increase of the number of users in the system. On the left hand side of the jump point where the user population is considered to be relatively small, the packet service time is much less than packet interarrival time. Hence, most of time a data user has at most one packet in its buffer. However, the increased user population leads to

Figure 8.11. Total throughput of data terminals

Figure 8.12. Mean queueing length of data terminals

excessive collisions and increased congestion, which result in more users at contention state and more collisions. It is this positive feedback effect which leads to the jump. On the right side of the jump point where the user population is sufficiently large, the mean packet service time may exceed the mean packet interarrival time. In this case, the data user buffer is almost always full and the data packet blocking probability is large.

8.7 Summary

In this chapter, TUA is used to evaluate the performance of a PRMA system with both voice and data users. Voice packet dropping probability is analyzed in detail. The performance of PRMA systems obtained here seems to be more optimistic than that obtained in [39] and [40]. The possible reasons for the difference are: (a) a different system configuration from [39] and [40] is used; (b) In the system model IFT principle is employed for voice users and DFT for data users, while DFT is employed for both voice and data users in [39] and [40]; (c) keeping in mind the difference in the requirement for voice and data users, the slot availability probability for voice users is greater than that for data users here, but both voice and data users are supposed to have the same slot availability probability in [39] and [40]; (d) a more accurate model for voice packet dropping scheme is used here.

An unsolved problem is how to analyze the performance of a voice data integrated PRMA system where a data user is allowed to reserve a slot if it has more than one packet available for transmission. This problem is expected to be solved using TUA. For instance, if the data users operate using the R-ALOHA protocol considered in Chapter 7, the system can be analyzed by applying the analysis developed in this chapter to voice users and the analysis developed in Chapter 7 on data users.

9 CONCLUSIONS

9.1 Summary of This Research

This research is focused on the performance analysis of buffered random multiple access protocols. By reviewing the literatures, we found that almost all of the available approaches lead to very complicated mathematical analysis and cannot handle systems with complicated queueing behaviors due to the large size and high dimension of the state space required for describing the system. Moreover, they do not employ the existing queueing theory for the system analysis though each user is indeed a queueing system because of the difficulty for obtaining the packet service time distribution. In one word, there is no unified approach available for the performance analysis of buffered random multiple access protocols.

The objective of this research is to develop a unified approach which is able to use the classical queueing theory to the system analysis. TUA is presented to serve this purpose. The key for using queueing theory in system analysis is to find the packet service time distribution, which is determined by the behaviors of all users in multiple access protocols. In TUA, an efficient approach, called state flow graph, is introduced.

Generally speaking, TUA is a framework where the packet service time distribution is obtained using state flow graph, and then applied to the user queueing analysis using avail-

able results in classical queueing theory. The basic principle of TUA is discussed in Chapter 3.

In Chapter 4, a buffered S-ALOHA without channel propagation delay is treated. The system stability is studied when buffer capacity is infinite. In addition, a complete equilibrium point analysis is also presented here for S-ALOHA systems with finite buffer capacity to identify the range of permission probability so that the optimal system performance is obtained.

To demonstrate how to apply TUA to systems with channel propagation delay, a S-ALOHA system operating in satellite communications environment is studied in Chapter 5. Channel propagation delay introduces memory which would further complicate the system analysis. However, this problem is handled well by TUA. The analysis results are verified by simulations.

In order to show how to apply TUA to systems operating in a wireless environment, a CDMA S-ALOHA system is also studied in Chapter 5. It shows that the analysis on multiple access protocols is independent of the analysis on channel PER. If the channel PER is given, the performance analysis is straightforward. To further demonstrate the advantage that TUA is able to handle very complicated systems, we assume that multiple packet message arrival be permitted in this CDMA S-ALOHA system. That is, each user has a bulk arrival process. By developing the corresponding queueing theory, the system is analyzed using TUA.

In Chapter 6, we apply TUA to the analysis of buffered CSMA/CD systems. Several examples are given. Firstly, we study the influence of permission probability on system

performance given user population and packet arrival rate. As shown by numerical results, too small or large permission probability would degrade the system performance. Secondly, an optimal CSMA/CD system in term of channel throughput is studied. In this case, the permission probability is not predetermined, but adapted to the system status. Thirdly, a CSMA system, a special case of CSMA/CD system, is studied. Finally, we analyze a heterogeneous CSMA/CD systems consisting of two groups of users.

In Chapter 7, buffered R-ALOHA systems are studied using TUA where the message size may have arbitrary distribution in the integral number of slots. Since R-ALOHA is designed for satellite communications, long channel propagation delay exists. So far, EPA is the only available approach on R-ALOHA. EPA has been used to analyze R-ALOHA systems where each user can hold at most one message, and each message consists of a geometrically distributed number of packets. These systems handled by EPA are only special cases of the more general systems treated by TUA here.

In Chapter 8, a PRMA system is studied using TUA. PRMA system is a heterogeneous system consisting of voice users and data users which have different demands. A main problem associated with the available approaches is that the voice packet dropping scheme is not modeled properly. This problem is solved using TUA, for the first time. In addition, the available approaches cannot be applied to PRMA systems where data users are allowed to reserve a slot even if they have multiple packets ready for transmission. Though we do not analyze this system using TUA, we have discussed the possibility and concluded that TUA is able to handle this system.

9.2 Summary of Contributions

A unified approach called TUA is developed. TUA is able to handle systems with very complicated queueing behavior as long as the corresponding queueing theory is available. Moreover, TUA is simple since the analyses on channel contention scheme and the user queueing behavior are decoupled. The amount of numerical computation in TUA is also small since state transition probability matrix is not required.

