



Reservation – Time Division Multiple Access Protocols for Wireless Personal Communications

Theodore V. Buot

B.S.Eng (Electro&Comm), M.Eng (Telecomm)

Thesis submitted for the degree of

Doctor of Philosophy

in

The University of Adelaide

Faculty of Engineering

Department of Electrical and Electronic Engineering

August 1997

Contents

Abstract	iv
Declaration	v
Acknowledgments	vi
List of Publications	vii
List of Abbreviations	viii
Symbols and Notations	xi
Preface	xiv
1. Introduction	1
<i>Background, Problems and Trends in Personal Communications and description of this work</i>	
2. Literature Review	12
2.1 ALOHA and Random Access Protocols	14
2.1.1 Improvements of the ALOHA Protocol	15
2.1.2 Other RMA Algorithms	16
2.1.3 Random Access Protocols with Channel Sensing	16
2.1.4 Spread Spectrum Multiple Access	17
2.2 Fixed Assignment and DAMA Protocols	18
2.3 Protocols for Future Wireless Communications	19
2.3.1 Packet Voice Communications	19
2.3.2 Reservation based Protocols for Packet Switching	20
2.3.3 Voice and Data Integration in TDMA Systems	23
3. Teletraffic Source Models for R-TDMA	25
3.1 Arrival Process	26
3.2 Message Length Distribution	29
3.3 Smoothing Effect of Buffered Users	30
3.4 Speech Packet Generation	32
3.4.1 Model for Fast SAD with Hangover	35
3.4.2 Effect of Hangover to the Speech Quality	38
3.5 Video Traffic Models	40
3.5.1 Infinite State Markovian Video Source Model	41
3.5.2 AutoRegressive Video Source Model	43
3.5.3 VBR Source with Channel Load Feedback	43
3.6 Summary	46
4. Performance Analysis of R-TDMA	48
4.1 System Model	48
4.1.1 Channel	49
4.1.2 Slot Reservation	49
4.1.3 Immediate First Transmission	50
4.1.4 Effect of Random Access Collisions	51
4.1.5 Single Carrier System	51
4.1.6 State (In)Dependence	51
4.2 Analysis Methods	52

4.3	Approximation of a Slotted Random Access Protocol	55
4.3.1	Finite Population ALOHA	57
4.3.2	Binary Exponential Back-off	58
4.4	Analysis of the Channel Allocation Queue	61
4.4.1	System Model (R-TDMA Channel Allocation)	61
4.4.2	Solving the Steady-State Occupancy	62
4.4.3	Analysis of the Queue	65
4.5	System Model for R-TDMA Protocols	67
4.5.1	S-G Analysis	69
4.5.2	Effect of Retransmission Probability	71
4.5.3	Stability of ATDMA	73
4.5.4	Mean Delay Analysis of ATDMA	75
4.6	Summary	78
5.	Reservation-TDMA Protocols for WPC	80
5.1	WPC with R-TDMA Multiaccess Protocol	80
5.1.1	Logical Channel Structure	83
5.1.2	R-TDMA Support for Voice Traffic	84
5.1.3	R-TDMA Support for Data Traffic	85
5.2	R-TDMA Performance with Packet Voice Traffic	85
5.2.1	Contention Process	86
5.2.2	Channel Allocation Process	87
5.2.3	Results based on the model	89
5.3	Reservation Policy for Data Users	91
5.3.1	Immediate Assignment Allocation Scheme	91
5.3.2	Performance Comparison	92
5.4	Enhancements to the R-TDMA Protocol	94
5.4.1	Effect of Capture and Forward Error Correction	94
5.4.2	Capture and Antenna Beam Overlap	96
5.4.3	R-TDMA with Dynamic Frame Configuration	98
5.4.4	Integrated Voice/Data ATDMA Protocol	101
5.5	Multipriority Channel Access	106
5.5.1	Stack Algorithm	107
5.5.2	Approach to Stack Prioritisation	110
5.5.3	Multipriority Stack Algorithm	115
5.6	ATDMA with Stack Algorithm	117
5.6.1	Analysis of ATDMA with Stack CRA using TFA	117
5.6.2	Voice and Data Prioritised Stack ATDMA	120
5.7	Random Access and Polling Solution	123
5.7.1	Protocol Description	123
5.7.2	State Transition Cycle	126
5.7.3	Throughput/Delay Approximation	126
5.7.4	Calculation based on the Polling Cycle	128
5.7.5	Stability	130
5.7.6	Simulation Model	131
5.7.7	Base Station Polling Control	132
5.8	Integrated-SCARP Protocol	132
5.9	Summary of Chapter 5	134

6. MultiMedia Access Protocol	135
6.1 Multislot Reservation for MultiMedia R-TDMA	136
6.1.1 Multislot Allocation Schemes and Fairness Criteria	137
6.1.2 Approximate Analysis using Birth & Death Markov Chains	138
6.1.3 Approximate Analysis using Discrete Markov Analysis	141
6.2 Multislot Reservation with Multiclass Users	151
6.2.1 Simple Algorithms for Multislot Systems with Heterogeneous Users	152
6.2.2 Best Effort Algorithms for Prioritised Multislot Systems	153
6.2.3 Simulation Parameters	156
6.2.4 Discussion	159
6.2.5 Summary	160
6.3 Multislot Reservation with Mixed Traffic	163
6.3.1 Reservation Policy for Mixed Traffic	163
6.3.2 Simulation Model	165
6.3.3 Simulation Results and Observations	166
6.4 Variable Coding Rate Multislot Multimedia System	172
6.4.1 Channel Model	173
6.4.2 Channel Coding	174
6.4.3 Simulation Model	175
6.4.4 Summary	177
6.5 QoS Maintenance for Wireless Video Transmission	181
6.5.1 System Model	181
6.5.2 Static Optimisation for Video/Data System	183
6.5.3 Video with Dynamic Load Feedback	184
6.5.4 Simulations and Observations	187
6.5.6 Summary of Chapter 6	187
7. Conclusions	190
7.1 Thesis Summary	190
7.2 Future Work	193
Appendices	
A Sample Data Source with Buffering	195
B ATDMA Performance	196
C Simulation on the Stability of S-ALOHA	200
D Rough Approximation of R-TDMA with Voice Traffic	202
E Derivation of the Stack Splitting Parameter	203
F Results of the Multiclass Stack Simulations	205
G Results of the SCARP Simulations	208
H Results of the Multiclass Multislot Simulation	210
References	211
Bibliography	223

Abstract

Packet switching technology is seen to provide more capacity and flexibility for future Wireless Personal Communications (WPC). A reservation based multiple access protocol has been considered to support packet-switched access in the wireless environment. This thesis is aimed to improve the design and performance of Reservation based Time Division Multiple Access (R-TDMA) protocols for WPC.

To provide an efficient voice and data integration, the R-TDMA protocol must support the following, a) fast channel access, b) variable rate transmission and c) fast error recovery. The first two design criteria are essential for all traffic types while the third criteria is required for delay sensitive and error sensitive services. R-TDMA protocol is chosen due to the ease of providing a steady, bursty and variable rate traffic after the design criteria are satisfied.

To provide a fast channel access, many possibilities were explored like the use of prioritisation at the random access, the use of stack algorithm, the exploitation of the capture probability by improving the topology and the use of a polling mechanism to support the random access. For the support of a variable rate transmission, the TDMA frame structure is exploited by employing a multislot reservation. It is required by services which are sensitive to the message delivery time and services characterised by multirate transmission. It is used in the provision of priority control for multiclass and mixed traffic.

The other aspect of this thesis is the derivation of some performance evaluation tools. The performance evaluation consists of a source traffic modelling and characterisation and the derivations of analytical procedures. Approximations were mainly used and were sustained with simulations. For voice traffic, the packet dropping rate is used as a design benchmark. Multiplexing gains for packet-voice in the order of 1.8 to 2.3 were found to be achievable. For the data traffic, both the mean and the cumulative delay performance benchmarks were considered. The support of video traffic is also investigated which suggests a coding scheme in order to transmit a high quality video. The overall assessment suggests that R-TDMA is a potential technology for WPC.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

SIGNED:

DATE: 15 AUGUST 1997

Acknowledgments

I would like to thank all individuals and institutions who on in some way or another has rendered valuable inputs and supports to bring this manuscript into completion. I would like to express my sincere thanks to my supervisor, Prof. Reg Coutts for his constant academic, financial and moral support during my three and a half years of research at the Centre for Telecommunications Information Networking (CTIN) of the University of Adelaide. His very long experience in both industry and research has provided me with valuable inputs regarding the direction of this research.

I am indebted to Fujio Watanabe for our constant collaboration during the entire period of this research. His comments and those of his colleagues of the Communications Research Laboratory (CRL, Japan) has provided me with good insights in conducting my performance evaluations. I would also to thank my peers in CTIN, Martin Ostrowski, Dohun Kwon, Yu-ShaoKai and others with there valuable discussions and comments. I am also grateful to John Leske, Derek Rogers, Tony Smith and Sergey Nesterov as well as all the staff of CTIN who answered my needs and not to forget Collette Snowden for editing this document.

I have to thank the Australian Government who supported me with the Overseas Postgraduate Research Scholarship as well and the University of Adelaide for providing me with financial support. Also to CTIN for sending me to overseas conferences.

I will not forget to mention the support of the Filipino students of South Australia despite my limited time with them. My sincere gratitude to the Caluya families who fostered me during my early years in Adelaide, also to the Del Castillo family, and to all my friends in Adelaide.

A am very grateful to my own family, especially my wife Cherry for her love and patience while I was away for three years. Her constant inspiration has provided me with extra effort in bringing this thesis into completion. I am thankful to my in-laws who cherished my family during my absence. Also to my own brothers and sisters who are always there to support me. I am very grateful to my parents, whose guiding principles has constantly reminded me of the value of higher education.

List of Publications

Journal Publications:

- 1) T. Buot and F. Watanabe, "*Random Access Algorithm for Users with Multiple Priorities*," Special Issue on Advanced Adaptive Radio Communications Technologies, **IEICE of Japan Transactions on Communications**, March 1996.
- 2) T. Buot and F. Watanabe, "*Channel Allocation Algorithm for Multislot TDMA with Multiclass Users*," Special Issue on Advanced Adaptive Radio Communications Technologies, **IEICE of Japan Transactions on Communications**, March 1996.

Conference Publications:

- 1) T. Buot, "*Random Access, Reservation and Polling Multiaccess Protocol for Wireless Data Systems*," **International Federation for Information Processing (IFIP'96) 14th World Computer Congress**, Canberra, Australia. September 2-6, 1996.
- 2) T. Buot, "*Channel Allocation Algorithm for TDMA with Multiclass Users*," **IEEE Vehicular Technology Conference (VTC'96)**, Atlanta, Georgia, USA, April 28 - May 1, 1996.
- 3) T. Buot, "*Priority Schemes for Mobile Data Access Employing Reservation*," **IEEE World Wireless Communications Symposium**, Long Island, NY, November 1995.
- 4) T. Buot, "*Channel Allocation Strategy for Voice/Data TDMA Systems*," **4th IEEE International Conference on Universal Personal Communications (ICUPC'95)**, Tokyo, Japan. November 6-10, 1995.

List of Abbreviations

ACTS	<i>Advanced Communications Technologies and Services</i>
AMPS	<i>Advanced Mobile Phone Systems</i>
ARDA	<i>Asynchronous-Reservation Demand-Assignment</i>
ATDMA	<i>Advanced Time Division Multiple Access</i>
BER	<i>Bit Error Rate</i>
BTMA	<i>Busy Tone Multiple Access</i>
CBR	<i>Continuous Bit Rate</i>
CDF	<i>Cumulative Distribution Function</i>
CDMA	<i>Code Division Multiple Access</i>
CDPD	<i>Cellular Digital Packet Data</i>
CRA	<i>Collision Resolution Algorithm</i>
CRI	<i>Collision Resolution Interval</i>
CSMA	<i>Carrier Sense Multiple Access</i>
DAMA	<i>Demand-Assign Multiple Access</i>
DFT	<i>Delayed First Transmission</i>
D-TDMA	<i>Dynamic Time Division Multiple Access</i>
EPA	<i>Equilibrium Point Analysis</i>
ETDMA	<i>Enhanced Time Division Multiple Access</i>
FCFS	<i>First Come First Served</i>
FEC	<i>Forward Error Correction</i>
FDMA	<i>Frequency Division Multiple Access</i>
FLMPLTS	<i>Future Public Land Mobile Telecommunications System</i>
FLR	<i>Frame Loss Rate</i>
FLT	<i>Frame Lookahead Technique</i>
GPRS	<i>General Packet Radio Services</i>
GSM	<i>Global Systems for Mobile</i>
ICMA	<i>Idle Casting Multiple Access</i>
IDC	<i>Index of Dispersion for Counts</i>
IFT	<i>Immediate First Transmission</i>
ISMA	<i>Idle Signal Multiple Access</i>

ISDN	<i>Integrated Services Digital Network</i>
ITU	<i>International Telecommunications Union</i>
IWLAN	<i>Integrated Wireless Local Area Networks</i>
LAN	<i>Local Access Network</i>
LCFS	<i>Last Come First Served</i>
LSTF	<i>Longest Service Time First</i>
LTl	<i>Linear Time Invariant</i>
MAC	<i>Medium Access Control</i>
MMPP	<i>Markov Modulated Poisson Process</i>
MOS	<i>Mean Opinion Score</i>
MRN	<i>Mobile Radio Network</i>
MSAP	<i>Mini-Slotted Alternating Priorities</i>
MSAT-X	<i>Mobile Satellite Experiment</i>
NMT	<i>Nordic Mobile Telephone</i>
PDF	<i>Probability Distribution Function</i>
PODA	<i>Priority On Demand Assignment</i>
PRMA	<i>Packet Reservation Multiple Access</i>
PRN	<i>Packet Radio Network</i>
QAP	<i>Quasi-random Arrival Process</i>
Q-CDMA	<i>Qualcomm Code Division Multiple Access</i>
QED	<i>Quantised Exponential Distribution</i>
QoS	<i>Quality of Service</i>
RACE	<i>R&D for Advance Communications in Europe</i>
RAM	<i>Radio Access Mobile</i>
RAMA	<i>Resource Auction Multiple Access</i>
R-ALOHA	<i>Reservation - ALOHA</i>
RMA	<i>Random Multiple Access</i>
R-TDMA	<i>Reservation Time Division Multiple Access</i>
SAD	<i>Speech Activity Detector</i>
SAP	<i>Service Access Point</i>
S-ALOHA	<i>Slotted ALOHA</i>
SCARP	<i>Silence Contention Acknowledgment Reservation with Polling</i>

SENET	<i>Slotted Envelope Network</i>
SSMA	<i>Spread Spectrum Multiple Access</i>
SRUC	<i>Split Reservation Upon Collision</i>
TACS	<i>Total Access Communications</i>
TFA	<i>Transient Fluid Approximations</i>
TASI	<i>Time Assigned Speech Interpolation</i>
UMTS	<i>Universal Mobile Telecommunications Service</i>
VBR	<i>Variable Bit Rate</i>
VRRA	<i>Variable Rate Reservation Access</i>
WINLAB	<i>Wireless Information Network Laboratory</i>
WLAN	<i>Wireless Local Area Network</i>
WPC	<i>Wireless Personal Communications</i>

Symbols and Notations

λ	arrival rate
P_{SS}	Poisson arrival distribution
$\log_e$	natural logarithm
$E[.]$	Expectation
$\beta_{in}(.)$	Binomial distribution
$\lceil . \rceil$	round-off to the nearest integer towards infinity
σ	Probability of a user becoming active
γ	Probability of a user becoming idle
P_{BS}	Quasi-random arrival distribution
L	average message length
$\mathfrak{D}(n, x(t))$	n-fold self convolution of $x(t)$
I	information or traffic slots
R	reservation slot
A	acknowledgment slot
G_a	load of reservation slots
S_a	Throughput of R slots
$e^{(.)}$	exponential
S	normalised throughput
A_o	offered traffic load
G	offered load per slot
N	number of slots per frame
N_a	number of R/A slots per frame
τ	frame duration
$\min\{.\}$	minimum
$\max\{.\}$	maximum
$ErlB(x, y)$	Erlang Blocking formula for x load and y channels
$ErlC(x, y)$	Erlang Delay for x load and y channels
p	retransmission probability

$\tilde{B}$	estimated backlogged users
∞	infinity
C_p^k	packet capture probability
M	number users or terminals
r.v.	random variable
i.i.d.	independent and identically distributed
T_{TALK}	average talkspurt duration
R_{sp}	average talkspurt rate
T_{SIL}	average speech silence duration
T_{gap}	average speech gap duration
χ	speech transmission efficiency
QoSv	quality of service for video
α	activity factor
μ	departure rate of a server, departure rate of contention state
ζ	departure rate of the channel allocation state
Da	access delay
Dm	message delay
P_{dc}	contention packet dropping rate
P_{da}	channel allocation packet dropping rate
Φ	fairness
P_{succ}	probability of success
$rand$	uniform deviates
h	hangover
$\mathbf{P}$	transition probability matrix
π	state occupancy for single slot
Π	state occupancy for multislot
$\Pr\{x\}$	probability of x
Σ	summation
Π	product

Note: Listed are only the frequently used symbols and notations

to my love ones,

my wife Cherry Clarissa,

and our offsprings Guiseppe Kevin and Ted John

Preface

Personal Mobile Communications has changed the way people communicate. As a result, the development of wireless personal communications has attracted many researchers as this area is not only technology motivated but is also market driven. This thesis deals with the design and performance of multiaccess protocols for the air-interface of a wireless communications system. This is aimed towards advancing the techniques in TDMA multiple access technology to offer more flexibility to the various wireless services. The main constraint in the advancement of mobile communications are the problems mostly encountered at the radio interface. Whilst it is true that wireless communications has started a century ago during the invention of the wireless telegraph, researches on how to achieve the Shannon limits of the radio channel capacity are still undergoing. Some recent claims suggest that present technologies are approaching the maximum capacity limits through the advances in modulation, coding and receiver design techniques. This left us the task to developed a fast, more efficient, and adaptive multiple access protocols.

In *wireless personal communications* applications, there are two more criteria to enhance the capacity and utilisation of the radio resource. One is to employ an efficient resource allocation scheme to a number of users in the coverage area. This criterion is necessary to accommodate the unpredictable traffic activity of mobile users. With the use of appropriate multiplexing schemes, a channel utilisation near unity is possible. This is achieve by employing a multiple access protocol and a channel assignment algorithm. In any case, the protocol is a major requirement in multiuser mobile communications. The other criterion to enhance the capacity is the network architecture. This time, the users' mobility pattern is anticipated by employing an adaptive cell layout or by employing intelligent antennae to tract the user density in the coverage area. Using advanced techniques in cell layout and dynamic configuration, significant improvements can be achieved on the radio resource. Overlaying networks will soon be adopted in personal communications.

While there are lot of debates on the choice of "superior" multiple access technology, the nature of the Personal Communications industry is very much market driven such that both the spectrum and economic efficiencies are to be considered. This makes the flexibility and the scalability of the underlying technology given of more

importance. However, this should not refrain us to further investigate the strengths and weaknesses of competing technologies as future wireless communications will gradually evolve from the present available technologies. Another important point to consider in the choice of technology is to assess the network's overall performance of which includes the aforementioned criteria such as *channel capacity*, *teletraffic efficiency* and *network adaptability*. These address the prime problems in mobile communications which are noisy channels, random user activity and rapid user mobility.

Theodore V. Buot

University of Adelaide, Australia
2 August 1997



Chapter 1

Introduction

Since the invention of the wireless telegraph in 1897 by Marconi, wireless communication has undergone a series of remarkable change and has now entered a new era of personal communications. Wireless Personal Communications or WPC was yesterday's dream but it fast becoming to a reality. It is the ability to communicate with anyone, anytime and anywhere by any means (voice, data, video, etc.). The growth of the cellular mobile communications market in many parts of the world is phenomenal and has encouraged the participation of industry, regulatory bodies and research communities in the numerous activities relevant to this area. The excitement of this industry has been propelled by the demand for more capacity and new services coupled with the advancement in technology making wireless personal communications both a technology push and a market driven activity. As a result, there are many research opportunities in this area.

To realise a truly ubiquitous personal communication, there are existing problems that must be resolved. These are the problems concerned with the network *capacity*, *flexibility*, transmission *speed*, and service *quality*. From the network's point of view, network capacity and flexibility are the important factors in providing Personal Communication Services or PCS, so that both spectrum and economic efficiencies must be considered in the choice of a wireless access technology. The need for a flexible network is a result of the rapid changes in the nature of new wireless services. On the other side, the user or subscriber's demands are focused on the speed and the quality of service criteria. As speed is synonymous to the transmission bandwidth, capacity remains to be the major problem in wireless communications.

With the success of the many cellular interface standards such as TACS, NMT, AMPS, and especially GSM, the demand for mobile services rapidly increases choking the available spectrum to congestion. This condition forces many researchers to consider higher spectrum bands with more available bandwidth to provide room for newer and superior technologies. This is accompanied with the increasing usage of wireless data services and wireless Internet. Moreover, increasing quantities of

multimedia traffic in the fixed network will make wireless multimedia support inevitable. Furthermore, the recent introduction of many client-server services such as interactive and transaction style communications demands a mix of both narrowband and wideband wireless access. The varied topological nature of these services i.e. asymmetric traffic, carries an additional burden on the design of WPC networks. Consequently, these problems must be addressed in the design of multiple access or multiaccess protocols at the radio-interface of a WPC network. Although a WPC network also consists of a fixed network infrastructure, the capacity problem at the radio interface is characterised as spectrum (resource) limited while the fixed network side is only cost limited.

Presently, there are three multiple access technologies that are widely adopted in cellular mobile networks. They are the Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA) and Spread Spectrum Multiple Access (SSMA) in a form of Code Division Multiple Access (CDMA). These technologies can be briefly described as follows:

- FDMA - the entire spectrum is divided into frequency units of uniform width and each user is assigned a specific frequency band of fixed bit rate.
- TDMA[†] - the spectrum is usually grouped into smaller bands as in FDMA. Each FDMA radio channel is further split in the time axis into smaller channels called timeslots. Each user is assigned with a specific timeslot.
- SSMA - has two versions. One is frequency hopping SSMA where users jump from one frequency unit to another. The other is CDMA where users are assigned unique codes used to spread the input signal (information).

Numerous debates on the selection of these technologies had occurred since wireless communications has been commercialised [Viter92] [Cox92] [Abra94] [Eve94] [Frull94] [Rait91]. Recently, both SSMA and TDMA received some credits as candidate protocols for the 3rd generation wireless information networks. However, these claims need further research and investigation for the exact architecture of the candidate technologies to accommodate the future services. This thesis is concerned only with TDMA technology.

[†] Both FDMA and TDMA were known as fixed assignment multiaccess protocols. In the WPC context, they are referred to as multiaccess technologies which could employ DAMA type multiaccess protocols

In TDMA and FDMA, the improvements in terms of the capacity are concentrated at the physical layer design. However, the access flexibility is left to the multiaccess layer. The main capacity improvement derived at the multiaccess layer of TDMA systems is the provision of time division multiplexing only to busy users. Prior to the discussion on this principle, we digress to the definition of multiple access protocols.

The entire communication system is structured into a concept of layers. The International Standards Organization (ISO) has developed an Open Systems Interconnection (OSI) Seven Layer Reference Model for interconnecting communication systems [Tanbm89]. In the OSI seven layer model, the lower the layers are more related to the hardware and the higher layers to software. The MAC is a sublayer sandwiched between the Second Layer (Data Link Layer) and the physical layer. The MAC sublayer is responsible for the terminals to initiate a connection whenever it has information to transmit. The MAC sublayer utilises some of the multiaccess protocols [Klie75] [Tob80] [BerGal92] to allow multiple terminals or users to share a common channel.

In the communications jargon, protocol is defined as *"The set of rules and agreements among the communication parties that dictate the behaviour of switches, and the channel is the physical medium over which signals, representing data, travel from one switch to another"* [RomSi90]. In simple terms, the protocols manages the available resources (channels) according to the users (communicating parties) demand by some form of signalling prior and/or during the transmission of the actual information. Multiaccess is to enable multiple users to share a common channel. It often relates to the link access procedure which is referred to as Medium Access Control (MAC) protocol. In the context of flexibility pertaining to the protocol's ability to handle varying traffic and number of users, multiaccess protocols are grouped into three types. They either belong to the *random access*, *fixed assignment*, or *demand assignment*.

In the random access protocols a user is allowed to transmit at its own will. Every time it has a new packet to send, it can transmit without explicit scheduling. Hence, packet collisions may arise resulting in a lower channel utilisation. Its advantage is its short access delays at lower load regions. In many cases, the main interest in the study of random access protocols is in the algorithms used to resolve the conflict between

contending users upon initiating a transmission in order to achieve higher throughputs. The algorithms used to resolve such conflicts are also called Collision Resolution Algorithms (CRA) or Random Multiple Access Algorithms [Tsyb85].

A fixed assignment protocol offers no contention but each user is limited to the channel permanently assigned to it and hence limited in speed. The main disadvantage of this type of protocol is that the fixed user environment is unsuitable for most applications. To increase the flexibility of this type of protocols, a combination of random access and fixed assignment is sometimes necessary. Protocols using this configuration are categorised as Demand Assignment Multiple Access or DAMA. A DAMA protocol is capable of supporting multiple users (usually much greater than the number of available channels) by allocating channels only to users that are busy. The protocol's flexibility is also increased through the use of a resource allocation management to satisfy every user's demand.

The need to provide a multiple access is due to the topological nature of a personal communications network. It is characterised by a group of independent users sharing a common pool of channels. Since every user is only active by only a small percentage of time, a single channel has a potential to support a large number of users. As an example, voice service with an activity factor of 0.02, one channel can be shared by a maximum of 50 users. Currently, TDMA and FDMA achieved this using the DAMA architecture. However, the current implementations of both TDMA and FDMA have slow access speeds and are limited in transmission rates. Thus further improvements are necessary.

In order to identify the suitable architecture and performance of WPC networks, the understanding of the nature of the future services is a must. In particular, the traffic characteristics as well as the performance criteria must be determined. Future services as classified into three main categories namely, voice, data and video (multimedia). Voice service currently dominate the present cellular systems and is expected to continue in the future. Data services are getting more importance. They are further classified into different types (i.e. interactive, FTP, messaging, browsing, etc.). The data services requires more flexibility than the current voice service because of the varied transmission characteristics such as synchronous or asynchronous. Apart from data, video traffic is much more difficult to support due to its large bandwidth requirements. Video sources have two distinct characteristics which are the high bit

rates and the large bit rate variation. With regards to performance requirements or quality of services (QoS), voice and video quality are measured in terms of information loss rate (i.e. speech frame dropping rate and video frame loss) and data QoS is measured in terms of its delivery time. Since data services requires an error-free end to end delivery, the addition of an error recovery mechanism in the design of the multiaccess protocol is necessary.

Efficient integration of the aforementioned services is a new task in the multiaccess protocol design. The on-off traffic characteristics of both voice and data services as well as the bit rate variation of video can be exploited in order to achieve higher multiplexing gains. The first requirement of an integrated system is a high speed channel. Research towards bit rates in the region of 1 to 2 Mbps (Megabits/second) are found to be achievable on the microcellular environment [RACE95]. Accompanied by the advances in video encoding, at this bit rate low speed variable bit rate video can already be supported. The second requirement is the use of *uniform traffic units* in order to multiplex the different services in a common radio channel. *Packet switching* technology has been widely accepted as an appropriate architecture for integrated services. It is known for its flexibility in supporting mixed or multimedia services currently implemented in the fixed network. Packet switching has been considered for the 3rd generation wireless networks [Goodm89].

A packet based wireless network requires a connectionless mode multiaccess protocol at the radio interface. In contrast to the circuit-switched oriented protocols adopted in most existing standards, connectionless oriented protocols can achieve faster link establishments and release by employing appropriate headers (addressing) in each packet. This scheme can achieve better multiplexing because the channels can be assigned dynamically at a faster rate through the use of appropriate channel assignment algorithms. Existing multiaccess protocols used in the data networks can be further developed to be suitable in the packet-switched wireless networks. Reservation based TDMA protocols are being considered in this thesis.

The main requirements in order to support a packet based wireless access at the multiaccess layer are 1) fast channel access and release, 2) the support of variable transmission rates and 3) fast error recovery for data services. The fast channel access and release is required in order to achieve higher multiplexing gains. For voice traffic, a talkspurt level multiplexing requires an end to end delivery of voice packets in the

region of 30 milliseconds (ms). Therefore the reverse or uplink access budget must be in the region of 10 ms delay. This is the minimum access speed required for the multiaccess protocol. Video services using the VBR encoding also changes the bit rates according to the sampling rate (around 100 ms). Thus the channel request and release changes at this interval. Fast channel reservation scheme is also required for the fast channel request and release of VBR video. Furthermore, interactive data and bursty data services requires faster message delivery and therefore requires multichannel reservation. Reservation-TDMA or R-TDMA protocols qualifies on these requirements but further improvements are required to improve its design.

Why TDMA?

Future wireless services will offer multimedia traffic and will require greater flexibility in the access protocol. Multimedia services are characterised by variable bit rate and varied QoS requirements. TDMA achieved this more simply than other technologies by allocating a variable number of slots to each user according to their requirements. The variable bit rate is achieved using a single which is the main advantage of TDMA. Very low rate services can also be supported by employing sub-multiplexing on the TDMA channels. This capability is also enhanced if packet-switched access is employed. TDMA also employs digital technology at the physical layer which does not limit the use of many capacity enhancement techniques like digital modulation, source coding, error control coding and other signal processing techniques. In addition, TDMA can implement frequency hopping[†] in order to improve its resistance to fading. The reliability of TDMA for data services can also be enhanced if variable coding rate is employed. In this case less tight power control is required. Additionally, a more stringent characteristics of TDMA is the initial synchronisation that the TDMA terminals can possess so that upon the arrival of their packets, the channel request can be executed immediately without first initiating a synchronisation procedure. This enables TDMA to allow fast access for delay sensitive services such as voice. And lastly, TDMA is a proven technology where in the channel access procedure, many collision resolution algorithms are suitable for slotted channels which may also be applicable in a TDMA. A sample TDMA system is shown in Figure 1.1.

[†] Frequency hopping is to allow a terminal to jump from one frequency or radio channel to another. This is one type of spread spectrum. Its advantage is to build resistance to fading since the effective channel quality in every terminal is an average of all the frequencies.

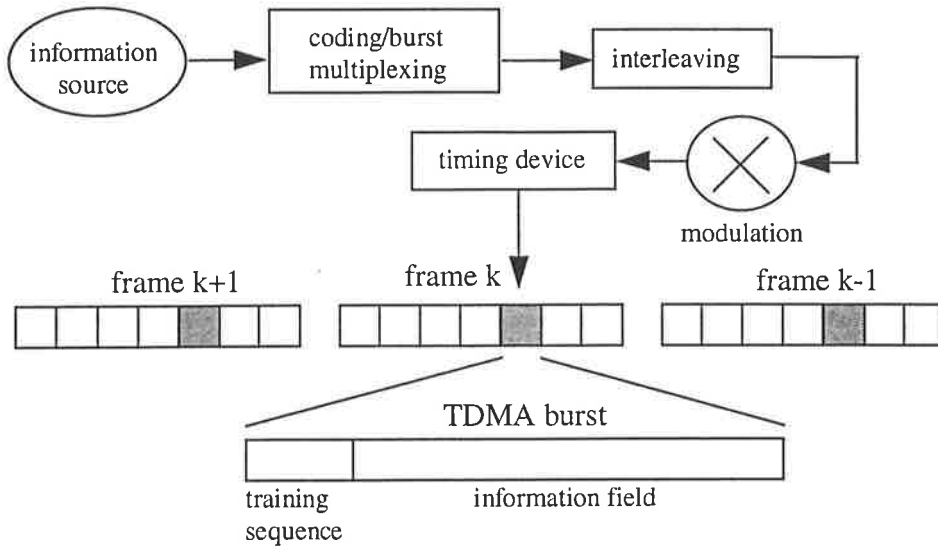


Figure 1.1 The TDMA System

Like any other technologies, TDMA has some disadvantages. For one, a TDMA channel has a speed limitation in a mobile environment. This is due to its less resistance to multipath fading as compared to SSMA. However, as mentioned earlier, recent claims suggests that bit rates above 1 Mbps are achievable. The other disadvantage is its smaller coverage area compared to FDMA and CDMA which results to a deployment of more base stations. However, future base stations are smaller (i.e. smaller antenna for higher frequencies) especially in a microcellular environment. The other main disadvantage of TDMA is the need for frequency planning as it limits the scalability of TDMA network during implementation. Dynamic channel assignment is the way to avoid this problem and is being included in the implementation of currently available TDMA standards.

The Reservation TDMA protocols

Reservation Protocols are DAMA type protocols which combine contention based protocols with time division multiplexing. The idea of using Reservation TDMA (R-TDMA) for WPC is borrowed from Crowther's Reservation ALOHA for packet radio networks. In contrast to the R-ALOHA, R-TDMA employs an explicit reservation which requires the transmission of reservation packets prior to the actual message. The reservation packet or request packet carries the user information and its necessary signalling information (destination address, amount of slots required, user priority, etc.) to inform the central control of its busy status. Once successful, a form of

acknowledgment is sent to the successful user. The user immediately reserves a channel if available and transmits its information or waits until a channel is available. As mentioned earlier, the main requirements to support a WPC based on R-TDMA are the fast channel access, variable rate transmission and a fast error recovery. Fast access is one requirement for packet-voice traffic in which a maximum access delay in the order of less than 10 ms is required. Voice packets that exceeds this delay are dropped and are counted against the voice quality measure. The need for variable rate transmission is for delay sensitive data services and the support of video transmission. The multislot reservation allows up to a maximum of reserving all the slots in one TDMA radio channel. Then the fast error recovery is essential for error sensitive data services as transmission errors is common in radio channels.

The R-TDMA requires a slotted channel structured into frames. Each frame consists of a number of basic channel units called timeslots. Each timeslot repeats in a cyclic fashion in every frame so that a user that reserves a particular timeslot can transmit in every frame. The size of the timeslot varies (usually in hundreds of bits) but is often optimised for voice transmission. The use of slotted channels is advantageous in orthogonal systems in order for the users to achieve initial synchronisation and thus reduce the amount of training sequence. Because of the difficulty in synchronisation, training sequence is still required in every timeslot and thus offers an overhead to R-TDMA protocols. Guard time between frames is also required in most cases. R-TDMA varies slightly from the Reserved Idle Signal Multiple Access (R-ISMA) because ISMA protocols does not require framing. It treats the whole channel as a single timeslot resulting in a M/M/1 queue configuration while R-TDMA has an M/M/N configuration. The access protocol for R-TDMA systems has a higher degree of independence to the physical layer when compared to CDMA systems where the multiaccess capability is tightly coupled to the physical layer. However, the capture effect of the contentions mechanism used in R-TDMA will inherit some Quasi-Orthogonal[†] properties. Some examples of protocols under the R-TDMA classification are the Packet Reservation Multiple Access [Goodm89], Advanced TDMA (ATDMA) [Urie95], Enhanced TDMA (ETDMA) [LiMer94] and many more.

[†] Quasi-orthogonal protocols allows a probability of reception of the transmitted information from one terminal even a collision with another terminal had occurred, a principle used in CDMA. Refer to Chapter 2 for more discussions.

Outline of this Thesis

This thesis deals with the design and performance of Reservation-TDMA protocols for wireless personal communications. Generic traffic models were derived and were used to test the performance of the protocols. The Advanced TDMA was carefully studied and its variants were introduced. Simulation and Approximate analyses were used in order to assess the performance of the proposed protocols.

The main text of this thesis is divided into five major parts. Chapter 2 is devoted to literature review starting from the ALOHA protocol and then discuss some of its variants. The differences between each type of protocols were also discussed. In particular is the description of reservation protocols as they evolve towards integrated protocols. Some key descriptions of reservation protocols were also described. In Chapters 3 & 4, some gaps in this area are being carried out in order to provide an efficient evaluation of R-TDMA protocols. Chapter 3 covers the source traffic statistics. It covers the two important parameters of source traffic which are the arrival process and the message length statistics. In addition, the speech packet generation is considered and a model to generate synthetic speech packets is described. In §3.5 the effect of the different values of speech hangover is examined which is later included in the optimisation. Then some generic models for video traffic is also discussed in §3.6.