In general, TUA has the following advantages:

- Packet service time distribution is easy to be obtained using state flow graph. The variable packet size, channel propagation delay and different transmission principle can be handled.
- The analyses on channel contention scheme and user queueing behavior are decoupled so that the results in classical queueing theory can be employed for the system analysis. Complicated queueing behaviors of users such as bulk arrival process and arbitrary service discipline can be handled as long as the corresponding queueing theory is available.
- A wide range of performance measurements can be obtained using TUA. In addition to mean values, the distribution of most measurements can also be obtained.
- TUA can be applied to non-ideal channel. No matter what is the channel condition, TUA can be used as long as the packet error rate (PER) is known.
- TUA is able to handle memory problem like propagation delay and message consisting of an arbitrarily distributed number of packets.
- TUA is able to analyze the stability of buffered random multiple access protocols.
- TUA can be applied to heterogeneous systems.
- The numerical algorithms of TUA can be developed easily.
- Calculation complexity of TUA is low since state transition probability matrix is not required for the system analysis.
- There is no numerical instability problem with TUA.
- The accuracy of TUA is good.

9.3 Suggestions for Future Research

In this thesis, TUA has been applied to the performance analysis of several typical protocols successfully. As a framework for the performance analysis of multiple access protocols, TUA does not and cannot cover the particular techniques for each protocols. It is these particular techniques which introduce extra errors into the analysis, in addition to the inherent errors of TUA by assuming that each user always operate with its equilibrium probabilities. Although the particular techniques developed in this thesis such as the method for finding the probability p_i in Chapter 3 work well, there may exist better ones. A challenging task is how to find them.

With the quick development of telecommunications industry, more and more contention based multiple access protocols are proposed for designing new systems. The most important problem is how to analyze the performance of these protocols using TUA.

In this thesis, all of the analysis work is focused on the single channel systems. However, multiple channels are employed in many multiple access systems [89]. How to apply TUA to multichannel multiple access protocols is also interesting.

In addition, all of the analysis work in this thesis is focused on the discrete-time systems. However, it is shown that state flow graph is applicable in continuous-time systems. It is therefore interesting to extend TUA to the performance analysis of continuous-time buffered random multiple access systems.

It is found that when a finite buffer system is bi-stable, TUA can analyze the system performance at each equilibrium point only, and cannot give the real performance of the

system. This is because the system occupancy probability in each equilibrium point can not be obtained using TUA. How to find the occupancy probability is a very interesting issue.

Regarding the principle of TUA, a few words should be mentioned about its mathematical grounds. TUA is an approximate approach, but there is no theoretical method available for evaluating the accuracy of its analytical results. The only way is to evaluate its accuracy using simulation. In order to establish TUA as a mathematically credible techniques, such theoretical evaluation has to be provided. Hence, a very important and challenging problem is how to evaluate the accuracy of TUA theoretically.

9.4 Conclusion

TUA is a unified approach for the approximate analysis of the performance of buffered random multiple access protocols. It has been applied successfully in S-ALOHA, CDMA ALOHA, Satellite ALOHA, CSMA, CSMA/CD, R-ALOHA and PRMA systems. TUA often has sufficiently accurate analysis results. However, we also see that TUA does not give satisfied accuracy when it is used to analyze some systems. Even in this case we can obtain at least a rough estimation of the system performance that can be useful at the first phase of the performance evaluation. TUA is not developed to replace the available approaches. Instead, it should be complementary with them. Generally speaking, some available approaches may have better accuracy than TUA if the analysis is tractable since they employ higher dimensional state space to describe the system. However, the analysis using these approaches may be intractable for more complicated systems. It is these complicated systems which are the main application areas of TUA.

BIBLIOGRAPHY

- [1] S. Tasaka, "Performance analysis of multiple access protocols," The MIT Press, 1986.
- [2] F. Tobagi, "Multiaccess protocols in packet communication systems," IEEE Trans. Commun., vol. COM-28, no. 4, pp. 468-488, 1980.
- [3] R. Prasad, "CDMA for wireless personal communications," Artech House, Boston, 1996.
- [4] R. Rom and M. Sidi, "Multiple access protocols: performance and analysis," Springer-Verlag, New York, 1990.
- [5] N. Abramson, "The ALOHA system-another alternative for computer communications," in 1970 Fall Joint Comput. Conf., AFIPS Conf. Proc., vol. 37, pp. 281-285, 1970.
- [6] L.G. Roberts, "ALOHA packet system with and without slots and capture," Comput. Commun. Rev., vol. 5, pp. 28-42, 1975.
- [7] L. Kleinrock and F. A. Tobagi, "Packet switching in radio channels: Part I-Carrier sense multiple access modes and their throughput delay characteristics," IEEE Trans. Commun., vol. COM-23, pp. 1400-1416, 1975.
- [8] H. H. Tan and K. Tsai, "Packet output processes of CSMA and CSMA/CD Protocols," IEEE Trans. Commun., vol. 44, no. 4, pp. 464-474, 1996.
- [9] F. A. Tobagi and V. B. Hunt, "Performance analysis of carrier sense multiple access with collision detection," Computer Networks, vol. 4, pp. 245-259, 1980.
- [10] J. L. Hammond and P. J. P. O'Reilly, "Performance analysis of local computer networks," Addison-Wesley Publishing Company, 1986.

- [11] F. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part II-The hidden terminal problem in carrier sense multiple access and the busy tone solution," *IEEE Trans. Commun.*, vol. COM-23, pp. 1417-1433, 1975.
- [12] L. Kleinrock and S. S. Lam, "Packet switching in a multiaccess broadcast channel: performance evaluation," *IEEE Trans. Commun.*, vol. COM-23, no. 4, pp. 410-422, 1975.
- [13] A. B. Carleial and M. E. Hellman, "Bistable behavior of ALOHA-type systems," *IEEE Trans. Commun.*, vol. COM-23, no. 4, pp. 401-410, 1975.
- [14] F. A. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part IV-stability considerations and dynamic control in carrier sense multiple access," *IEEE Trans. Commun.*, vol. COM-25, no. 10, pp. 1103-1119, 1977.
- [15] T. S. Rappaport, "Wireless communications: principle and practice," Prentice Hall, 1996.
- [16] A. G. Konheim and B. Meister, "Service in a loop system," *J. Ass. Comput. Mach.*, vol. 19, pp. 92-108, 1972.
- [17] J. F. Hayes, "An adaptive technique for local distribution," *IEEE Trans. Commun.*, vol. COM-26, 1978.
- [18] F. A. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part III-Polling and (dynamic) split channel reservation multiple access," *IEEE Trans. Commun.*, vol. COM-24, pp. 832-845, 1976.
- [19] I. Widipangestu, A. J. 'T Jong and R. Prasad, "Capture probability and throughput analysis of slotted ALOHA and unslotted np-ISMA in a Rician/Rayleigh environment," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 457-465, 1994.
- [20] J. C. Arnbak and W. V. Blitterswijk, "Capacity of slotted ALOHA in Rayleigh fading channels," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 261-269, 1987.
- [21] K. J. Zdunek, D. R. Ucci and J. L. LoCicero, "Packet radio performance of inhibit sense multiple access with capture," *IEEE Trans. Commun.*, vol. 45, pp. 164-167, 1997.
- [22] J. H. Wen and J. F. Chang, "The effect of multipath interference on the performance of packet radios," *IEEE Trans. Commun.*, vol. 38, pp. 740-743, 1990.
- [23] G. Wu, K. Mukumoto, and A. Fukuda, "Performance evaluation of reserved idle signal multiple-access scheme for wireless communication networks," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 658-658, 1994.

- [24] W. Stallings, "Data and computer communications," 5th Edition, Prentice Hall, 1997.
- [25] W. Stallings, "Local and metropolitan area networks," 5th Edition, Prentice Hall, 1997.
- [26] I. E. Pountourakis and E. D. Sykas, "Stability and throughput optimisation of multichannel slotted nonpersistent CSMA random choice protocol," IEE Electronics Letters, vol. 30, pp. 1387-1388, 1994.
- [27] M. Ajmone and F. Neri, "A simulation study of delay in multichannel CSMA/CD protocols," IEEE Trans. Commun., vol. 39, pp. 1590-1602, 1991.
- [28] T. N. Saadawi, M. H. Ammar and A. E. Hakeem, "Fundamentals of telecommunication networks," John Wiley & Sons, 1994.
- [29] I. M. Jacobs et al., "CPODA-A demand assignment protocol for SATNET," in Proc. 5th Data Commun. Symp., Snowbird, Utah, pp. 2-5 to 2-9, Sept. 1977.
- [30] L. G. Roberts, "Dynamic allocation of satellite capacity through packet reservation," in Proc. AFIPS conf., vol. 42, pp. 711-716, 1973.
- [31] R. Weissler et al., "Synchronization and multiple access protocols in the initial satellite IMP," in Proc. COMPCON 78, Washington D. C., pp. 356-362, 1978.
- [32] W. R. Crowther, R. Rettbert, D. Walden, S. Ornstein and F. Heart, "A system for broadcast communication: Reservation-ALOHA," in Proc. 6th Hawaii Int. Syst. Sci. Conf., 1973.
- [33] S. S. Lam, "Packet broadcast networks - A performance analysis of the R-ALOHA protocol," IEEE Trans. Comput., vol. C-29, no. 7, pp. 596-603, 1980.
- [34] S. Tasaka, "Stability and performance of the R-ALOHA packet broadcast system," IEEE Trans. Comput., vol. C-32, no.8, pp. 717-726.
- [35] S. Tasaka, "ALOHA-type satellite packet communication networks with buffering capacity," (in Japanese) Trans. IECE Japan, vol. J65-B, pp. 62-69, 1982.
- [36] D. J. Goodman, R. A. Valenzuela, K. T. Gayliard and B. Ramamurthi, "Packet reservation multiple access for local wireless communications," IEEE Trans. Commun., vol. 37, no. 8, pp. 885-809, 1989.
- [37] D. J. Goodman and S. X. Wei, "Efficiency of packet reservation multiple access," IEEE Trans. Veh. Technol., vol. 40, no. 1, pp. 170-176, 1991

- [38] H. Y. Chung and D. J. Goodman, "Transmission of speech and data using packet reservation multiple access," in IEEE ICC'91, pp. 0099-0140, 1991.
- [39] S. Nanda, D. J. Goodman and U. Timor, "Performance of PRMA: a packet voice protocol for cellular systems," IEEE Trans Veh. Technol., vol. 40, no. 3, pp. 584-598, 1991.
- [40] S. Nanda, "Analysis of packet reservation multiple access: voice data integration for wireless networks," in IEEE Globecom'90, pp. 1984-1988, 1990.
- [41] X. Qiu and V. K. Li, "On the capacity of packet reservation multiple access with capture in personal communication systems," IEEE Trans. Veh. Technol., vol. 45, no. 4, pp. 666-675, 1996.
- [42] F. Borgonovo and L. Fratta, "SRUC, A technique for packet transmission on multiple access channels," in Proc. Int. Conf. Comput. Commun., Kyoto, Japan, pp. 601-607, 1978.
- [43] L. Kleinrock and Y. Yemini, "An optimal adaptive scheme for multiple access broadcast communication," IEEE ICC'78, Toronto, pp. 7.2.1-7.2.5, 1978.
- [44] 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, 1982.
- [45] S. Tasaka, "Dynamic behavior of a CSMA-CD system with a finite population of buffered users," IEEE Trans. On Comm., vol. COM-34, no.6, 1986, pp. 576-586.
- [46] G. Wu, K. Mukumoto and A. Fukuda, "Analysis of an integrated voice and data transmission system using packet reservation multiple access," IEEE Trans. Veh. Technol., vol. 43, no. 2, pp. 289-297, 1994.
- [47] H. Qi and R. Wyrwas, "Performance analysis of joint voice-data PRMA over random packet error channels," IEEE Trans. Veh. Technol., vol. 45, no. 2, pp. 332-345, 1996.
- [48] M. Sidi and A. Segall, "Two interfering queues in packet radio networks," IEEE Trans. Commun., vol. COM-31, pp. 123-129, 1983.
- [49] H. C. Yu and R. L. Hamilton, "A buffered two-node packet radio network with product form solution," IEEE Trans. Commun., vol. 39, pp. 62-75, 1991.
- [50] H. Takagi and L. Kleinrock, "Mean packet queueing delay in a buffered two-user CSMA/CD system," IEEE Trans. Commun., vol. COM-33, pp. 1136-1139, 1985.