Chapter 4 is about the analyses methods for reservation protocols. Most of the methods are the once being used in this study. In §4.2, the different analyses for ATDMA is described. In §4.3 a general approximate model for slotted random access is devised and was tested in the ALOHA protocols. For the channel allocation queue, a two-moment delay approximation is introduced (§4.4). It described a procedure to calculate the delay distribution of a discrete queue. The last portion described the basic principles of S-G analysis and EPA/TFA analysis. Then the stability of ATDMA is considered. These techniques were all applied to the Advanced TDMA protocol.

Chapter 5 deals with the R-TDMA design and performance as a candidate protocol for WPC. All throughout this thesis, we used ATDMA as the test protocol because it is well known that most R-TDMA protocols have comparative performance. In §5.2 a performance evaluation of ATDMA with voice traffic is considered. It includes the effect of hangover values of speech talkspurts in the optimisation. Then in §5.3 a reservation policy for data users is proposed which achieves some degree of fairness. A round robin reservation is modified using a flow control. After knowing the

limitations of ATDMA, some enhancements were introduced in §5.4 which considers the effect of the S-ALOHA retransmission probability and eventually its effect on the protocol's stability. The effect of the physical layer like packet capture and random bit errors were also investigated. Then two variants of ATDMA were proposed which are the R-TDMA with dynamic frame configuration and the Integrated voice/data ATDMA. In §5.5, the inclusion of prioritisation at the random access level is being considered. A multi-priority collision resolution algorithm based on stack algorithm is proposed. Later it is applied to ATDMA in §5.6 and a performance analysis is developed. Then in §5.7, a novel protocol which combines random access, polling and reservation is developed which attains higher flexibility and stability for varying traffic. The proposed protocol, SCARP is also analysed.

In Chapter 6, the resource allocation problem in multimedia wireless access is considered. First, it discussed about the problems in multimedia systems which are centred on the resource allocation and QoS maintenance. First a performance measure is derived which suggested a two moment analysis for multislot systems. Then performance analysis for multislot resource allocation is developed using a Birth and Death Markov chain and Discrete Markov Analysis. Then in §6.2, the implementation issues of multislot reservation is considered and some procedures were proposed evaluated. Later in §6.3 the performance of a multislot reservation for different traffic mix is evaluated. Then the effect of a time varying channel to multislot reservation is evaluated in §6.4. It shows the advantage of multislot reservation in mitigating the bit rate reduction as an effect to the use of lower coding rates during poor channel conditions. Lastly, the transmission of variable bit rate video is discussed in §6.5 and a video coding scheme is introduced. Then followed by a conclusion.

Original Contributions of this Thesis

Some parts of this thesis are published in conference proceedings and journals as shown in the *List of Publications*. Throughout this work, the contributions are in the design improvements of multiple access protocols for WPC as well as in the performance analysis. After a careful investigation of the characteristics of a reservation protocol, some protocols were proposed which exhibit good throughput-delay performance and accordingly, their analyses were developed.

In the improvements of the quality of service based on the nature of the source traffic generation, two new capacity optimisations were identified. First is for the voice traffic in which the maximum capacity of R-TDMA with speech multiplexing requires a selection of the optimal hangover value which has never been considered in all other analyses of PRMA and ATDMA.. The other aspect of source coding is that of VBR video. It suggests that QoS improvement can be achieved by providing an association between the multiaccess layer and the video encoder in order to negotiate for the optimal QoS. This is in contrast with most of the studies in wireless video transmission which only focus on the channel coding.

In the area of multiaccess, a simple but accurate method to calculate the delay distribution of S-ALOHA with random retransmission was developed. For the channel allocation queue of R-TDMA, a Markov analysis for the delay distribution was applied. In addition, the stability of ATDMA was investigated. Then the analysis of ATDMA using Transient Fluid Approximations was developed. The unique characteristics of the analysis is the combination of TFA and M/M/N/p queue in order to include the channel allocation queue in the analysis. The technique achieves a high degree of accuracy.

In chapter 5 the new protocols that were proposed are the, R-TDMA/DFC, Integrated ATDMA, Multipriority Stack, Integrated ATDMA/Stack, ATDMA/Tree and the SCARP protocols. These are accompanied with their corresponding performance analysis and/or simulations. The multipriority stack algorithm is designed to improved the stability the system while providing an excellent prioritisation. In addition, a method to improve the performance of R-TDMA by exploiting the capture effect and its relation to the topology of the network is being identified.

In chapter 6, the multislot reservation is carefully studied. The first contribution is in the time domain throughput-delay analysis of multislot R-TDMA based on Markov models. The second contribution is in the provision of prioritisation to multislot R-TDMA by exploiting the multislot reservation capability. Then the investigation of the effect of code rate switching together with multislot reservation indicates a performance improvement to data services. Other minor contributions will be identified along the text and most of these contributions are concerned with the design principles in improving the performance of R-TDMA.

Chapter 2

Multiaccess Protocols – a Brief Review

Since the pioneering work of Norman Abramson in 1970, many protocols have been proposed and implemented both in the wired and wireless networks. Our concern in this study are the multiple access protocols used as a mechanism of allocating channels to busy users in a wireless environment. There are several classifications of multiple access protocols [RomSi90] [Pras96] [Wu94a], but a more general way to classify them is shown in Figure 2.1. According to Abramson [Abra94], protocols are either orthogonal or quasi-orthogonal. Quasi-orthogonal protocols are governed by probabilities reception caused by overlapping transmission on a common radio channel. Presently, direct sequence spread spectrum multiple access protocols are examples of this protocol. ALOHA with high capture characteristics is also in my view an example of this type of protocol. Meanwhile, orthogonal protocols assume that any overlapping transmission will result in a complete loss of information by any colliding users. Most protocols are under this category. Orthogonal protocols are further classified into contention based or contentionless. By its definition, contentionless protocols use access scheduling or fixed assignment. The access scheduling is achieved by polling the users individually or interrogating its status by means of token passing. On the other hand, contention based protocols allow collision because a channel is shared by multiple users. Examples of this protocols are also shown in Figure 2.1. The Reservation based protocols are combination of contention based protocols and fixed assignment protocols.

The study of multiaccess protocols was pioneered by N. Abramson at the University of Hawaii for packet radio transmission of multiple computer nodes to a central control. The network was called the ALOHANET and a simple multiaccess protocol called ALOHA was developed [Abra70]. The ALOHA protocol being a *random access* type was considered because of the non-steady nature of the packet radio traffic and the network topology of the ALOHANET. The success of the ALOHANET played an important role in the rapid development of satellite and digital Packet Radio Networks (PRN).

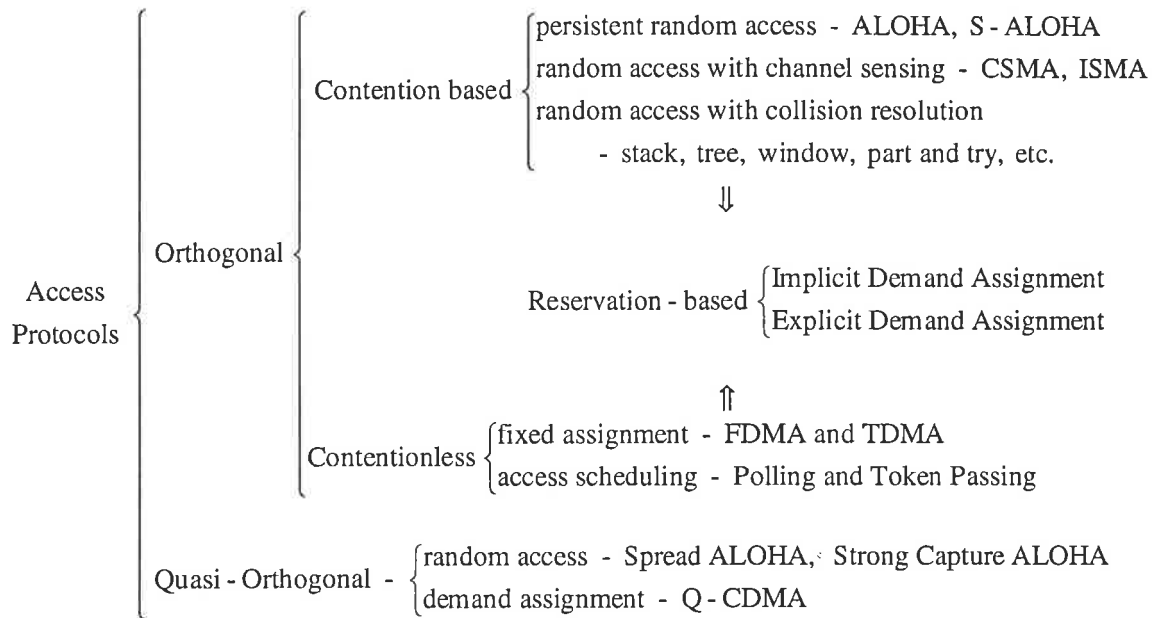


Figure 2.1 Classifications of Multiaccess Protocols

The efficiency of the ALOHA protocol is relatively poor when used in steady traffic conditions. Therefore, the need for a more efficient protocol for steady traffic naturally led to *fixed assignment* protocols. When the number of accessing terminals is quite few and the traffic from each terminal is steady (continuous), the *fixed assignment* protocols are more appropriate. It was at this point that TDMA and FDMA were considered as an alternative to the ALOHA system. It is achieved by allocating smaller bandwidth (frequency or timeslot respectively) permanently assigned to each terminal so that no contention exists whenever a terminal has information to transmit. Both FDMA and TDMA were used in satellite networks. The performance of such protocols depends heavily on the traffic characteristics. But the main drawback of these protocols is that the number of nodes or terminals that can be supported is fixed and can only be increased with additional infrastructure (more channels).

A more flexible architecture is to accommodate a variable number of users within the multiaccess environment. This is achieved by allocating resources only to users that are ready to transmit known as the Demand Assigned Multiple Access architecture or (DAMA). The DAMA architecture requires some allocation of resources to either 1) schedule the transmission by some form of signalling or 2) use for sending channel request signal in a random access mode. The DAMA concept is implemented in both

satellite and packet radio networks. In cases where the traffic is varies rapidly, such as in personal communications, a compromise between the DAMA and the random access is necessary. Thus, hybrid protocol architecture is one of the means of achieving it.

In this chapter, a review of these different types of protocols is presented alongside their properties. We also review the recent developments of protocol design for Wireless Personal Communications. More detailed descriptions of the history and developments of access protocols are also discussed in much of the literature [RomSi90] [Tanbm89] [Stall85] [Li87] [Abra94].

§2.1 ALOHA and Random Access Protocols

There are two main random access protocols, the ALOHA for orthogonal access and the random access Spread Spectrum Multiple Access (SSMA) for quasi-orthogonal access. The SSMA has two implementations which are the CDMA [Gil89] and the Spread ALOHA [Abra94][Abra96]. We first describe the behaviour of ALOHA protocol since it is the basis of the study of multiaccess protocols. In fact ALOHA is a family of protocols using different collision resolution procedures to improve its performance.

In the ALOHA protocol the terminals transmit their messages in packets of fixed length. A busy terminal or user transmits its packet immediately after it is generated. Since the terminal transmit independently of each other, two possibilities could eventuate for the transmitted packet, either it is successfully transmitted or it is garbled due to collisions. A collision occurs if two or more users transmit within a time frame such that a portion of their packets collide. Users that experience collision will know the status of its transmission either by listening during the transmission or preferably from the central control through a feedback channel. After knowing the status of the transmission, collided packets are retransmitted at a random time and the procedure is repeated until the packet is transmitted successfully. Users that undergo collision are said to be backlogged until the packet is successfully transmitted. Backlogged users assume that they cannot generate new packets until the collision is resolved. The maximum throughput of the ALOHA protocol is $1/2e$.

§2.1.1 Improvements of the ALOHA Protocol

The ALOHA protocol suffers from low throughput [Kob77] [Tsyb85] [Sheik90] and instability [Carl75] [PlaL90]. An improvement to the ALOHA was introduced by Roberts a few years after it was introduced. The protocol is called *Slotted ALOHA* (s-aloha) [Rob75]. Instead of using a continuous channel, a slotted channel is used where each packet information is equivalent to one slot length. The main idea of the Slotted ALOHA is to reduce the vulnerable time of a transmitted packet against collisions since packets are transmitted only at the beginning of each slot. By reducing the vulnerable time to half that of the Pure ALOHA, the throughput of a S-ALOHA is 1/e or twice that of the original ALOHA or *Pure ALOHA* [Klie75] [Git75] [Tsyb85] [Ben94]. Nevertheless, S-ALOHA still suffer from stability problems and requires some access control procedures [LamKl75].

Both the Pure ALOHA and the S-ALOHA protocols are the basis for the study of random access protocol. The basic assumption of ALOHA applies to most multiple access protocols which are:

1. Multiple Independent Users or Terminals
2. Poisson Arrival
3. Shared Channel (continuous or slotted)
4. Immediate Feedback for the status of previous transmission
5. Immediate First Transmission (non-blocked)
6. Random Retransmission after a Collision

As mentioned earlier, the main characteristic of a random access protocol is in the management of backlogged users governed by the CRA. There are three CRA's that are widely used in the ALOHA protocols. They are the 1) *fixed retransmission probability* 2) *binary exponential back-off* by Mikhailov and 3) Revist's *pseudo-bayesian algorithm* [BerGal92]. Both the binary exponential back-off and the pseudo-bayesian algorithms exploit the feedback information to arrange the retransmission of backlogged users. As a result they outperform the fixed retransmission probability in terms of the throughput/delay and stability¹ characteristics. Both algorithms require a binary feedback (*collision* or *no-collision*) to update the retransmission parameters. The binary exponential back-off listens to the result of its own transmission and count the number

¹ Multiaccess protocols are subject to instability where the probability of success approaches to zero. This occurs when the increase of the number of backlogged users is faster than the success rate.

of collisions while the pseudo-bayesian algorithm requires to listen the feedback channel continuously to update the number of backlogged users estimation.

§2.1.2 Other RMA Algorithms

In addition to the CRA algorithms mentioned, there are other types like the Stack Algorithm [Tsyb85], Window Control Algorithms [PaKaz89], Tree Algorithms [Cap79] [BerGal92] and the Part and Try or First Come First Served Splitting Algorithms [Tsyb85]. James Massey further classified Stack algorithm into Blocked Stack and Non-blocked Stack Algorithms. There are also meshed stack algorithms for multiple independent stack algorithms. For the Tree algorithms, there is a Static tree and Dynamic Tree Algorithms [Cap79]. The mentioned CRAs have varied performance characteristics and degrees of complexity. However, there are limitations on the amount of complexity in a CRA in order to be implemented in the high speed wireless access environment due to synchronisation, signalling and other impairments. It is also evident that when we consider reservation techniques, the need for very complex CRAs is no longer of much importance.

§2.1.3 Random Access Protocols with Channel Sensing

Another family of protocols that performs better than the ALOHA protocols are the protocols that employ carrier sensing for access control. The main principle of channel sensing is for the terminals to *listen first before transmitting*. The first of this type of protocol is the Carrier Sense Multiple Access (CSMA) [KliT75]. The CSMA protocol behaves similar to ALOHA except that a terminal listens first whether the channel is idle or another terminal is transmitting. Once an idle channel is detected, it transmits immediately or after some random time. Like the ALOHA protocols, collisions may result. Collided packets are retransmitted according to the RMA Algorithm. There are also many versions of CSMA for improving the collision resolution. They are CSMA with collision resolution (CSMA/CR) and a CSMA with collision detection (CSMA/CD). A slotted implementation of CSMA also exists.

The performance of CSMA is better than that of the ALOHA protocols (maximum throughput around 0.6). This of course assumes that all terminals are able to detect the status of the channel. Due to the nature of the radio environment, it is impossible for a terminal to listen to all other terminals which will result to a degradation of the system performance. This is termed as the *hidden terminal problem*

in random access protocols [KliT75]. A solution to this problem is to provide a separate broadcast channel from the central base station to transmit a busy tone indicating the status of the channel. The protocol is called Busy Tone Multiple Access (BTMA) [TobK75]. A digital version of such procedure is also implemented in CDPD networks known as the Digital Sense Multiple Access (DSMA).

In contrast to BTMA, another method to resolve the hidden terminal problem in CSMA is by broadcasting a status signal on the forward (downlink) channel whenever the channel is idle. In Idle Signal Multiple Access (ISMA) [MukF81], idle signals are transmitted periodically at intervals greater than twice the propagation delay and processing time whenever a channel is idle. A terminal which has packets ready-to-transmit listens to the status signal. After identifying an idle signal, it transmits after the idle signal. Once a collision occurs, an idle signal will still be broadcast by the central control. If no collision occurs, then the central control will stop transmitting the idle signals. The ISMA and the busy tone has practically the same performance and they differ only in the assumptions. The ISMA protocol has also many versions, Slotted-ISMA and ISMA/CD [Wu94a].

§2.1.4 Spread Spectrum Multiple Access

Spread Spectrum techniques have evolved from the military applications to personal communication. It is a widely accepted technology much superior to that of TDMA and FDMA [Purs87] [Gil89] [Khon95] [Torr92]. With the capability of multiaccess, SSMA is also a random access protocol in the sense that terminals are competing on a shared channel with probabilistic transmission. Spread Spectrum in the personal communications applications has two versions of implementation. First is the Spread ALOHA [Abra94] which is a truly random access protocol in the sense that users are competing on a shared channel (Single Code SSMA). The other implementation is the CDMA as previously mentioned and it provided multiaccess capability since users are assigned with unique codes (paired-off fashion). Both Spread ALOHA and CDMA may co-exist in a single network as a DAMA implementation of SSMA. A method to increase the bit rate of CDMA is to combine CDMA with Orthogonal-FDMA by employing multiple transceivers in each mobile terminal [LiG95]. This is another area to consider in protocol design.