- [51] W. Szpankowski, "Bounds for queue lengths in a contention packet broadcast system," *IEEE Trans. Commun.*, vol. COM-34, pp. 1132-1140, 1986.
- [52] S. S. Kamal and S. A. Mahmoud, "A study of users' buffer variations in random access satellite channels," *IEEE Trans. Commnu.*, vol. COM-27, pp. 857-868, 1979.
- [53] T. Saadawi and A. Ephremides, "Comments on 'A study of users' buffer variations in random access satellite channels,'" *IEEE Trans. Commun.*, vol. COM-28, pp. 1436-1437, 1980.
- [54] K. K. Mittal and A. N. Venetsanopoulos, "Comments on 'A study of users' buffer variations in random access satellite channels,'" *IEEE Trans. Commun.*, vol. COM-29, p. 919, 1981.
- [55] D. V. Ramana and V. V. S. Sarma, "Performance of slotted ALOHA satellite channels with finite buffer," *IEE Proc.*, vol. 129, Pt. F, pp. 19-24, 1982.
- [56] K. K. Mittal and A. N. Venetsanopoulos, "Buffer analysis in random multiple access broadcast communication systems," *Can. Elect. Eng.*, vol. 10, no. 4, pp. 158-162, 1985.
- [57] T. N. Saadawi and A. Ephremides, "Analysis, stability, and optimization of slotted ALOHA with a finite number of buffered users," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 680-689, 1981.
- [58] A. O. Fapojuwo, W. C. Chan and D. Irvine-Halliday, "An iterative approximation for modelling buffered CSMA-CD LAN's," *Journal of Electrical and Electronics Engineering, Australia*, vol. 8, no. 2, pp. 102-111, 1988.
- [59] A. Ephremides and R. Z. Zhu, "Delay analysis of interacting queues with an approximate model," *IEEE Trans. Commun.*, vol. COM-35, pp. 194-201.
- [60] J. Silvester and I. Lee, "Performance modeling of buffered CSMA - an iterative approach," *Globecom 1982*, pp. 1195-1199.
- [61] T. K. Apostolopoulos and E. N. Protonotarios, "Queueing analysis of buffered slotted multiple access protocols," *Comput. Commun.*, vol. 8, pp. 9-21, 1985.
- [62] Y. Matsumoto, Y. Takahashi and T. Hasegawa, "The effects of packet size distributions on output and delay processes of CSMA/CD," *IEEE Trans. On Comm.*, vol.38, pp. 199-214, 1990.
- [63] T. Takine, Y. Takahashi and T. Hasegawa, "An approximate analysis of a buffered CSMA/CD," *IEEE Trans. On Comm.*, vol. 36, pp. 932-941, 1988.

- [64] P. D. Welch, "On a generalized M/G/1 queuing process in which the first customer of each busy period receives exceptional service," *Oper. Res.*, vol. 12, pp. 736-752, 1964.
- [65] Y. Matsumoto, Y. Takahashi and T. Hasegawa, "Probability distributions of interdeparture time and response time in multipacket CSMA/CD systems," *IEEE Trans. On Comm.*, vol. 38, pp. 54-66, 1990.
- [66] Y. Matsumoto, Y. Takahashi and T. Hasegawa, "Performance analysis of CSMA/CD networks with a buffered gateway," *IEEE Trans. Commun.*, vol. 38, pp. 1404-1418, 1990.
- [67] E. D. Sykas, D. E. Karvelas and E. N. Protonotarios, "Queueing analysis of some buffered random multiple access schemes," *IEEE Trans. On Comm.*, vol. COM-34, no.8, pp. 790-798, 1986.
- [68] T. K. Apostolopoulos and E. N. Protonotarios, "Queueing analysis of buffered CSMA/CD protocols," *IEEE Trans. On Comm.*, vol. COM-34, pp. 898-905, 1986.
- [69] A. and I. Chlamtac, "A linear solution to queueing analysis of synchronous finite buffer networks," *IEEE Trans. On Comm.*, vol. 38, pp. 440-447, 1990.
- [70] A. Ganz and I. Chlamtac, "An efficient solution to finite buffer non-persistent carrier sense multiple access protocol," *Computer Networks and ISDN Systems*, vol. 18, pp. 263-274, 1989/90.
- [71] F. A. Tobagi, "Analysis of a two-hop centralized packet radio network - Part I: Slotted ALOHA," *IEEE Trans. Commun.*, vol. COM-28, pp. 196-207, 1980.
- [72] F. A. Tobagi, "Analysis of a two-hop centralized packet radio network - Part II: Carrier sense multiple access," *IEEE Trans. Commun.*, vol. COM-28, pp. 208-216, 1980.
- [73] L. Kleinrock and Y. Yemini, "Interfering queueing processes in packet-switch broadcast communications," in *Proc. IFIP Congr. 80*, Tokyo, Japan and Melbourne, Australia, pp. 557-562, Oct. 1980.
- [74] F. A. Tobagi and L. Kleinrock, "On the analysis and simulation of buffered packet radio systems," in *Proc. 9th Hawaii Int. conf. Syst. Sci.*, pp. 42-45, Jan. 1976.
- [75] M. Lo and T. P. Yum, "The tone sense multiaccess protocol with partial collision detection (TSMA/PCD) for packet satellite communications," *IEEE Trans. Commun.*, vol. 41, pp. 820-824, 1993.
- [76] R. A. Howard, "Dynamic probabilistic systems, Volume II: Semi-Markov and decision processes," John Wiley & Sons, 1971.