§2.2 Fixed Assignment and DAMA Protocols

In cellular wireless communications, voice service is the dominant traffic. The choice of the DAMA scheme for these networks is due to the steady traffic characteristics of speech. The idea of using TDMA and FDMA for cellular networks was borrowed from the scheme used in satellite networks. As they are well understood technologies, their implementation in the cellular networks did not face many problems. The concept of using TDMA and FDMA for satellite networks was presented simultaneously in the 1974 IEEE Electronics and Aerospace Systems Convention, a few years after the ALOHANET was established. They were conceptualised based on the assumptions that most satellite networks carry long distance telephone traffic from a fixed number of earth stations. With the limited flexibility of TDMA and FDMA systems, the DAMA architecture naturally evolved. Since the request channel for the DAMA protocols are often ALOHA or S-ALOHA for slotted systems, these protocols were classified as Reservation-ALOHA. Examples of these protocols are defined in [Tanbm89]. Both TDMA and FDMA are candidates for reservation protocols by using an explicit reservation (explicit demand assignment). Examples of such protocols for satellite networks are the SENET, PODA and the MSAT-X as described in [Li87].

Aside from the combination of random access and fixed assignment to provide a DAMA architecture, a scheduled transmission can be achieved by means of a polling or token passing protocols. In roll-call polling, each terminal is polled individually until a busy terminal is identified by the central control. In this case, only busy terminals are allocated with a channel for the transmission of its packets. Once a terminal has transmitted its packets, the channel is seized back by the centralised control and begins the polling cycle again. An alternative to this procedure is the use of token passing. In this protocol, a token in a form of signalling information is passed from terminal to terminal until a busy terminal will hold the token and begin its transmission. Once a token is held by a busy terminal, all other terminals are not allowed to transmit. After the transmission is over, the token is passed to the next adjacent terminal. Both the token passing and the polling offers no collision during transmission but has a relatively longer access delay compared to the reservation protocols so that most of their applications are in high speed LANs.

§2.3 Protocols for Future Wireless Communications

Challenges in the design of access protocols for WPC is due to the unique characteristics of the multiaccess environment, which may have the capability to be configured both as a Packet Radio Network (PRN) and as a Mobile Radio Network (MRN). Therefore, it must be capable of handling data packets as well as voice calls. The major criteria to meet both traffic types are good throughput/delay characteristics and short call set-up times at the radio interface. Most protocols that fit these criteria are reservation based protocols.

§2.3.1 Packet Voice Communication

One of the significant improvements in digital mobile radio is the packetisation of speech in order to allocate channels only to users that are busy. This is achieved by the use of speech activity detectors (SAD) in the vocoder. This technique enables TDMA or FDMA systems to multiplex users on a talkspurt level and hence increase the potential number of users that can be supported. If talkspurt multiplexing is not utilised, the use of SAD can reduce the co-channel interference into the adjacent cells. This scheme is being adopted in GSM [GSM 6.31] called the discontinuous transmission. In SSMA, it has a dramatic improvement on the capacity because the use of SAD will decrease the multiple access interference (MAI) within the cell [Gil89] [Khon95]. In TDMA systems, the exploitation of the on-off patterns in speech was first introduced in Packet Reservation Multiple Access (PRMA) [Goodm90]. It was an extension of the application of Time Assigned Speech Interpolation (TASI) in undersea cable for long distance communications. The achievable multiplexing gains in packetised speech is obtained in [Yue94].

Voice traffic (speech) which is generated from two or more speaking parties is traditionally classified as a continuous type of traffic in the analogue network. With the advent of digital technology, voice is sampled and coded into digital units called frames. The frames are of uniform lengths (ie. 20,10 or 5 ms) which represents the instantaneous samples of speech. When a speaker is active it generates a series of frames called talkspurts. When inactive, no talkspurt will be generated and the speaker is said to be in a silence state. Thus, the traffic for a single speaker is said to be periodic. Hence, a channel is active less than half of the time.

Modelling of the On-Off patterns of speech was pioneered by P.T. Brady [Brad69]. It was discovered by Brady that the voice activity factor is only 44%. Brady's model was too conservative since it did not include the intersyllabic gaps less than 200 milliseconds (ms) by employing a fill-in operation (concatenation of talkspurts separated by less than 200 ms of silence). It was confirmed in [Grub82] that even in monologue speech a channel is still less than 60% active if no fill-in or hangover (appending additional frames after the last speech frame) operation is employed. Later, Lee and Un [LeeUn86] confirmed a more accurate result which indicated a 27% voice activity factor for no fill-in or hangover. This result is achievable in mobile systems employing fast activity detectors.

§2.3.2 Reservation based Protocols for Packet Switching

After W. Crowther and his colleagues introduced the concept of channel reservation (Reservation-ALOHA) for packet broadcast channels, several variants of their protocol emerged as candidate protocols for the third generation TDMA based wireless networks. The idea was to develop a packet based wireless access to enable efficient integration of speech packets and data packets. Reservation seems to be the appropriate technique because it can provide many types of connections (packet or circuit switched). After scanning through the literature, there are two implementations of explicit reservation protocols. They are the two-step and the three-step reservation access [Buot95a]. The two-step reservation process involves a contention phase and a reservation phase in the channel access procedure. On the other hand, the three-step reservation process involves a contention, channel allocation and a reservation phase in the channel access procedure. These protocol types differ slightly in the frame structure. The protocols widely considered for the future wireless systems are as follows

a) Crowther's Reservation - ALOHA ^[Lam80]

The R-ALOHA protocol is a S-ALOHA with reservation implemented in a TDMA scheme. It consists of a number of terminals competing for a fixed number of timeslots on a frame by frame basis. Users are allowed to reserve one or more slots per frame (multislot). In the R-ALOHA protocol, the slots are either reserved or free. A slot is said to be free if it is empty or a collision had occurred in the preceding frame. A reserved slot means a packet was successfully transmitted in the preceding frame. Users that are reserving a slot(s) in the current

frame will continue to reserve the slot(s) in the next frame provided they have ready-to-send packets. Newly active users that are willing to transmit will contend on any of the free slots the same as in the S-ALOHA protocol. Upon successful contention on a particular slot, the slot will be off limits to all other users. R-ALOHA has two versions. One employs an end-of-use flag and the other does not employ an end-of-use flag appended to the last packet. A detailed performance analysis of R-ALOHA is found in [Tas83] [Lam80].

b) PRMA [Goodm89]

The PRMA protocol is similar to R-ALOHA in the sense that a user has to transmit a reservation packet (RP) during the contention process. The RP consists of all necessary information attributed to the accessing terminal so that whenever a collision occurs, only the RP is to be retransmitted. The procedure of PRMA is as follows: 1) a user that has ready-to-send packets waits for a free slot in the frame. Once a free slot is identified, it sends a RP on that slot. The successful terminal will immediately reserved the slot and start transmitting in the next frame otherwise it retransmit in any of the free slots according to the retransmission probability. Users that reserves a slot will loose it reservation only when there is no more packet to transmit. PRMA is originally designed to meet the criteria of a packet voice system. It is described in [Goodm89] [Goodm90] and analysed in [Nand92] [Frull93] [QiWyr94]. PRMA for voice/data integration was also analysed in [Wu94b] [Gril93]. Some papers describing the PRMA control procedures are described in [QiWyr94b] [Frull94] [Bianch94] [Bolla95] [Wang94]. PRMA is geared towards the third generation wireless networks and the criteria is to offer ten times more capacity than the first generation systems with voice/data and video integration capability, service mobility and terminal mobility. Development of these third generation systems is carried out in WINLAB at Rutgers University, IMT2000 (formerly RACE, ACTS) in Europe, FLMPLTS of the ITU and in many commercial research laboratories.

c) Variants of PRMA

PRMA like the R-ALOHA is subject to instability [QiWyr94b]. This is due to the S-ALOHA contention process that leads to the bistable characteristics of the PRMA. To improve the stability and access speed of PRMA, a reservation multiple access protocol was proposed in [Mitro90] to provide separate slots for

control purposes. It improves the delay performance as well as stability compared to the PRMA. The same protocol is extended to voice/data integration [Mitro93]. Later it was modified in [Dunl93] [Dunl94] called PRMA++ or Advance TDMA (ATDMA) as an air interface proposal for the European third generation personal communications system. The ATDMA protocol is similar to the one in [Mitro90] where some slots are allocated for control purposes. The control slots in the uplink is intended for sending reservation packets and paired by an acknowledgment slot in the downlink together with a fast paging channel. ATDMA was analysed in [Dev93] and simulated in [Dunl95]. In parallel with the ATDMA project is another reservation protocol called Enhanced TDMA (E-TDMA) which is an enhancement of the North American cellular standard [LiMer94].

d) RAMA ^[Ami92]

Another recent scheme to provide fast channel assignment in wireless networks is the Resource Auction Multiple Access (RAMA) which is also considered for the third generation systems. The RAMA protocol is proposed in [Ami92] and evaluated in [Ami94]. In RAMA the available resources (free slots) are auctioned based on the users' ID. Users that are willing to transmit can join the auction process by listening the status of the auction through the feedback channel. The auction process is done one digit (eg. 10 digit ID) at a time until a single user will successfully win the auction process (ie. a user with the highest ID value). RAMA has claimed to provide faster resource allocation compared to the other access schemes but is seen to suffer from some synchronisation problems.

e) Dynamic-TDMA (D-TDMA) ^[Wils93]

The Dynamic-TDMA is similar to the R-ALOHA protocol except that minislots are provided at the front of each frame for random access purposes [Wils93]. Like in ATDMA, active users will request for a slot on any of the request slots (minislots). Successful users are allocated with a slot through the downlink acknowledgment. If collision occurs, a colliding user retransmits on any of the request slots until it succeed. In the case of a voice terminal, a maximum waiting time is observed otherwise a call is blocked. For data users, a random retransmission delay is employed.

f) Reservation-Busy Tone Multiple Access (R-BTMA)

BTMA is known for solving the hidden terminal problem in CSMA [TobK75]. Reservation technique is used to enhance the system capacity. R-BTMA also employs a slotted channel consisting of a data channel and a busy tone channel. The busy tone channel is used to transmit the status of the transmission of the data channel (i.e. idle, collision or busy) as well as acknowledgment. The procedure of R-BTMA is described in [Tabb92].

g) Idle Signal / Idle Casting / Inhibit Sense Multiple Access Protocols

These protocols are similar in nature which employ reservation and idle channel sensing. In Idle Signal Multiple Access, the polling signal is used to identify the status of the channel (idle or reserved). Users monitor the idle signal and if an idle channel is detected, ALOHA random access is used to contend for the channel. R-ISMA is a variant of CSMA intended for an unslotted radio channel. The Idle Casting Multiple Access [LeeUn96] and the Inhibit Sense Multiple Access [Linn94] are similar to the ISMA. An improve version of ISMA is the Reservation-ISMA or R-ISMA.

§2.3.3 Voice and Data Integration in TDMA Systems

The possibilities of data transmission over mobile channel was first investigated at the AT&T Labs [Karim86]. Since then mobile data has become another area of interest in wireless communications. A number of services are currently deployed for the niche market of mobile data like the ARDIS, RAM, MOBITECH, DataTac, CDPD, etc. These networks are mostly packet radio networks. The challenge for the next generation of mobile service is full integration of voice and data using packet-switch technology.

The design of voice/data integrated protocols for personal communications is again an extension of the studies conducted in the satellite and fixed networks. Numerous papers were published in this area such as [Fisch76] [Tuck88] [Grub83] [Srir83] [Falk83] [Hob83] [Apos93]. These studies have shown the possibility of employing a statistical multiplexing of speech or Digital Speech Interpolation (DSI) and data integration. One of the early works in the voice and data integration in mobile radio was demonstrated in [Mah83]. It covers both physical layer issues as well as access methods. The main concept in early works of voice and data integration was to transmit data in the silent gaps of a voice channel. This idea was carried out later in

[Stern90] [Akaiw92] [Mad91] [Mad92]. The main drawback of these type of protocols is the signalling overhead caused by detecting the silent gaps on a talkspurt level of voice channels. Secondly, a QoS for data is not guaranteed due to its dependence on voice traffic. The work in [Stern90] has similarities to that of integrated R-ISMA [Wu94c] except that the latter has a better signalling method to avoid collisions between data and voice packets. Recently, modified polling mechanisms were also investigated in [Lu94] [Cho95] for wireless networks claiming good performance for voice/data integration.

A more efficient scheme to integrate voice and data is in TDMA based protocols. Better efficiency is achieved by allowing voice and data users to compete for slots in the frame. Priority handling can also be implemented through the channel allocation process. These protocols are found in [Merak92] [Wies95] [Chang94] [Wang94]. Performance investigation for some integrated TDMA protocols are also found in [Wies95] [Cleary94].

Chapter 3

Teletraffic Source Models for R-TDMA

Teletraffic source models play a crucial part in the design of multiaccess protocols as they are tailored to specific applications or services. Hence, they also set the performance benchmarks of the system. This chapter is devoted entirely to the modelling and replicating of traffic sources generated by users within the wireless environment. The modelling of a traffic source depends on the performance evaluation method being employed (e.g. simulation or analytical). Usually, simple models are used in the analytical methods due to the amount of complexity that will be contributed if sophisticated models are used. In essence, realistic and sophisticated models are applicable in simulations methods and hence more accurate results becomes available compared to the analytical methods.

In the study of R-TDMA, three traffic types must be considered namely data, voice and video. Voice traffic refers to speech conversation while the video traffic refers to moving picture rather than still images. All the rest are classified under data so that data can be further classified into different types, priority and message statistics. Both voice and video traffic requires real-time transmission and are loss sensitive in which the information loss rate is a primary QoS parameter. On the other hand, data traffic is delay sensitive and error sensitive. Since wireless channels are naturally erroneous, the effect of transmission errors will also lead to an increase in transmission delays. This problem is aggravated by the limited speed of wireless channels

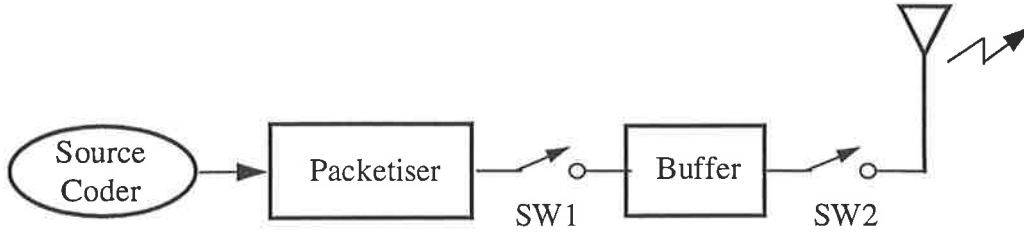
In all traffic types, the modelling requires two important parameters which are the *information arrival* and the *information size* statistics. The former relates to the process of occurrence of the messages while the latter is related to the message length distribution. Often, a one-moment statistics is required to describe these two parameters but sometimes higher moments are required (i.e. mean and variance) in order to determine the worst case and long term performance. Since R-TDMA uses a slotted channel, the interarrival time units must be selected to be appropriate in the performance evaluation. Interarrival time units in *slots* or *frames* are often used so that the random arrivals coincide at the beginning of each slot or frame.

Most sources are also equipped with message buffers right after the source coders. Traffic statistics are changed whenever a buffering capability is employed in the user terminals. This will then require buffering time-out threshold dependent to the application in order to reduce the delay incurred at the traffic source. A simple source model is shown in Figure 3.1. To differentiate the source models, we describe the behaviour of SW1 and SW2. The first two source models are due to [Lam80]. The third model is necessary to implement some transmission policies. In the first model, the message is generated at once and the source shuts off until the message is successfully transmitted. In the second model, the new packets may join the buffered packets and take part in the currently transmitted message. The terminal will attempt to transmit immediately after the arrival of the first message. In contrast, the third model waits some time to attempt transmission or if the buffer has enough packets to send whichever come first. The last model is more realistic as it can control the average message length and the access rate. The last model is essential in most protocol carrying non-real time information with bursty arrivals. The buffering mechanism employed in the last model provides a smoothing effect which will be considered in the analysis of delay.

In this chapter, some relevant teletraffic models for voice, data and video are proposed and evaluated. In particular, voice and video are taken with more emphasis due to the strict nature of their QoS. Schemes to improve the QoS in the multiaccess environment are also presented. In the subsequent sections, the generic arrival processes as well as the message size characteristics are discussed. Then followed by the discussion on the buffering of data traffic, modelling of speech traffic and the modelling of video traffic.

§3.1 Arrival Process

The most commonly used arrival process in communication networks modelling is the *Poisson* process because of its simplicity and inherent properties. A Poisson arrival process represents random arrivals from a large (infinite) number of users. For a random arrival process, it should be noted that the number of arrivals from t to $t+\Delta t$ is only dependent to Δt . This is referred to as the *memoryless property* of the process. To generate a Poisson arrival process, we are only interested in the interarrival time



Source Model	Idle Switch Position	Busy Switch Position
Single Message Arrival	SW1=On, SW2=Off	SW1=Off, SW2=On
Queued Users	SW1=On, SW2=Off	SW1=On, SW2=On
Partially Queued Users	SW1=On, SW2=Off	SW1=On, SW2=On*

Figure 3.1 Source Traffic Model

* SW2 switches "On" only when the number of packets in the buffer exceeds a specified threshold or when a certain time has elapsed since the arrival of the first packet. Other control procedures are possible.

distribution. The Poisson arrival is described as the probability of exactly x arrivals in a period T for an arrival rate λ and is expressed as

$$P_{SS}(x, T) = \frac{(\lambda T)^x}{x!} e^{-(\lambda T)} \quad (3.1)$$

The interarrival time distribution is then taken with the probability of only one arrival in T . From Eq 3.1 it is evident that the interarrival time distribution has a negative exponential distribution with the parameter λ . Thus the generation of Poisson arrival is obtained from the exponential Cumulative Distribution Function (CDF) expressed as

$$t'(\lambda) = \frac{-1}{\lambda} \log_e(rand) \quad (3.2)$$

where t' is the interarrival time and *rand* is a computer generated uniform deviates [0-1]. For a slotted channel, this means that there could be $\{0, 1, 2, \dots\}$ number of random arrivals in one slot or frame taken at each starting point.

The second model for the arrival process is the *Markov Modulated Poisson Process* (MMPP) [Heff86]. Like any Markov model, the process is characterised by a series of states in a Markov chain. Each state corresponds to an arrival rate λ_k ,

$k=1,2,..K$ and a sojourn time Sj_k . The main feature of MMPP compared to Poisson is its index of dispersion for counts, IDC (variance to mean ratio of arrivals). Unlike Poisson which has an IDC of unity, MMPP has an IDC greater than one which means its randomness¹ is greater than Poisson or it is simply called *bursty*. The resulting mean arrival rate for MMPP is obtained by weighting the sojourn times in the next equation.

$$E[\lambda] = \frac{\sum_k \lambda_k Sj_k}{\sum_k Sj_k} \quad (3.3)$$

For a two-state MMPP, the IDC is expressed as [Raaf94] [Heff86]

$$IDC(t) = 1 + \frac{2(\lambda_1 - \lambda_2)^2 r_1 r_2}{(r_1 + r_2)^2 (\lambda_1 r_2 + \lambda_2 r_1)} \left(1 - \frac{1 - \exp(-(r_1 + r_2)t)}{(r_1 + r_2)t} \right) \quad (3.4)$$

where r_k is the reciprocal of Sj_k . For most applications, the use of two-state MMPP is sufficient to provide bursty traffic arrivals. The bursty nature of the traffic depends on the traffic source generator of the user.

Both the Poisson and MMPP arrival models are concerned with the arrival process of a single traffic generator, whether it consists of a single user or many users. Normally in multiaccess protocol performance evaluation, the traffic statistics are taken from a finite number of sources on a shared channel. Each source terminal could be a user, or a group of users. In this case the traffic arrival becomes a *Quasi-random Arrival Process* (QAP). The QAP is very simple and was adopted in much of the literature for protocol performance. It is generated from a number of homogeneous users which alternates between idle and busy periods. The sojourn time a user spends in each state in the QAP are simply negative exponentially distributed. For a binary user state model (see Figure 2) and independent uniform-user assumption, the number of users becoming active in T is binomially distributed expressed as

$$P_{BS}(x, T) = \text{bin}(M_i, x, \sigma_T) \quad (3.5)$$

¹ This term is confusing. However in Teletraffic terms, randomness is relative. It is usually compared with Poisson by using the variance to mean ratio. Peakier or more random traffic has IDC greater than 1.

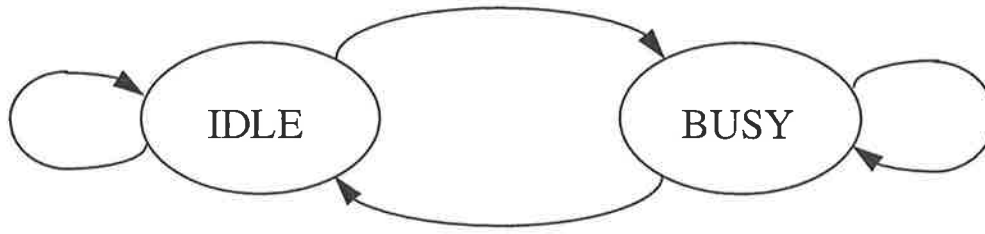


Figure 3.2 Binary State Model for User Activity

where

$$\sigma_T = 1 - e^{-T/T_{idle}}. \quad (3.6)$$

T_{idle} is the mean sojourn time in the idle state, M_i is the number of idle users and $\beta_{in}(s, j, p)$ is the binomial probability for s trials, j successes with probability of success p or

$$\beta_{in}(s, j, p) = \binom{s}{j} p^j (1-p)^{s-j}. \quad (3.7)$$

§3.2 Message Size Distribution

For the message length distribution, the most commonly used model is the negative *exponential* distribution. In this model, the random length of a synthetic message can be generated based on

$$l(L) = -\lceil L \log_e(rand) \rceil \quad (3.8)$$

where L is the mean message length and $\lceil x \rceil$ is the nearest integer from x towards infinity. In a slotted channel where the lengths have integer values, this is sometimes called a Quantised Exponential Distribution (QED). A similar distribution also used in many protocol modelling in the *geometric* model. For the geometric distribution, the random message length is generated from its CDF as

$$l(L) = \left\lceil \frac{\log(rand)}{\log\left(1 - \frac{1}{L}\right)} \right\rceil. \quad (3.9)$$

There is no significant difference in the message length statistics of an exponential and a geometric model as they both inherit a memoryless property. The simplicity of these two models is the fact that only the mean message size (length) is required. Other models which include the skewness of the distribution are also available such as the Pareto, Weibull, Cauchy, and many more. However, for the purpose of comparison, the exponential and geometric models are sufficient.

§3.3 Smoothing Effect of Buffered Data Users

In most systems, mobile terminals require some buffering capabilities. In data applications, it is almost essential because a terminal does not request a channel once the packets are formed as organised by the upper layers of the protocol stack. The third source model which is the Partially Queued Users (with Buffer Threshold), is a method to control the effective message generation rate of the arrival process as the buffer eventually decreases the effective arrival rate of the source by accumulating packets until a threshold is satisfied. We will examine the resulting traffic characteristics at the output of the buffer by calculating the effective message length distribution. An increase in the average message length has a dramatic improvement in the performance of reservation protocols because the load of the access mechanism can be reduced.

To be more general in our analysis, let us assume that the arrival process from an unbuffered source is Poisson with a negative exponential message size distribution. The arriving messages are accumulated in a buffer until 1) the waiting time of the first message exceeds the waiting time threshold or time-out or 2) if the length of the accumulated message(s) exceeds the buffer threshold before the time-out expires. In this problem, the main interest is to determine the access time or the buffering delay. The process is represented in Figure 3.3. From the figure, the access time is measured from the arrival of the first packet until the terminal request for a channel. Any arrival after the allocation of the requested channel is considered part of the next request. Thus every channel request incorporates one or more packets in the buffer.

When there are two or more packets in the buffer during which the messages are on the process of transmission, then we have to determine the effective message length distribution when the messages are concatenated. Let $x_1, x_2, \dots, x_n$, be n independent continuous random variables (r.v.), then the distribution of their sum is the convolution of their original distribution expressed as

in which $x(t)$ is a quantised exponential PDF of the packet length. Similarly, the average request delay of the terminal in transmitting the accumulated messages is

$$T_{sw} = \sum_{t=1}^{tv} t f_R(t) \quad (3.14)$$

where $f_R(t) = [F_R(t) - F_R(t-1)]$ and tv is buffer waiting time threshold. The effective interarrival distribution is then obtained by convolving the channel request time distribution with the interarrival time distribution. For the message length distribution, it is calculated by

$$f_L(l) = \sum_k \sum_{tv} [D_L(k, l) P_{SS}(k, tv)] f_R(tv) \quad (3.15)$$

The mean effective message length will then be calculated as $\frac{\lambda}{\mu} \lambda_{eff}$ where λ_{eff} is calculated from the effective interarrival time distribution. A plot showing the effect of the buffering and time-out is shown in Figure 3.4. It describes the effective interarrival time and the effective message length distribution as different from the original exponential distributions and the Gaussian-like distributions resulted from the smoothing operation in the buffer. The shift of the distributions to the right indicates an increase of the mean values for both distributions which is important in the design of access protocols. It can be confirmed later that this behaviour occurs because the access protocol has its inherent access delays. The plots also includes simulation values. Other simulation results are shown in the Appendix A.

§3.4 Speech Packet Generation

Telephony has been the prime service in a cellular network and will continue to be a dominant traffic. As wireless personal communications advances, various coding techniques have been developed mainly for digital cellular applications [Rab94]. The aim of the coding is simply to reduce the amount of traffic to be transmitted while achieving a high voice quality. It is important in order to determine the suitable channel speed (i.e. slot size and frame duration). Independent to the coding scheme is the detection of the voice signal in order to quantify the amount of traffic (speech activity) necessary for the allocation of a voice channel to a particular user. This requires a

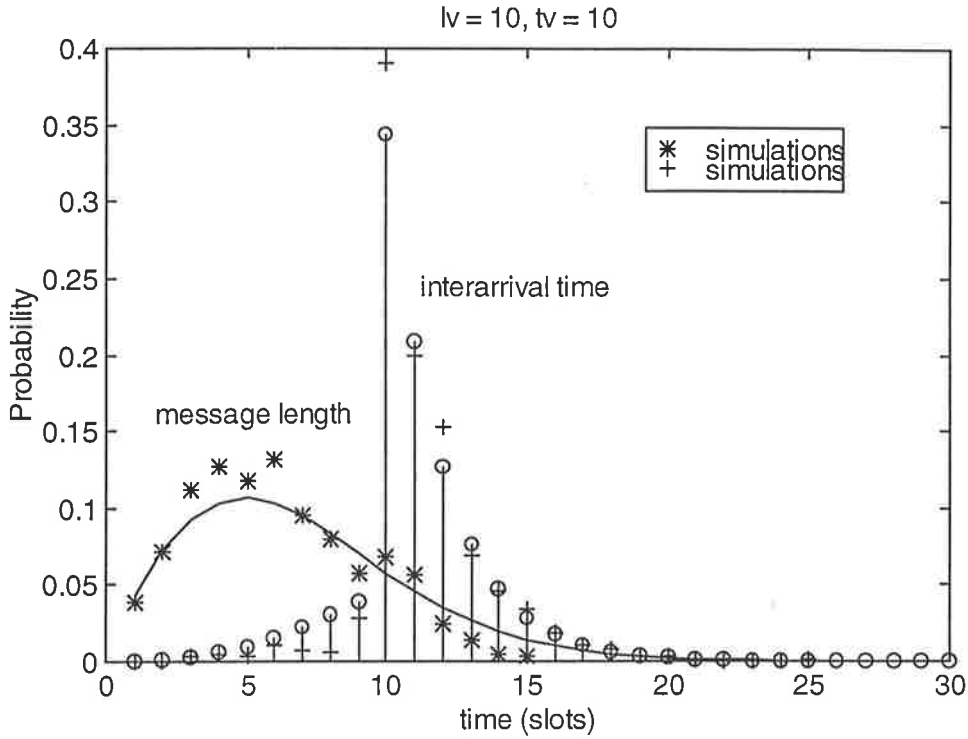


Figure 3.4 Smoothing Effect of a Source Model with Buffer

The stem plot with o - ticks represents the effective interarrival time distribution and the solid line for the effective message length distribution. Since the effective message length distribution is a result of an n -fold self convolution of n messages, the resulting distribution is Gaussian-like. $\lambda=0.5$, $L=1$.

speech activity detector (SAD) prior to the coder to detect the presence of speech information. As a result of using a SAD, an on-off speech statistics which describe the user activity can be extracted from which the traffic pattern is based. There are two kinds of SAD, a Fast SAD and a Slow SAD. The SAD will classify the series of idle and busy periods of the speech source as a result of the conversation. Thus an alternate of talkspurts and silences will be detected by the SAD for which we are interested of the parameters. The slow SAD will only detect the principal talkspurts and silences while the fast SAD can detect the mini-spurts and short pauses and intersyllabic gaps inherent in a speech conversation. The use of slow SAD will result in a relatively longer mean talkspurt and silences not detecting the short intersyllabic gaps and mini-spurts. A speech codec using slow SAD can be modelled by a two-state Markov chain while using the fast SAD requires more than two states.

Studies on the speech statistics of a telephone conversation which employ speech activity detectors had been carried out in [Brad69], [Brad69b], [Brad67], [Grub82] and [LeeUn86]. In the study of P.T. Brady [Bra69], there are two methods that were

employed in speech coders to control the temporal speech parameters. One is the *hangover operation* and the other is the *fill-in operation*. In the hangover operation, a number of frames equivalent to the hangover period are appended at the end of every talkspurt, eliminating all short silence periods less than the hangover period while shortening the rest of the silences greater than the hangover. In this operation, the short intersyllabic gaps that occur in speech are also minimised. The hangover frames are also required in speech transmission to carry the background noise or comfort noise during which a talker pauses. This is essential in order for the conversant not to confuse the pauses and silences for a dropped call. The effect of the increase of the hangover value is similar to that of the reduction in the SAD sensitivity. By decreasing the sensitivity, the switching time of the SAD will be delayed and thus automatically concatenate talkspurts that are separated by small gaps.

The fill-in operation requires an observation period equivalent to the fill-in time. During this period, every silence period or gaps less than or equal to the fill-in time are filled-in leaving only the silences greater than the fill-in time. The fill-in operation inserts silence descriptor frames into the gaps. At the same time, the talkspurts separated by silences less than the fill-in time are concatenated. Fill-in is effective in terms of the resulting traffic characteristics because it does not shorten the long silences in which will not significantly result in a higher voice activity factor. However, fill-in operation requires a buffering period equal to the fill-in time in order to detect the length of the gaps. Since it is impractical in real-time transmission it is favourable to employ a hangover operation for a statistically multiplexed speech in a wireless access protocol. But first, the statistics are to be determined.

To obtain the mean talkspurt and silence duration their probability distribution functions (PDF) are required. From [Brad69] and [Grub82] the talkspurt and silence PDF's are expressed in Eq 3.16a and Eq 3.16b respectively having a two-weighted geometric distribution

$$f_T(k) = Wt_1(1 - u_1)u_1^{k-1} + Wt_2(1 - u_2)u_2^{k-1} \quad (3.16a)$$

$$f_S(k) = Ws_1(1 - w_1)w_1^{k-1} + Ws_2(1 - w_2)w_2^{k-1} \quad (3.16b)$$

for $k \geq 1$ which represents the length in frames where 1 frame corresponds to 5, 10 or 20 ms as used by the speech coder. The values of the constants u_1 , u_2 , w_1 , w_2 and Ws_1 , Ws_2 , Wt_1 , Wt_2 are subject to the language and the talker. Introducing a hangover period h to this fitted model, it is evident that the resulting distribution of the silence PDF is given in Eq 3.17 as [Grub82]

$$f_S^h(k) = \frac{f_s(k+h)}{\sum_j f(j+h)} \quad h = 0, 1, 2, \dots \quad k = 1, 2, 3, \dots \quad (3.17)$$

and the mean silence duration is calculated from the sum of the weighted probabilities which results to

$$T_{SIL}(h) = \sum_k k f_S^h(k) \quad (3.18)$$

To calculate the mean talkspurt duration, it is necessary to calculate the talkspurt rate as described in [Grub82]. The equations for talkspurt rate, R_{sp} and mean talkspurt duration, T_{TALK} are shown in the following equations.

$$R_{sp}(h) = \frac{1}{T_N} \left[1 + (N_t - 1) \sum_k f_S^h(k) \right] \quad (3.19)$$

$$T_{TALK}(h) = \frac{1}{R_{sp}(h)} - T_{SIL}(h) \quad (3.20)$$

where N_t is the number of talkspurts generated in the observation period. The voice activity factor can then be determined as the ratio $T_{TALK} / (T_{TALK} + T_{SIL})$.

§3.4.1 Model for Fast SAD with Hangover ^[Buot95b]

When the temporal speech parameters are obtained, then we can generate synthetic speech packets based on the Markov transition probabilities. We will consider the model in Figure 3.5 where two silences and one talkspurt are used. In the model, one type of talkspurt statistics is used because of a good fit of the talkspurt distribution from a two-weighted geometric to an exponential distribution. So we first express the mean silence duration as

$$T_{SIL} = \frac{R_{gap}T_{gap} + R_{listen}T_{listen}}{R_{gap} + R_{listen}} \quad (3.21)$$

The rate of listening silences as well as its average duration are not sensitive to the hangover operation if the hangover period is short. However, the gaps are affected so that the values of T_{gap} and R_{gap} are to be evaluated. Since the sum of R_{gap} and R_{listen} is equal to the talkspurt rate, we can calculate T_{gap} as

$$T_{gap} = \frac{T_{SIL} R_{sp} - R_{listen} T_{listen}}{R_{sp} - R_{listen}} \quad (3.22)$$

where R_{gap} and R_{listen} are the rate of the gaps and the long silences respectively whose sum is equal to the talkspurt rate. T_{listen} is the mean listening silence duration and both T_{listen} and R_{listen} are taken from the mean silence duration and talkspurt rate with very large hangover and T_{gap} is the mean duration of the silence gaps. The sojourn time in each state (except for the hangover) in Figure 3.5 is exponentially distributed.

The model above is realistic as it depicts exactly how speech is generated in a speech coder. Since every talkspurt is appended with a hangover, then the resulting talkspurt distribution can be approximated as a constant plus exponential. This is also observed in [LeeUn86] where the talkspurt and the silence distributions are constant plus exponential and exponentially distributed respectively. For simplicity, the talkspurt distribution is assumed to be exponential since the length for hangover is usually much smaller than that of the average talkspurt length (one order of magnitude). The other feature of the model is that we can vary the traffic statistics by just changing the hangover value for the purpose of optimisation in contrast to other models used in [Nand92] [Cleary94] [Goodm89] which used fixed parameters for their analysis. Some results in this model are shown in Figure 3.6. In some studies like [Stern94] [Brad69b] the interaction between speaking parties is taken into account which of course demonstrate some correlation between the successive transitions. In the study of access protocols where multiple users are often assumed, the interaction is not necessary but rather concentrate on the talkspurt and silence statistics to obtain the effective access rate and traffic load to the system. Also, [Kim83] confirmed that the arrival process for statistically multiplexed speech packets approaches Poisson if multiple users are considered.