- [77] CDPD Forum Inc., "Cellular digital packet data system specification: Release 1.1," January 19, 1995.
- [78] H. Takagi, "Queueing analysis - A foundation of performance evaluation, vol. 2: Finite Systems," North-Holland Publishing Company, 1993.
- [79] H. Takagi, "Queueing analysis - A foundation of performance evaluation, vol. 3: Discrete-Time Systems," North-Holland Publishing Company, 1993.
- [80] R. B. Cooper, "Introduction to queueing theory," Second Edition, North-Holland Publishing Company. Republished by the Continuing Engineering Education Program, The George Washington University, 1990.
- [81] S. J. Mason, "Feedback theory - some properties of signal flow graphs," Proc. IRE, vol. 41, pp. 1144-1156, 1953.
- [82] W. A. Wolovich, "Automatic control systems: basic analysis and design," Saunders College Publishing, 1994.
- [83] F. H. Raven, "Automatic control engineering," 5th Edition, McGraw-Hill, Inc., 1995.
- [84] L. Kleinrock, "Queueing systems, Volume 1: theory," John Wiley & Sons, 1975.
- [85] H. Bruneel and B. G. Kim, "Discrete-time models for communication systems including ATM," Kluwer Academic Publishers, 1993.
- [86] W. Crowther, R. Rettberg, D. Walden, S. Ornstein and F. Heart, "A system for broadcast communications: Reservation-ALOHA," in Proc. 6th Hawaii Int. conf. Syst. Sci., pp. 373-374, 1981.
- [87] T. G. Robertazzi, "Computer networks and systems: queueing theory and performance evaluation," 2nd Edition, Springer-Verlag, 1994.
- [88] H. Okada, T. Yamazato, M. Katayama and A. Ogawa, "Queueing analysis of CDMA slotted ALOHA systems with finite buffer and finite population assumptions," IEEE ICC'98.
- [89] M. A. Marsan and F. Neri, "A simulation study of delay in multichannel CSMA/CD protocols," IEEE Trans. Commun., vol. 39, no. 11, pp. 1590-1602, 1991.
- [90] Tao Wan and Asrar U.H. Sheikh, "A unified approach to performance analysis of multiple access protocols with buffered users," submitted to IEEE Trans. on Veh. Tech., U.S.A., 1998.

- [91] Tao Wan and Asrar U.H. Sheikh, "Analysis of finite population buffered slotted ALOHA protocols using tagged user analysis (TUA)," IEEE Globecom'98, Sydney, Australia, 1998.
- [92] Tao Wan and Asrar U.H. Sheikh, "Stability analysis of buffered slotted ALOHA protocols using Tagged User Approach," accepted for publication in IEEE Trans. on Veh. Tech., U.S.A., 1998.
- [93] Tao Wan, Ioannis Lambadaris, Mike Devetsikiotis and Asrar Sheikh, "Queueing analysis of buffered slotted DS/CDMA ALOHA protocols using Tagged User Approach (TUA)," to be presented in IEEE ICC'99, Vancouver, Canada, 1999.
- [94] Tao Wan and Asrar Sheikh, "Performance analysis of CSMA/CD systems," submitted to IEEE Trans. on Veh. Tech., U.S.A., 1999.
- [95] Tao Wan and Asrar U.H. Sheikh, "Performance analysis of buffered R-ALOHA systems using Tagged User Approach," submitted to IEEE Trans. on Veh. Tech., U.S.A., 1999.
- [96] Tao Wan, Asrar U.H. Sheikh, Yonghai Gu, and Abdulbasset Atweri, "Performance analysis of PRMA with integration of voice and buffered data users," IEEE PIMRC'98, Boston, U.S.A., October 8-11, 1998.
- [97] Tao Wan, Asrar U. H. Sheikh, Paul Tse and Abdulbaset Atweri, "Performance analysis of cellular digital packet data (CDPD) system," IEEE VTC'98, Ottawa, Canada, 1998.
- [98] Tao Wan and Asrar U.H. Sheikh, "performance analysis of packet reservation multiple access (PRMA) systems using Tagged User Approach," revised and re-submitted to IEEE Trans. on Veh. Tech., U.S.A., 1998.

A FLOW GRAPH ANALYSIS OF DISCRETE-TIME MARKOV CHAINS AND SEMI-MARKOV CHAINS

In the thesis only slotted multiple access protocols are concerned, where each user is a discrete-time queueing system. Here we present a brief introduction on the flow graph analysis [76] of discrete-time Markov and semi-Markov chains. We further focus our analysis on homogeneous chains only. These chains are such that their transition probabilities are stationary with time.

A.1 Flow Graph Analysis of Markov Chains

The discrete-time Markov chain has the property that the transition from current state to some other state, including back to the same state, must occur in each time unit [84]. Let p_{ij} be the state transition probability of a discrete-time Markov chain from state i to state j . The system behavior can be described using a state flow graph where a branch with direction from i toward j and weight $p_{ij}z$ is associated with the states i and j , where z indicates that the required time for the transition is one time unit. Normally, if a state transition from i to j needs n time units in a semi-Markov processes, the corresponding weight, called transfer function, is $p_{ij}z^n$. Since discrete-time Markov processes are concerned in this section,

the time needed for each state transition is always one time unit due to the memoryless property of Markov processes.

For example, the state flow graph for a two-state Markov chain is given in Figure A.1, where the transition probability matrix is given by

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}. \quad (\text{A.1})$$

Figure A.1. State flow graph of a two-state Markov chain

Suppose we wish to calculate the PGF $\psi_{11}(z)$ of the time required until the process comes back to state 1 for the first time, given the current state is state 1. To solve this problem, we modify the diagram in Figure A.1 to fit our task, as shown in Figure A.2. It is clear that the required time is just the time duration from point A to point B in Figure A.2. It is proved that the PGF of the time duration from point A to point B is just the transfer function from point A to point B. Therefore, the required PGF is obtained using Mason's formula [81], which is given by

$$\psi_{11}(z) = \frac{0.06z^2}{(1 - 0.8z)(1 - 0.7z)}. \quad (\text{A.2})$$

Figure A.2. State flow graph for calculating the required PGF

Now we consider another example. Suppose we wish to calculate $\phi_{11}(n)$, the probability that the process will be in state 1 after n transitions, given that it starts in state 1. The transform $\phi_{11}(z)$, defined as

$$\phi_{11}(z) = \sum_{n=0}^{\infty} \phi_{11}(n)z^n, \quad (\text{A.3})$$

can be calculated using the state flow graph given in Figure A.3.