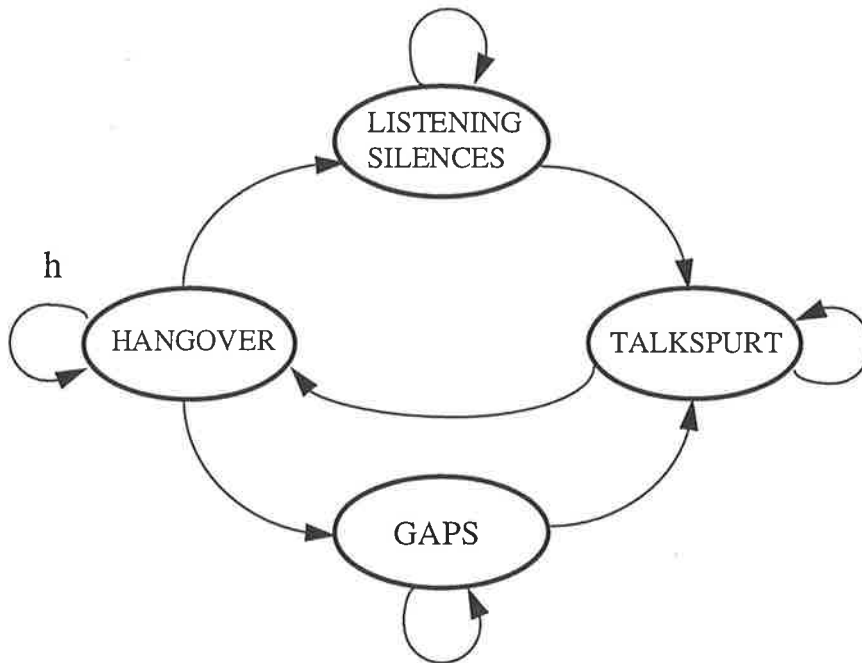


Figure 3.5 Speech Model with Short Hangover Period

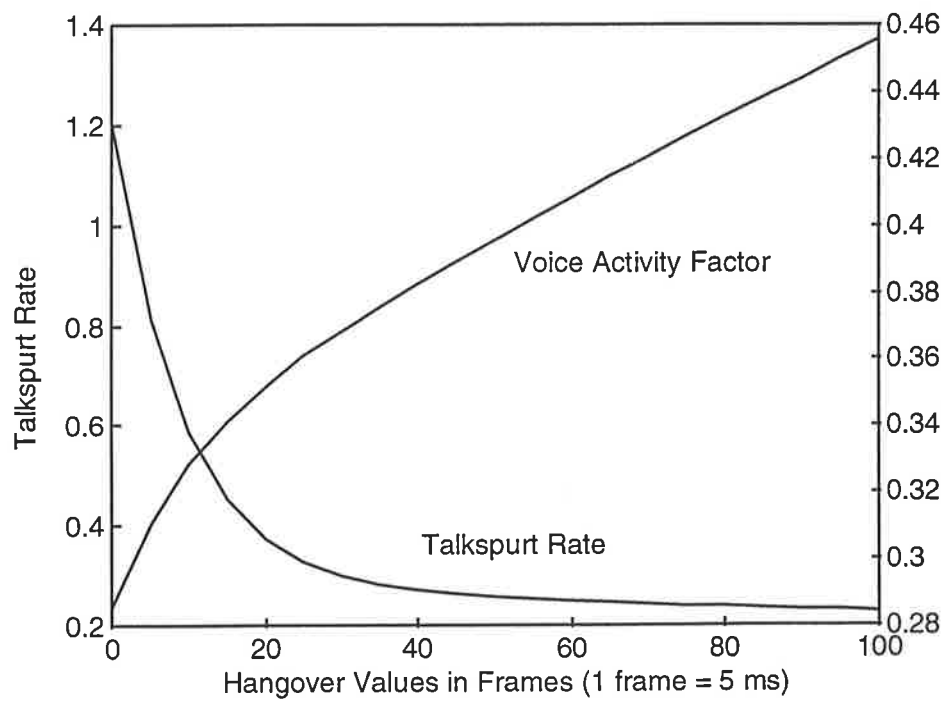


Figure 3.6 Speech Packet Statistics for Different Hangover Values
The temporal speech parameters are taken from [LeeUn86]

§3.4.2 Effect of Hangover to the Speech Quality

Speech quality is determined by a subjective parameter called *Mean Opinion Score* (MOS). This MOS parameter is subject to many parameters like signal to noise ratio, codec speed, coding rate, etc. However, in the access protocol, these parameters are disregarded and concentrate on the packet dropping probability which can be quantised. But when it comes to the effect of hangover values, two things can be considered. One is the effect on speech quality because shorter speech hangovers will reduce the spontaneity of the syllables. No study has yet considered how the fast and slow SAD vary the MOS. The other issue is in the calculation of the percentage of packet dropped in the statistical multiplexing of packetised speech. If the hangover value is not taken into account, the packet dropping probability is slightly misleading because longer mean talkspurt length as a result of large hangover values carries both the speech packets and the silent descriptor frames. Therefore, a correction must be made in the calculation of the packets dropping rate as most of the dropped packets are actually speech packets at the beginning of the talkspurts (see Figure 3.7). If we have h as the hangover, then we need to calculate the mean value of the gaps when talkspurts are concatenated by the hangover period. Let $\tilde{h}$ be the mean value of the gaps, then we have

$$\tilde{h} = \frac{\int_0^h t \lambda_0 e^{-\lambda_0 t} dt}{\int_0^h \lambda_0 e^{-\lambda_0 t} dt} \quad (3.23)$$

where λ_0 is the talkspurt rate with zero hangover. The numerator can be solved using the integration by-parts resulting to:

$$\tilde{h} = \frac{\frac{1}{\lambda_0} - e^{-\lambda_0 h} \left(h + \frac{1}{\lambda_0} \right)}{1 - e^{-\lambda_0 h}} \quad (3.24)$$

The previous equation suggests that the value of $\tilde{h}$ is approximately $h/2$ if the hangover period is very short. To calculate the average number of talkspurts without hangover being concatenated, we start with the average talkspurt length as

$$T_t(h) = (T_t(0) + \tilde{h})(n_T(h) - 1) + T_t(0) + h \quad (3.25)$$



Figure 3.7 Example of Concatenated Talkspurts due to Large Hangover

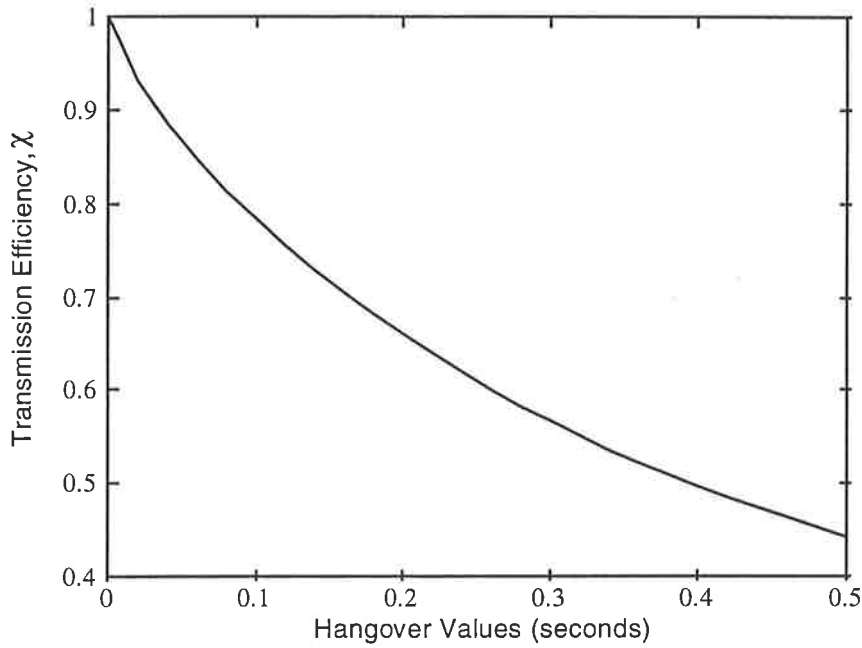


Figure 3.8 Effect of Hangover Value to Speech Quality

Rearranging we result to

$$n_T(h) = 1 + \frac{T_t(h) - T_t(0) - h}{T_t(0) + \tilde{h}} \quad (3.26)$$

where n_T is the number of talkspurts without hangover that are concatenated. The transmission efficiency, $\chi(h)$ which is the ratio of a number of speech packets to the total packets (including silence descriptor frames) in a talkspurt is expressed as

$$\chi(h) = \frac{n_T(h) T_t(0)}{(T_t(0) + \tilde{h})(n_T(h) - 1) + T_t(0) + h} \quad (3.27)$$

The transmission efficiency is then plotted in the figure above. It shows that the amount of speech information in a speech with large hangover is only a fraction of the talkspurt size. It is therefore necessary to limit the hangover length but sufficient to carry the background noise information.

§3.5 Video Traffic Models

Video traffic characteristics vary due to the many types of coding schemes. This is a results from the different quality requirements of different applications as well as the use of more advanced coding and compression techniques that are currently being developed (see [Streit95]). These requirements range from a low bit bit rate image to a variable bit rate moving picture. As in the generation of speech packets we are not concerned with the details of the coding scheme but focus on the bit rate variation in representing the video traffic. As a result we will deal with two main video encoding schemes that are widely adapted. They are the Constant Bit Rate (CBR) and the Variable Bit Rate (VBR) video.

In the CBR encoding, a buffer is placed after the coder to smoothen the varying bit rate and the buffer size and is used as a feedback to the coder in order to adaptively vary the quantisation size of the video information [Dalg95]. On the other hand, the VBR encoding uses a fixed quantisation level to maintain the picture quality while producing a varying information rate. The common video standard is the ITU H.261 CBR encoding designed for datarates which are multiples of 64 Kbps. A sample mapping of this information rate to the channel will result in

$$6.4 \frac{Kb}{frame} \times 10 \frac{frames}{second} \times 10 \frac{ms}{TDMA frame} = 640 \frac{bits}{TDMA frame}$$

If a slot in a TDMA frame can carry 128 bits, then it is a valid assumption that video traffic is 5 traffic units (TU) relative to speech traffic. The minimum bit rate of a CBR video is 64 Kbps as being adapted in one ISDN bearer channel operating in a circuit-switched mode. CBR encoding generally suffer from changes in quality but require constant bit rate channels and thus easier to transmit in most access protocols.

For the VBR video, the bit rate varies from by a factor V_f from the basic bit rate. A most prominent characteristic of VBR encoding is that instantaneous changes in the picture scene can produce large variation of the bit rate. It is indeed difficult to transmit this kind of traffic unless a large bandwidth channel is available (i.e. VBR sources thus require VBR transmission). In order to maintain the quality set by the encoder, the access protocol must be able to support the maximum bit rate of the encoder (i.e. during changes in picture scenes).

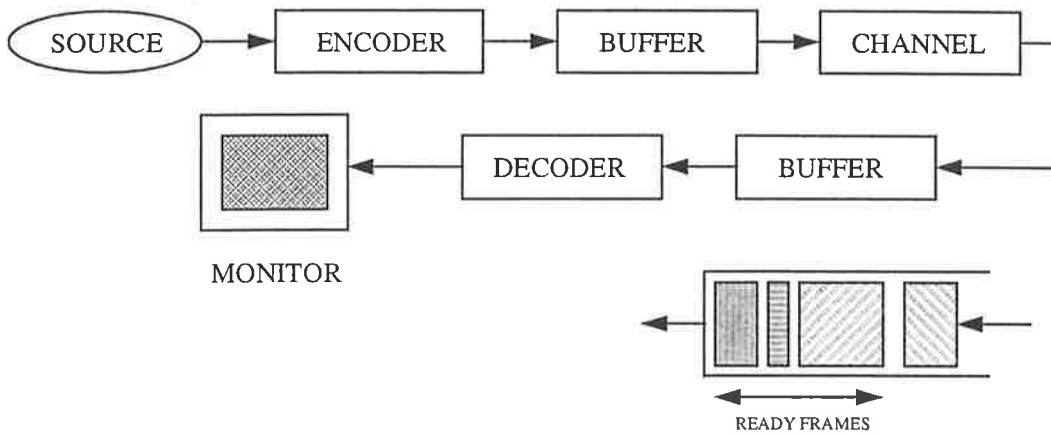


Figure 3.9 System Model for VBR with Limited Speed Channel

In cases where the maximum bit rate of the channel is less than the maximum encoder rate, buffering is required in both the encoder and the decoder (receiving end) to mitigate the problem of shortage in the channel. This problem also occurs in an integrated system where video traffic contend with other traffic like voice and high priority data. A system model for VBR video is shown in Figure 3.9. The figure depicts the delay/quality trade-off. Selection of an acceptable buffering delay also depends on the variance and correlation of the frame size. In this case, highly correlated frame size sequences during channel high loads is detrimental to the quality of the video transmission.

Numerous attempt to replicate and model VBR video were found in the literature. However, most of the traffic characterisation as limited to some video sample (e.g. Miss America[†] video sequence) and therefore not a representative a realistic sequence which is a mixture of different picture scenes and sequences. Since the picture scene can vary indefinitely, the Markovian model requires an *infinite number of states* which will be discussed in §3.5.1.

§3.5.1 Infinite State Markovian Video Source Model

No exact model can replicate VBR video sources but there are some important considerations in modelling such as 1) basic or minimum rate 2) rate variation and 3) sequence correlation. In R-TDMA, the rates are quantified in terms of the number of TDMA bursts generated in every TDMA frame based on the picture sampling rate and

[†] Miss America is a video scene that is relatively stationary. This has been the basis for video quality reference in many studies in low speed VBR coding.

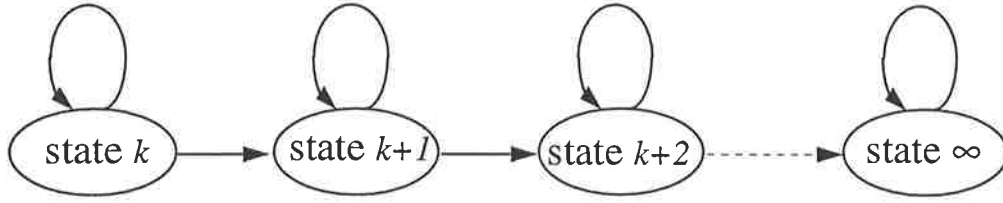


Figure 3.10 Infinite State Markovian VBR Video Model

encoding (compression). Thus the frame size sequences consist of deterministic and random components. For simplicity we can model a VBR video by a multistate Markov source. Each source in the model is characterised by a corresponding frame size distribution and sojourn time. To maintain the Markovian property, the sojourn time is exponential while the frame size could be a chosen distribution. The correlation between the frame sequences is determined from the sojourn time of each state. Multiple states are required to obtain good models so that an *Infinite State Markovian* model is more appropriate (see Figure 3.10). The mean feature of a Markovian model is that the next state of the system depends on the previous state (i.e. $X_n = f(X_{n-1})$). The correlation property of video traffic can be represented by a small variation of the traffic characteristics between two succeeding states. An example of an *Infinite State Markovian* video source is as follows. The video is represented by a *scene* and the *motion* within a scene. At the s^{th} scene, the video rate is represented by a fixed rate $Vo(s)$ and a Gaussian random component $E(s)$ of known variance which represents the motion within the scene. The duration of each scene is exponentially distributed to retain the Markovian property while the scene parameters must be of known distributions. When the sojourn time of a scene expires, the video moves to another scene and updates the scene parameters. To maintain the correlation property, $Vo(s+1) = f(Vo(s))$. The instantaneous frame size of the video frame at the s^{th} scene is described as

$$I_s(n) = X_s(n) + E_s(n) \quad (3.28)$$

where the mean component is a series of Markovian transitions given as

$$X_s = \begin{cases} X_{(s-1)} + f(d) & \text{if } y \leq e^{\left(\frac{-|X_{(s-1)} - X_0|}{X_m} \right)} \\ X_{(s-1)} - f(d) & \text{otherwise} \end{cases} \quad (3.29)$$

and $f(d)$ is a uniformly distributed r.v. which accounts for the deviation of the mean size. The expression $y \leq e^{\left(\frac{-|X_{(s-1)} - X_0|}{X_m} \right)}$ is a test for the state of the system for controlling the transition. In the expression, y is the uniform deviates and, X_0 is the minimum mean rate and X_m is the mean for an exponential control of the state transition. $Es(n)$ is a Gaussian distributed random variable whose standard deviation is also Gaussian and is updated in every scene. The resulting requeencies in terms of the number of slots required for R-TDMA is shown in Figure 3.12 and is compared with the widely used model described in §3.5.2.

§3.5.2 *AutoRegressive Video Source Model* ^[Nom87]

In the AutoRegressive Model (AR) (see Figure 3.11), the scenes are not treated individually but the correlation between the frame sequences is emphasised. The AR model is described in the following equations [Wat97][Nom89]

$$I(n) = X(n) + I_0 \quad (3.30)$$

$$X(n) = \sum_{w=1}^W K(w)X(n-w) + E(n) \quad (3.31)$$

where $I(n)$ is the frame size in the n^{th} frame, I_0 is the average frame size, $A(w)$ is the AR coefficient and $E(n)$ is a Gaussian distributed random component. Thus the model accounts for the fixed, random and correlation components.

§3.5.3 *VBR Source with Channel Load Feedback*

The main objective in video transmission is to achieve a constant quality picture so that the VBR encoding was introduced. However, in the multiaccess environment where the channel availability is not predictable, no encoding scheme can guarantee a constant quality video. Moreover, the VBR encoding scheme by itself cannot provide constant quality video due to some physical limitations (e.g. minimum quantiser scale, sampling rate, etc.). Here we introduce a coding scheme which incorporates the instantaneous channel availability to the coding rate by adjusting the quantisation level to the best achievable video quality.