Figure A.3. State flow graph for calculating $\phi_{11}(z)$

The required transform $\phi_{11}(z)$ is just the transfer function from point C to point D, which is given by

$$\phi_{11}(z) = \frac{1 - 0.7z}{1 - 1.5z + 0.5z^2}. \quad (\text{A.4})$$

It is easy to get $\phi_{11}(n)$ using (A.4). It should be noticed that $\phi_{11}(z)$ is not a PGF.

A.2 Flow Graph Analysis of Semi-Markov Chains

The semi-Markov chain can be thought as a process whose successive state occupancies are governed by the transition probabilities of a Markov chain, but whose stay in any state is described by an integer-valued random variable that depends on the current state and on the next state to which the transition will be made. Thus at imbedded transition instants, the semi-Markov chain behaves just like a Markov chain. This process is called imbedded Markov chain. In discrete-time Markov chains, the times between transitions is geometrically distributed. However, in discrete-time semi-Markov chains, the time duration between transitions is permitted to obey an arbitrary probability distribution (in multiple time units).

Let p_{ij} be the state transition probability of a semi-Markov chain from state i to state j , and τ_{ij} be the time duration between transitions from state i to state j . The state flow graph of a general two-state semi-Markov chain is given in Figure A.4. The idea for the flow graph analysis of semi-Markov chains is same as that for Markov chains. For example, it is easy to find that in this case the PGF $\psi_{11}(z)$ of the time required until the process comes back to state 1 for the first time, given the current state is state 1, is given by

$$\Psi_{11}(z) = \frac{p_{12}p_{21}z^{\tau_{12} + \tau_{21}}}{(1 - p_{11}z^{\tau_{11}})(1 - p_{22}z^{\tau_{22}})}. \quad (\text{A.5})$$

We can also find that in this case $\phi_{11}(z)$ is given by

$$\phi_{11}(z) = \frac{1 - p_{22}z^{\tau_{22}}}{(1 - p_{11}z^{\tau_{11}})(1 - p_{22}z^{\tau_{22}}) - p_{12}p_{21}z^{\tau_{12} + \tau_{21}}}. \quad (\text{A.6})$$

Figure A.4. State flow graph of two-state semi-Markov chains

B DISCRETE-TIME QUEUEING SYSTEMS

B.1 Discrete-time Models

A discrete-time queueing system is characterized by time-slotted service. The time axis is divided into equal intervals, called slots. Customer arrivals, and the starting and completing service of customers are supposed to occur only at slot boundaries.

Due to the discrete nature of the model, different assumptions can be made on the exact times of customer arrivals and departures with respect to time boundaries. Without loss of generality, we normalize the slot length to unit time. Slots are sequentially numbered in nonnegative integers so that slot j is located in time $(j-1, j]$, $j = 1, 2, 3, \dots$. Denoting by j^- and j^+ two time points immediately before and after the time j , respectively. Throughout the thesis, we use the following convention: (i) a customer whose service starts in slot $j+1$ commences the service sometime in (j, j^+) ; (ii) a customer completing service in slot j is considered to depart from the system sometime in (j^-, j) . Though there is no restriction on customer arriving times in general, we assume that the customer arrival occurs also at slot boundaries for analysis convenience since system states are measured typically at slot boundaries in discrete-time systems.

There are two models, called late arrival model and early arrival model, associated with customer arrivals. In the former, we assume that customer arrivals over entire slot j be supposed to occur at time j^- . In the latter, we assume that customer arrivals in slot $j + 1$ occur at time j^+ . Here we are interested in later arrival model only.

When the system has finite buffer capacity, we need to specify a rule for the users concerning the acceptance of a customer that arrives in the last slot of a service time when the system is full [79]. In the late arrival model, two rules are defined regarding to this problem. In the first rule, called late arrival model with delayed replacement, the arriving customer is rejected even though a customer would depart from the queue in the same slot. In the other one, called late arrival model with immediate replacement, this customer is accepted. Only late arrival model with delayed replacement is considered in this thesis.

B.2 Notation

In order to represent the characteristics of queueing models in a concise manner, Kendall's notation in the form of $A/B/m/K/M$ has been widely employed, where A and B denote the interarrival time and service time distribution, respectively, m the number of servers, K the storage capacity of the system, and M the customer population in the system. If either of the last two descriptors is absent, we assume that it take value of infinite.

Hence, we denote by Geo/Geo/1/K a single server discrete-time queueing system with finite storage capacity, whose customer interarrival time and service time distribution are geometrically distributed. Similarly, a single server discrete-time queueing system with infinite storage capacity, a Bernoulli arrival process and generally distributed service time is denoted by Geo/G/1/K. In the above two queueing models, at most one customer arrival

is permitted in a slot. When more than one customers are allowed to arrive in a slot, the system is called bulk arrival system. In the thesis, we follow the notations in [79]. A bulk arrival Geo/G/1/K queueing system is denoted by $Geo^{[x]}/G/1/K$.

It should be mentioned that unlike continuous-time queueing system, there is no unified notation for discrete-time queueing system. For example, a model denoted by $Geo^{[x]}/G/1/K$ in [79] may be denoted by GI/G/1/K in [85].

B.3 Analysis of Geo/Geo/1/K System

A Geo/Geo/1/K or Geo/Geo/1 system is indeed a discrete-time birth-death process. Here we are interested in only the system with delay replacement model. Suppose the system storage capacity is L customers, the customer arrival rate and service rate are denoted by λ and μ , respectively. The most interest in queueing analysis is to find the equilibrium probability p_k , $k = 1, 2, \dots, L$. The analysis can be carried out using the diagram shown in Figure B.1. The transitions from a state into itself are not drawn since they

Figure B.1. State transition rate diagram for Geo/Geo/1/K

do not affect the analysis.

Since the total probability flow through any closed boundary must be zero, it is easy to obtain the following relations:

$$\begin{aligned}\lambda p_0 &= \mu(1-\lambda)p_1, \\ \lambda(1-\mu)p_i &= \mu(1-\lambda)p_{i+1}, \quad i = 1, 2, \dots, L-2 \\ \lambda(1-\mu)p_{L-1} &= \mu p_L,\end{aligned}\tag{B.1}$$

From (B.1), all p_i , $i = 1, 2, \dots, L$, can be expressed using p_0 as

$$\begin{aligned}p_i &= \left[\frac{\lambda(1-\mu)}{\mu(1-\lambda)} \right]^{i-1} \frac{\lambda}{\mu(1-\lambda)} p_0, \quad i = 1, 2, \dots, L-1 \\ p_L &= \frac{\lambda}{\mu} \left[\frac{\lambda(1-\mu)}{\mu(1-\lambda)} \right]^{L-1} p_0.\end{aligned}\tag{B.2}$$

Substituting (B.2) into the total probability law

$$\sum_{i=0}^L p_i = 1,\tag{B.3}$$

we have

$$p_0 = \left(\frac{\mu}{\mu-\lambda} - \frac{\lambda^2}{\mu(\mu-\lambda)} \left[\frac{\lambda(1-\mu)}{\mu(1-\lambda)} \right]^{L-1} \right)^{-1}.\tag{B.4}$$

Now, the system equilibrium probabilities are obtained. The performance measurements of the system can be developed based on the equilibrium probabilities. The customer blocking probability p_B is given by

$$p_B = p_L\tag{B.5}$$

since p_B is just the probability that an arriving customer finds the queue full and therefore gets lost, which happens to be p_L . The system throughput θ , defined as the average number of customers served in a slot, is therefore given by

$$\theta = \lambda(1 - p_B) \quad (\text{B.6})$$

since the average number of customers served in the system should be equal to the average number of customers accepted into the system in each slot.

Now we derive the PGF of the waiting time, defined as the number of slots which a customer spends in the system before being served. Because of BASTA (Bernoulli arrival see time average) property, an arriving customer would see k customers in the system with probability p_k . For a customer accepted into the system, with probability $p_0/(1-p_B)$ it will see the queue empty and be served immediately upon arrival. That is, the waiting time is zero. With probability $p_i/(1-p_B)$ a customer in the queue will see i , $1 \leq i \leq L-1$, customers in the queue when arriving. In this case, the waiting time is the total service time of all the customers lining ahead of it at the beginning of next slot, which is $i-1$ with probability μ , and i with probability $1-\mu$ since the service of the customer in the server will be completed in each slot with probability μ . Hence, the PGF $W(z)$ of the waiting time is given by

$$W(z) = \frac{p_0}{1-p_L} + \sum_{i=1}^{L-1} \frac{p_i}{1-p_L} [\mu B^{i-1}(z) + (1-\mu)B^i(z)], \quad (\text{B.7})$$

where

$$B(z) = \frac{\mu z}{1 - (1 - \mu)z} \quad (\text{B.8})$$

is the PGF of the service time for customers.

The system response time of a customer is defined as the time duration which a customer spends in the system, which is equal to the waiting time plus the service time. Hence, the PGF $R(z)$ of system response time is given by

$$R(z) = B(z)W(z). \quad (\text{B.9})$$

Note (B.1) holds for any multiple access protocols.

In some systems the output process, the inter-departure time distribution, of a queueing system may constitute a part of the arrival process for another queueing system. In that case, output process is interesting. Here we develop the output process or the PGF $O_u(z)$ of the customer inter-departure time of Geo/Geo/1/K system.

A departed customer would leave the system empty or not with probability π_0 or $1 - \pi_0$, respectively, where π_0 is the probability that a departure packet leaves queue empty, which is given by [79]

$$\pi_0 = \frac{\lambda}{\mu} \cdot \frac{P_0}{1 - P_0}. \quad (\text{B.10})$$

In the former case, the interdeparture time equals the interarrival time plus service time. In the latter case, it is equal to the interarrival time. Hence, we have

$$\begin{aligned} O_u(z) &= \pi_0 B(z) + (1 - \pi_0) B(z) \frac{\lambda z}{1 - (1 - \lambda)z} \\ &= \frac{B(z)}{1 - (1 - \lambda)z} [\lambda z + (1 - \pi_0)(1 - z)]. \end{aligned} \quad (\text{B.11})$$

As to the mean values, they can be obtained easily if the corresponding distributions are known. The same idea here can be used for the analysis of Geo/Geo/1 system, on which the analysis is much simpler. However, the system is ergodic only if $\lambda < \mu$.

B.4 Analysis of Geo/Geo/1 System

Here we consider a system with the same model as that in Section B.3 except that the system has infinite buffer capacity. This Geo/Geo/1 system can be considered as a special case of the above Geo/Geo/1/K system with $L = \infty$.

It is well known that we must have $\lambda < \mu$ in order that the system is stable. From (B.2), we have

$$p_i = \left[\frac{\lambda(1-\mu)}{\mu(1-\lambda)} \right]^{i-1} \frac{\lambda}{\mu(1-\lambda)} p_0, \quad i = 1, 2, \dots, \infty. \quad (\text{B.12})$$

Using total probability law and noting $[\lambda(1-\mu)]/[\mu(1-\lambda)] < 1$, we have

$$p_0 = 1 - \rho, \quad (\text{B.13})$$

where

$$\rho = \lambda/\mu \quad (\text{B.14})$$

is called utilization.