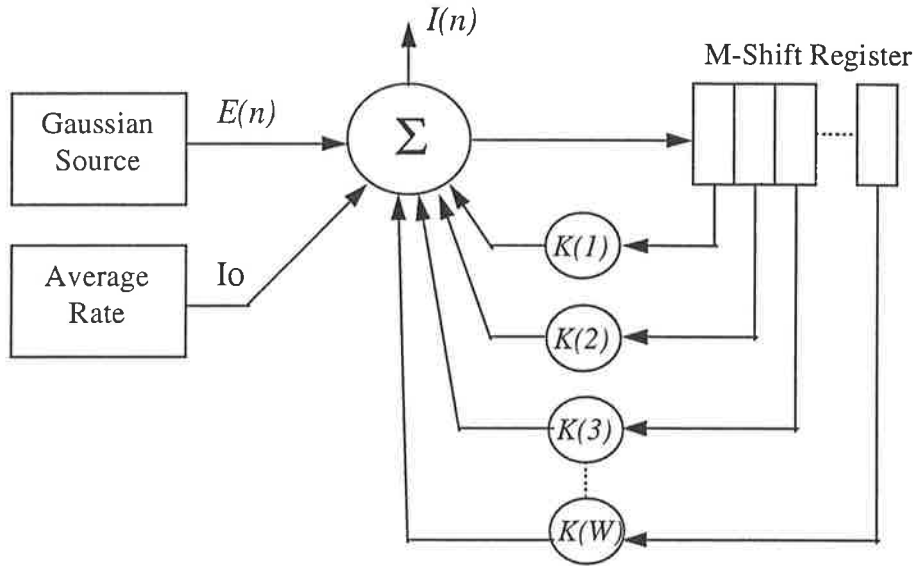


Figure 3.11 Auto-Regressive Video Source Model

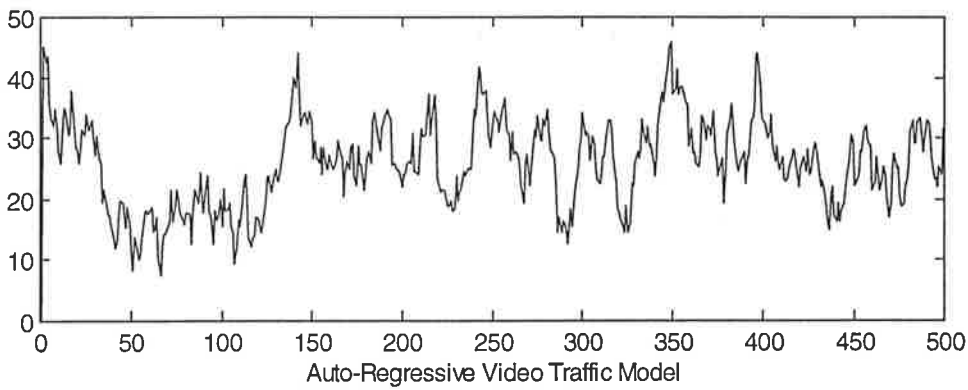
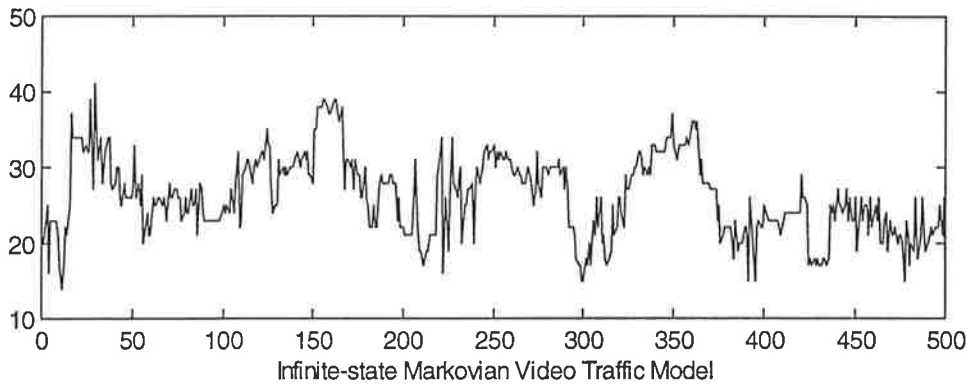


Figure 3.12 Sample Histogram of Generated Video Traffic

(y-axis refers to the number of channels required)

Infinite State Markov : E_s = Gaussian with zero mean and standard deviation = x , x = Gaussian with zero mean and standard deviation = 2, $X_m = 20$, $X_o = 12$, $f(d)$ = uniform from 0 to 3, $X_{smin} = 12$ and $X_{smax} = 50$. $I_s(n)_{min} = 5$.

AutoRegressive : $I_o = 12.8$, $W = 1$, $K = 0.88$, E = Gaussian with zero mean and standard deviation = 15.36.

Video quality is subjective so that mathematical quantification often leads to inaccurate measure. However, in the context of the effect of transmission the coded video information can be quantified in terms of relative quality measure from the coder output. In this study, the relative video quality due to the encoding scheme, QoS_v is simply expressed as

$$QoS_v = \left[\frac{Rq - Rq_{th}}{(1 - Rq_{th})} \right]^{kv} \quad \text{where } Rq \leq 1 \quad (3.32)$$

where kv is the quality degradation factor and Rq_{th} is the relative quantisation threshold and Rq is the relative quantisation which is the reciprocal of the ratio of the actual quantisation to the target or optimum quality size (see Figure 3.13). Rq is set to a maximum value of unity since the maximum quality is equal to the value achieved by the target quantisation size. Hence it is only a waste of resource if the actual quantisation ratio is greater than unity (more bits to send). For the quantisation size the value depends on the actual encoder being used but it is directly related to the number of bits per video frame and hence to the number of channels required or reserved. So the quantisation ratio can be replaced by the ratio of the actual frame size to the target frame size and the quantisation threshold is replaced by the minimum frame size. It is evident that for a CBR coding scheme, the frame loss rate is zero so that the video quality entirely depends on the variation of the quantisation size. Hence, varying video quality is expected. In contrast, VBR coding quality depends on the channel availability only because the video terminal requires variable number of channels (speed).

Since both schemes do suffer from quality degradation, a coding scheme which negotiates for the best quality requires further flexibility to adjust the actual quantisation to the channel availability and/or channel quality. Therefore, an adaptive encoder requires a feedback from the resource allocator regarding the channel availability in order to optimise the video quality by adjusting the quantisation size. Since both the relative quantisation and the frame loss rate contribute to the video quality, a balance of both parameters will achieve the optimal video quality. Detailed investigation of this problem is deferred to Chapter 6.

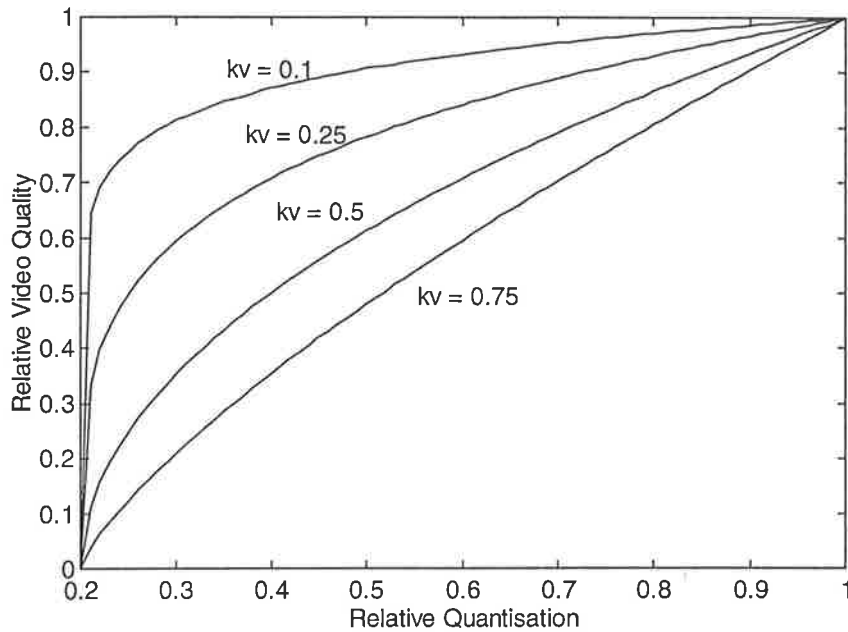


Figure 3.13. Plot of the Relative Video Quality with no Frame Loss.

The figure shows the rapid reduction in the video quality near the threshold value. ($R_{qth}=0.2$)

§3.6 Summary

Throughout this chapter, the source traffic generation is carefully studied. We have seen that the buffering of messages either in the source or in the terminal is an important factor that has to be considered in modelling the traffic sources. To account this factor, the source traffic behaviour is simplified in Figure 3.1. The essential models for traffic generation are the Poisson and MMPP arrival processes and the negative exponential and geometric message size distributions. Combination of these models together with the source coding and buffering characteristics can replicate some complex traffic sources. A good example is the source modelling of voice packets with hangover in §3.4.1. As buffering will affect the resulting traffic statistics, a simple method to evaluate the output distribution is shown using discrete analysis (ie. TDMA channel). This is necessary in order to compare the actual distribution as a result of buffering at the terminal. This is applicable in the traffic generators with queued users. For the single message arrival model, simplistic user activity model will result. The most commonly used user activity model is that of the Binary State Model in §3.1.

By noting carefully the characteristics of the speech packet from the fast SAD with hangover model, the talkspurt rate as well as the voice activity factor varies significantly with the hangover values. The increasing value of the voice activity factor means, longer hangover is not favourable. In contrast, the rapid reduction of the talkspurt rate in the region of hangover values less than 200ms, favours for some hangover. Eventually, the use of hangover must take into account that the resulting talkspurts do carry some silent descriptor frames and therefore the intelligibility of the speech (conversation) must take into account the ratio of actual speech packets to the silent descriptor frames. The calculation of the intelligibility factor was provided in §3.4.2.

Lastly, generic video source models were considered. In particular, the VBR video source is described in more detail. Two source models were proposed which are the Infinite State Markovian model and the VBR Source with Channel Load Feedback model. More consideration is required for the second model because it negotiate the QoS of the video source with the multiaccess layer. This is discussed in Chapter 6 in which for a given QoS criteria, the best video QoS can be achieved.

Chapter 4

Performance Analysis of R-TDMA

In the performance analysis of R-TDMA protocols, there are three important things to consider. They are the *input traffic* which is described in Chapter 3, the *processes* within the protocol, and the *system model* relating to the behaviour and structure of the entire system. This chapter deals with the performance analysis of R-TDMA protocols. In particular, approximate methods are considered in evaluating the throughput/delay performance. Throughout this study, the analyses were only concerned with the uplink due to the nature of the multiple access.

There are two notions of capacity in a multiaccess environment. One is the maximum capacity or throughput which is based on the assumption that delay is unlimited. The other is the delay limited capacity which considers the channel quality and service quality to the capacity of the channel. Thus the objective in the analysis of multiaccess protocols is to obtain the *throughput against the moments of the delays*. For most applications, the performance criteria based on a two-moment solution (mean and variance) is sufficient to describe the throughput-delay characteristics in contrast to the mean value analysis often encountered in the literature. This is due to the QoS criteria of most data services which is expressed usually in terms of *delay percentile*. Another key performance parameter of multiaccess protocol is the *stability*. Some protocols (especially the Random Multiple Access protocols) have inherent non-linear throughput/delay characteristics. Hence, the protocol stability as a function of the channel load must be determined. In this chapter, we describe the basic procedures in evaluating the performance of Reservation-TDMA protocols. We use ATDMA as an example because its generic frame structure is similar to most reservation protocols.

§4.1 System Model

Every protocol consists of distinct processes. For R-TDMA, the protocol consists of an access mechanism (i.e. random access, fixed assignment, access scheduling, etc.) for the users to request a channel or to inform the base station of its status. It also maintains a queue to manage the available resources among the competing users. Other

protocols require a polling process as part of the central resource allocator when preemption or buffering is supported in the case of multimedia access support. In addition, some protocols require a multislot request queue when multi-channel reservation is supported for variable bit rate users. Consequently, a system model and its assumptions are required in order to analyse such complex protocol processes. The next subsections describe the relevant assumptions used in the performance evaluation of R-TDMA.

§4.1.1 Channel

In the study of multiaccess protocols with voice traffic, noiseless channels are always assumed. However, for data traffic the effect of noise is very important. Simple models for noisy channels are sufficient to incorporate the effect of noise into the multiaccess performance. This is because errors are often measured in terms of frame of burst errors rather than the instantaneous signal to noise ratio being used in the study of the physical channel itself. The use of noisy channels in the performance evaluation is deferred in Chapter 6. Here, we focus on the frame structure only. Some channel structures of R-TDMA are shown in Figure 4.1. Minislots for access request is employed in the first two frame structures. The main drawbacks of the approach are 1) availability of feedback where the request cannot be acknowledged until the next frame due to the insufficient time interval and 2) synchronisation problems. When the result of the current access attempt will not be available immediately in the next access slot (delayed feedback), the users cannot fully utilise the benefits of multiple access slots. To this effect, a contending user will randomly select one of the minislots for the transmission of the request packet. Secondly, due to the small duration of the access slots, lesser synchronisation bits can be allocated in the access burst and hence limit the efficiency of the access mechanism.

§4.1.2 Slot Reservation

Two schemes are used in this study. They are the single slot reservation and the multislot reservation. In the case of multimedia transmission, multislot is often required. Single slot reservation is used for voice service. The multislot reservation is tightly coupled with the QoS performance. There are two schemes for multislot allocation. One is for the fixed bit rate multislot. In contrast, the second scheme is the

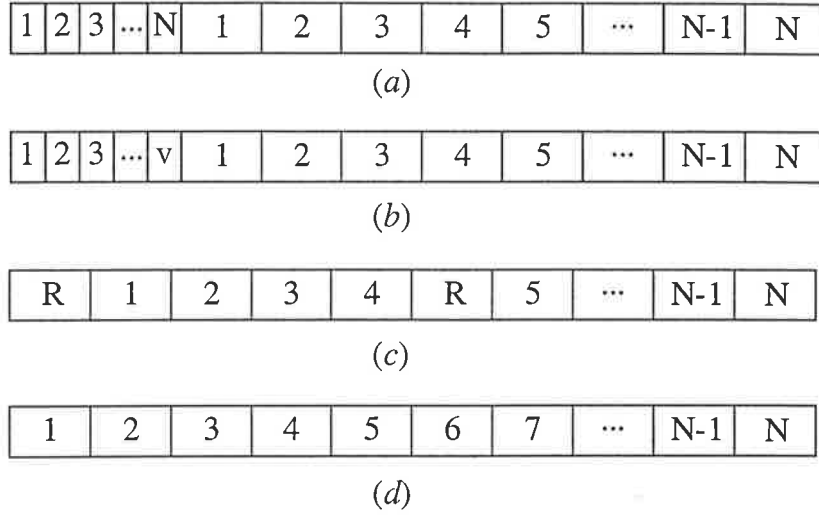


Figure 4.1 Basic Channel Structures used in R-TDMA protocols

- (a) Conventional TDMA for satellite systems
 (b) Dynamic TDMA ($v < N$)
 (c) Advanced TDMA and E-TDMA*
 (d) PRMA and Crowther's Reservation Protocol

variable bit rate multislot and the slot allocation is often based on the best QoS achievable (best effort allocation).

There are also two reservation policies in R-TDMA. They are the *single user per channel* and the *multiple users per channel*. The single user per channel allows only one user to reserve a slot until all its packets are successfully transmitted. This scheme is effective if the average message size is relatively short and the channel load is low. The multiple user per channel employs a form of sub-multiplexing whereby users are allowed to reserve only a specified amount of slots in the current transmission. This allows multiple users to transmit in turn within a single slot. Reservation policies to schedule the users are required in this case. This policy is effective when the channel load is on or near the overload region.

§4.1.3 Immediate First Transmission (IFT)

Upon the arrival of the first packet, a user immediately requests a channel in the case of a random access mechanism. Otherwise, the system is said to be a Delayed First Transmission (DFT) where a user waits for a random time before it first requests a channel. In [Wu94a], the DFT scheme is used in order to simplify the analysis. It is

* E-TDMA requires multiple carriers with staggered request slots.

done by having the first request probability equal to the retransmission probability as a result of collision. However, if the random delay is quite large, it will dramatically increase the mean access delay resulting in a large overhead as a result of the random access mechanism. Assuming that a user has no knowledge of the current status of the system upon its request of a channel, an IFT scheme is always recommended. The difference between the IFT and DFT schemes is the amount of random time a DFT requires during the initial transmission.

§4.1.4 Effect of Random Access Collisions

Again, for protocols using a random access mechanism, collision cannot be avoided. In the event that a collision occurs, three assumptions can be drawn namely: 1) Collision is catastrophic, 2) The base station can capture one user with a certain probability, and 3) The base station can capture more than one user with known probabilities. The third assumption is based on a system with multiple receivers/antennae with some degree of independence. Unless specified, a catastrophic collision was assumed where all colliding users lost their packets. Users that experience collisions will retransmit at a random time interval or in a prearranged sequence.

§4.1.5 Single Carrier System

In the actual implementation, multiple carriers can operate with a centralised controller. It will eventually increase the multiplexing gain of the system. For simplicity and conservatism, a single carrier operation was assumed throughout the study. One feature of reservation based protocols is the ease of hopping from one radio channel to another due to the fast cycle of the reservation process. In this case the load can be distributed evenly among the carriers (radio channels). Moreover, when each carrier consists of multiple timeslots, higher multiplexing gains can already be achieved in a single carrier operation.

§4.1.6 State (In)Dependence

In the system analysis of R-TDMA, each process is identified by a unique state of the user. However, it is inevitable that some states are dependent. The inter-dependence of each state in the access protocol must be carefully examined in order to obtain some simplifications in the modelling and analysis. For states or processes that are

independent to the rest of the system, they can be isolated and examined individually thereby reducing the complexity of the analysis. Furthermore, the processes that are highly dependent have to be combined and modelled as a single complex process. A good example is the dependence of the queue and the reservation states whereby the departure rate of the latter dictates the movement of the queue.

§4.2 Analysis Methods

The system model for R-TDMA as shown in Figure 4.2 indicates the amount of complexity involved in the analysis so that we often have resort to approximated analysis or simulations methods. In some cases, over-simplified models will be used and simple analysis can be derived. In that case, a simulation method is required to validate the accuracy of the analysis. The techniques that are applicable for the analysis of R-TDMA are as follows:

1. **S-G** Analysis
2. Simulation
3. Birth and Death Markov Chains
4. Discrete Markov Analysis
5. Equilibrium Point Analysis or Transient Fluid Approximations

Stability and load sensitivity can be analysed using the S-G analysis (throughput-load). It identifies the favourable operating region of a protocol by scanning the performance at various input load characteristics. The S-G analysis can determine only the steady-state (static) performance of a protocol. If further analysis is required, a more reliable approach to evaluate a system performance is to use a dynamic analysis. For systems that are very complex, the simulation method is the most favourable method to use. The main advantage of simulation is the ease of evaluating most performance parameters by replicating the entire system. The simulation method is the main performance evaluation tool used in this study.