Using (B.12) and (B.13), the PGF $P(z)$ of queue length is given by

$$P(z) = \sum_{i=0}^{\infty} p_i z^i = \frac{(1-\rho)(1-\lambda+\lambda z)}{1-\lambda-\rho(1-\mu)z}. \quad (\text{B.15})$$

Similar to (B.7), the PGF $W(z)$ of the waiting time is given by

$$\begin{aligned}
W(z) &= p_0 + \sum_{i=1}^{L-1} p_i [\mu B^{i-1}(z) + (1-\mu)B^i(z)] \\
&= \frac{(1-\rho)}{1-\lambda-(1-\mu)z} \cdot [1-(1-\mu)z].
\end{aligned} \tag{B.16}$$

In infinite buffer system, no customer is rejected upon arrival, the system throughput is equal to the customer arrival rate. Hence, we have

$$\theta = \lambda. \tag{B.17}$$

In Geo/Geo/1 system, we have $\pi_0 = p_0$. Substituting (B.8) and (B.13) into (B.11), the PGF of customer interdeparture time is given by

$$O_u(z) = \frac{\lambda z}{1-(1-\lambda)z}. \tag{B.18}$$

That is, the interdeparture time has the same distribution as interarrival time. This property holds also for continuous time M/M/1 system.

B.5 Analysis of Geo/G/1/K System

The system behavior of Geo/G/1/K system is described by a discrete time semi-Markov chain. The analysis of Geo/G/1/K system is much hard than that of Geo/Geo/1/K system. In this section, we are not going through the analysis process. Only the results are presented. Interested readers are referred to [79] for details.

We consider a Geo/G/1/K system with Bernoulli arrival rate λ , system capacity of L customers, and a generally distributed service time whose PGF is denoted by $B(z)$.

Let π_i , $i = 0, 1, \dots, L-1$ be the probability that a departure customer leaves i customers in the system, we have

$$\pi_k = \pi_0 a_k + \sum_{j=1}^{k+1} \pi_j a_{k-j+1}, \quad 0 \leq k \leq L-2, \quad (\text{B.19})$$

and

$$\sum_{i=0}^{L-1} \pi_i = 1, \quad (\text{B.20})$$

where a_k is the probability that k customers arrive during a service time. The PGF $A(z)$ of the number of customer arrivals during a service time is given by

$$A(z) = \sum_{k=0}^{\infty} a_k z^k = B(1 - \lambda + \lambda z). \quad (\text{B.21})$$

An efficient algorithm for computing π_k , $k = 0, 1, \dots, L-1$, can be found by defining

$$\pi'_k = \pi_k / \pi_0, \quad k = 0, 1, \dots, L-1. \quad (\text{B.22})$$

Then it is found that π'_k , $k = 0, 1, \dots, L-1$, can be recursively calculated as follows:

$$\begin{aligned} \pi'_0 &= 1 \\ \pi'_{k+1} &= \frac{1}{a_0} \left(\pi'_k - \sum_{j=1}^k \pi'_j a_{k-j+1} - a_k \right), \quad 0 \leq k \leq L-2, \end{aligned} \quad (\text{B.23})$$

and π_0 is found using total probability law (B.20) as

$$\pi_0 = \left(\sum_{i=0}^{L-1} \pi'_i \right)^{-1}. \quad (\text{B.24})$$

Let p_k , $k = 0, 1, \dots, L$, be the probability that there are k packets present in the system at an arbitrary time (slot boundary). Then we have

$$\begin{aligned}
 p_k &= \frac{\pi_k}{\pi_0 + \rho}, \quad 0 \leq k \leq L-1 \\
 p_L &= p_B = 1 - \frac{1}{\pi_0 + \rho},
 \end{aligned}
 \tag{B.25}$$

where

$$\begin{aligned}
 \rho &= \lambda b, \\
 b &= B'(1),
 \end{aligned}
 \tag{B.26}$$

where b is the mean service time of customers and p_B is the blocking probability.

Now, the system equilibrium probabilities are obtained. The distributions of the waiting time, response time and interdeparture time can be derived. The PGF of the waiting time is given by

$$W(z) = \frac{\pi_0(1-z)}{1-\lambda-z+\lambda B(z)} \left(1 - \left[\frac{\lambda B(z)}{z-1+\lambda} \right]^L \right) + [B(z)]^{L-1} \sum_{j=0}^{L-1} \pi_j \left(\frac{\lambda}{z-1+\lambda} \right)^{L-j}. \tag{B.27}$$

Note that equations (B.6), (B.9), (B.10) and (B.11) are applicable in Geo/G/1/K system here.

Denoting by t_w and t_r the waiting time and system response time of an arbitrary customer, respectively. Though the mean waiting time can be obtained using (B.27) because of $E(t_w) = W'(1)$, more convenient way is to calculate the mean system response time by Little's law and keep in mind that the mean waiting time is equal to the mean system response time minus the mean service time. Thus we have

$$E(t_w) = \frac{E(l_q)}{\lambda(1-p_B)} - b, \tag{B.28}$$

where l_q is the queue length at any time and its mean $E(l_q)$ is given by

$$E(l_q) = \sum_{k=0}^L k p_k. \quad (\text{B.29})$$

B.6 Analysis of Geo/G/1 System

Here we consider a system with the same model as that in Section B.5 except that the system has infinite buffer capacity. Though this Geo/G/1 system can be considered as a special case of the above Geo/Geo/1/K system with $L = \infty$, we do not have to treat it this way. In fact, the analysis on Geo/G/1 is much simpler than that on Geo/G/1/K system.

We use the same notation as in Section B.5. From [79], the PGF $P(z)$ of the queue length at any time (slot boundary) and PGF $\Pi(z)$ of the queue length at customer departure time are equal and given by

$$P(z) = \Pi(z) = \frac{(1-\rho)(1-z)B(1-\lambda+\lambda z)}{B(1-\lambda+\lambda z)-z}. \quad (\text{B.30})$$

The PGF of waiting time is given by

$$W(z) = \frac{(1-\rho)(1-z)}{1-\lambda-z+\lambda B(z)}. \quad (\text{B.31})$$

Note that equations (B.9), (B.11) and (B.17) are applicable here.