Another method suitable for the analysis of reservation protocol is the *Markov Analysis* method. Two Markovian techniques are wide adopted in the study of protocols. They are the *Birth and Death Markov Chains* and the *Discrete Markov Analysis*. The suitability of the Discrete Markov analysis in R-TDMA is due to the use

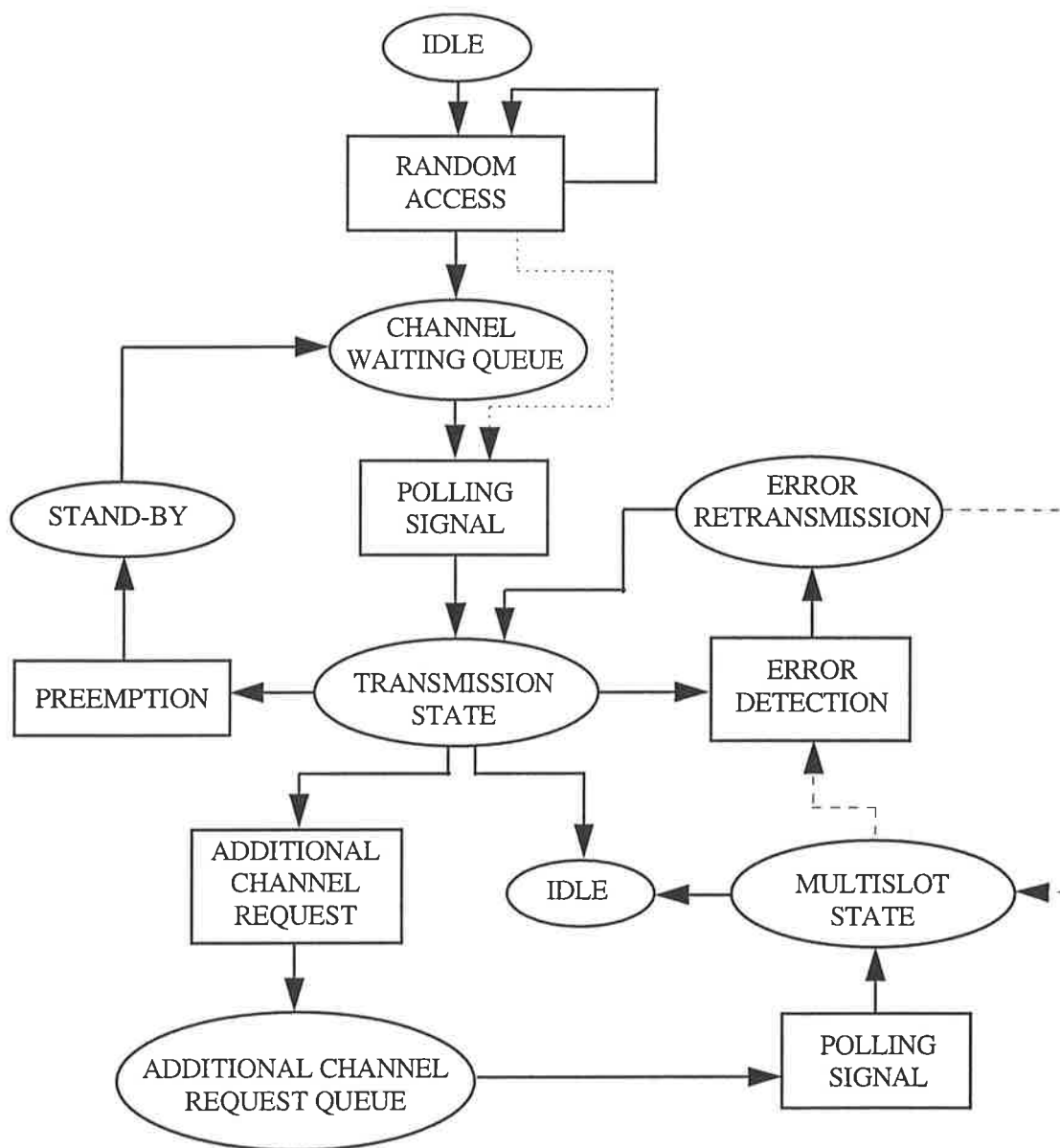


Figure 4.2 Sample System Model of R-TDMA

of slotted channels resulting in a system that can be modelled by discrete state and transitions at fixed time intervals. If the channel is framed, a more convenient analysis is to use a frame by frame state transition so that regardless of the channel structure, a uniform solution can be employed. The advantage of the Markov Analysis is to be able to obtain the user distribution in each state and the use of linear programming techniques to enhance the calculation speed. It is convenient to use the Markov Analysis when the number of states are only few (i.e. less than four). The Markov Analysis is necessary if the transition rate from one state to another is high (i.e. the probability of 2 or more users departing/arriving to each state is high). However, if the probability of more than one user is departing and/or arriving in all states is very small, the transitions of the Markov Analysis can be simplified and the computing time is dramatically reduced. In this case, the reservation state can be analysed by using *Markov Chains*.

Sometimes, for protocol comparison, a first moment solution is only required for the performance criteria. In this case, the *Equilibrium Point Analysis* (EPA) [Fuk83] is applicable. The EPA method is simple but accurate. First, a Markov model of the system must be obtained with the transition probabilities. Then the equilibrium state equations must be determined by equating the inflow equal to the outflow of each state. Then the mean number of users in each state can be solved from the equilibrium state equations. As in the Markov Analysis, a fixed number of users must be assumed (Conservation of Mass).

The EPA method is sometimes cumbersome if the system is non-linear and multi-stable because finding the correct pair of throughput delay is not a trivial task. Mukumuto, devised a numerical approach based on EPA [Muku90]. The method is called *Transient Fluid Approximation* (TFA). In TFA, instead of calculating the mean number of users in each state by simultaneous equations, an iteration method is used. The iteration will start with randomly chosen values of each state and the values are updated based on the transition probabilities. The iteration is done on fixed time intervals until the inflow is equal to the outflow in each state or if the difference between the mean values of the number of users in each state approaches to zero. The TFA method has some limitations since the transition probabilities are sometimes difficult to obtain. Like most other methods, it calculates the number of users in each state in the system. In calculating the throughput/delay values, established results from queuing theory can be used.

§4.3 Approximation of a Slotted Random Access Protocol

Since S-ALOHA is often involved in reservation protocols, we need a better approximation for random access protocols. The accuracy of both EPA and TFA in analysing a R-TDMA protocol is sensitive to the retransmission probability of the S-ALOHA contention process. In most papers of Gang Wu, small values of the retransmission probability were often used. In doing so, the system will maintain a relatively larger number of backlogged users which increases the accuracy of the analysis. In fact in the analysis of PRMA where the retransmission probability, p is high, a Markov Analysis is recommended. Here, we use another approximation based on *successive trials* which then determine the delay distribution of Slotted Random Access protocols.

To approximate the delay distribution of a Slotted Random Access protocol (e.g. S-ALOHA) we can model the user's behaviour as a series of independent trials until its packet is successfully transmitted. Since our channel is slotted, we need to determine the distribution of the number of slots passed by a user from the generation of its packet until the successful transmission of the said packet. The following assumptions are necessary:

1. Immediate First Transmission (IFT) - for random access without channel sensing, every busy user makes their first attempt right after the generation of its packet to avoid unnecessary delay as in the delayed first transmission.
2. Random Retransmission upon collision - if a user experiences a collision, it retransmits after a random interval.
3. Capture Model - describes the receiver's capability to capture packets in the event of a collision
4. Packet Arrival Model - usually a Poisson process.

The probability that a packet is successfully transmitted on the current slot is the product of the probability that the user transmits, the packet is captured and it satisfies the residual bit error requirements based on the channel characteristics. Thus we have

$$\Pr\{\text{success}\} = \Pr\{\text{transmit}\}\Pr\{\text{captured}\}\Pr\{\text{decoded}\}. \quad (4.1)$$

The first term in the equation refers to the access protocol itself and it specifies what collision resolution algorithm is used. The second term is tied up with the receiver's

ability to receive the intended packet and is very much related to the topology of the network [PolSil87]. The last term is related to the channel characteristics and error control coding. In the context of multiaccess protocol, we limit our assumptions to the first two terms only.

The probability that the packet is successfully transmitted right on the first attempt (IFT) is simply the capture probability given as

$$\Pr\{\text{delay} = 1 \text{ slot}\} = \sum_{K=1}^{\infty} \frac{\Pr\{K \text{ users}\} \Pr\{1 \text{ user captured out of } K\}}{K} \quad (4.2)$$

Then the probability that a user is retransmitted is $1 - \Pr\{\text{delay} = 1 \text{ slot}\}$. The probability of successful retransmission on the current slot given that it is not the first attempt is simply

$$\Pr\{\text{success}\} = \Pr\{\text{retransmit}\} \sum_{K=1}^{\infty} \frac{\Pr\{K \text{ users}\} \Pr\{1 \text{ user captured out of } K\}}{K}. \quad (4.3)$$

To calculate the delay distribution, the process is a series of *geometric* trials. We will use this model to the S-ALOHA protocol because it satisfies our assumptions.

If we define Cp^k as the probability of one packet being captured for K users transmitting, the success probability of the S-ALOHA is as follows:

$$\Pr\{\text{success}\} = \Pr\{\text{transmit}\} \left\{ \begin{array}{l} \Pr\{\text{no other user transmits}\} Cp^1 + \\ \Pr\{\text{one other user transmits}\} Cp^2/2 + \\ \Pr\{\text{two other users transmit}\} Cp^3/3 + \\ \text{etc} \end{array} \right\} \quad (4.4)$$

$$\Pr\{\text{success}\} = p \sum_{k=1}^{\infty} \frac{G^{k-1}}{(k-1)!} e^{-G} \frac{Cp^k}{k} \quad (4.5)$$

simplifying, we have

$$\Pr\{\text{success}\} = p e^{-G} \sum_{k=1}^{\infty} \frac{G^{k-1}}{k!} Cp^k. \quad (4.6)$$

For a non-capture case, the probability of success on the first trial ($delay=0$) is e^{-G} .

Thus the probability of success after the r^{th} retry ($delay=r$) is $(1 - e^{-G})(pe^{-G})(1-pe^{-G})^{r-1}$.

§4.3.1 Finite Population ALOHA

For a finite population S-ALOHA, the probability of successful contention from a given terminal is given in [Tanbm89] as the probability of exactly one terminal contending (Eq 4.7). Similarly, for a finite number of users the probability of success for a single terminal is the probability that it transmit and none of the other contending terminals transmit in the current R slot (Eq 4.8).

$$S_i = G_i \prod_{j \neq i} (1 - G_j) \quad (4.7)$$

$$\Pr\{success\} = p \left(1 - \frac{G}{M}\right)^{M-1} \quad (4.8)$$

The input load from the M users for a given speech statistic is equivalent to the throughput S_i . Assuming identical users, then probability of exactly k trials before a success (including the success) is geometrically distributed given as,

$$P_{tr} = \Pr\{K = k\} = \left[p \left[1 - \frac{G}{M}\right]^{M-1} \right] \left[1 - p \left[1 - \frac{G}{M}\right]^{M-1} \right]^{k-1} \quad (4.9a)$$

If p is large (>0.3), P_{tr} for $k>1$ must include the contention of backlogged users. For light loads, collisions are often caused by only two users, in which our tagged user will most likely collide with a new user. Hence the probability of success will be:

$$P_{tr} = \Pr\{K = k\} = \left[p \left[1 - \frac{G}{M}\right]^{M-2} (1-p) \right] \left[1 - p \left[1 - \frac{G}{M}\right]^{M-2} (1-p) \right]^{k-1} \quad (4.9b)$$

The $(1-p)$ term is the back-off probability of the colliding user. With the Immediate First Transmission (IFT) assumption, the first attempt has $p=1$. Therefore we have the following equations for probability of the number of retries, P_{ret} .

$$P_{ret}(r) = \left(1 - G/M\right)^{M-1} \quad \text{for } r = 1 \quad (4.10a)$$

$$P_{ret}(r) = [1 - P_{ret}(1)][P_{tr}(r-1)] \quad \text{for } r \geq 2 \quad (4.10b)$$

and

$$E[r] = \sum_{r=1}^{\infty} r \cdot P_{ret}(r) \quad (4.11)$$

The delay distribution of the S-ALOHA is plotted in Figure 4.4 together with simulation results for typical values of the arrival rate. It shows a linear PDF for the retransmissions in the logarithmic scale. The simulations also confirm the accuracy of the approximations.

Another famous CRA for S-ALOHA is the Binary Exponential Backoff . In this CRA, a user retransmit with a probability $1/2^k$ where k is the number of failed attempts experienced by the user. This protocol is used in the ethernet and proposed in many system (e.g. VRRM for GPRS, ETDMA, etc.). The approximation of this type of protocol differs slightly from that of the S-ALOHA previously analysed.

§4.3.2 Binary Exponential Back-off

Based on the method of successive trials, the access attempts occur as a random process. The length of each trial is a geometrically distributed random variable (see Figure 4.3). For Poisson arrival and no capture case, the probability of success, $P_{succ} = e^{-G}$. Thus the probability of delay equal to 1 slot is e^{-G} . Let P_n the probability that the success occur on the n^{th} trial, then we have

$$P_n = \begin{cases} e^{-G} & \text{if } n = 1 \\ \left(1 - e^{-G}\right) \left[\left((1-p)e^{-G}\right) \left(1 - e^{-G}(1-p)\right)^{n-2} \right] & \text{if } n \geq 2 \end{cases} \quad (4.12)$$

The term $(1-p)$ should have been $(1-p)^k$ where k is the number of colliding users excluding our tagged user. At low load region, k is usually one and most likely a new user (with a probability $\cong 0.9$) so that $(1-p)$ is $\cong 0.5$. The number of slots passed in each trial is geometrically distributed expressed as

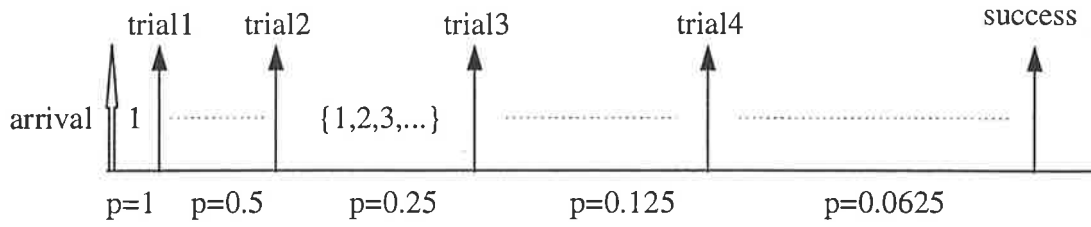


Figure 4.3 Diagram of the Binary Exponential Back-off

$$D_n^r = \Pr\{r \text{ slots occur within the } n\text{th trial}\} = 2^{1-n} \left(1 - 2^{1-n}\right)^{r-1} \quad (4.13)$$

where $D_n^0 = 0$.

Therefore the distribution of the delay after the n^{th} trial is the convolution of all the distribution of the preceding trials. Having $\mathbf{D}(f(1), f(2))$ as the discrete convolution of the discrete random variables $f(1)$ and $f(2)$, we have a recursion for the distribution as

$$D_n^s = \mathbf{D}(D_{n-1}^s, D_n^r) \quad (4.14)$$

where $D_1^s = \{0, 1\}$. We will then obtain the delay distribution of each user as:

$$D^s = \sum_n D_n^s P_n \quad (4.15)$$

The value $(1-p) = 0.5$ is compromise. Real values of p is less than 0.5 when $k \geq 1$. As k increases, p decreases which means this term is adaptive. When comparing the binary exponential back-off with the fixed retransmission probability it demonstrated a better PDF. However, the problem of this algorithm is the large delay variance due to the unfairness of the scheme.

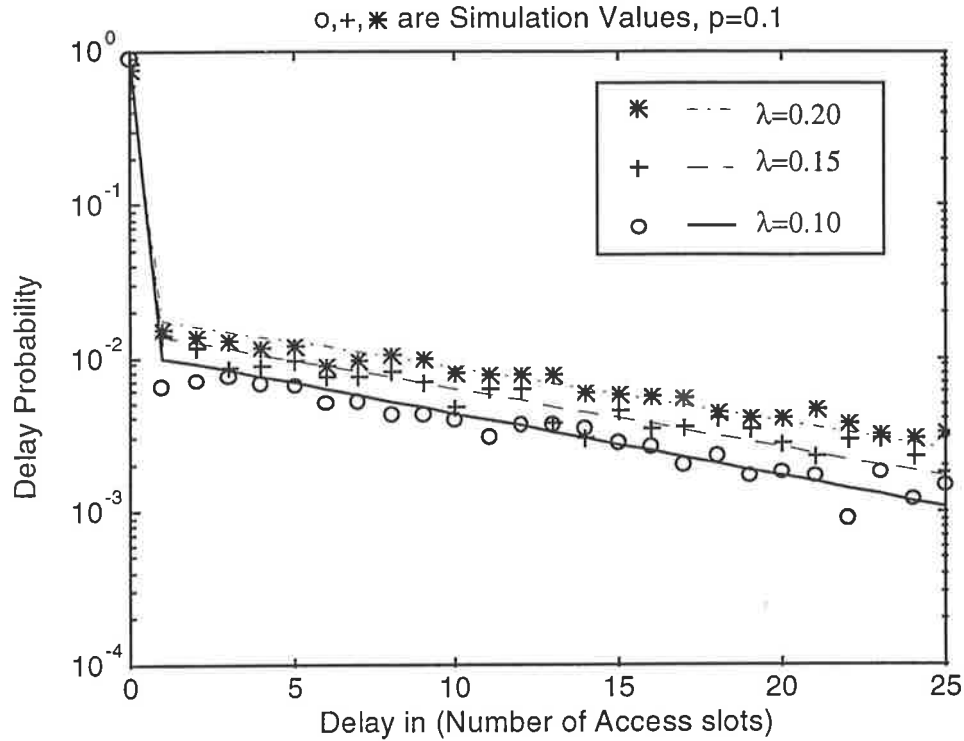


Figure 4.4 Delay Distribution of a Finite Population S-ALOHA

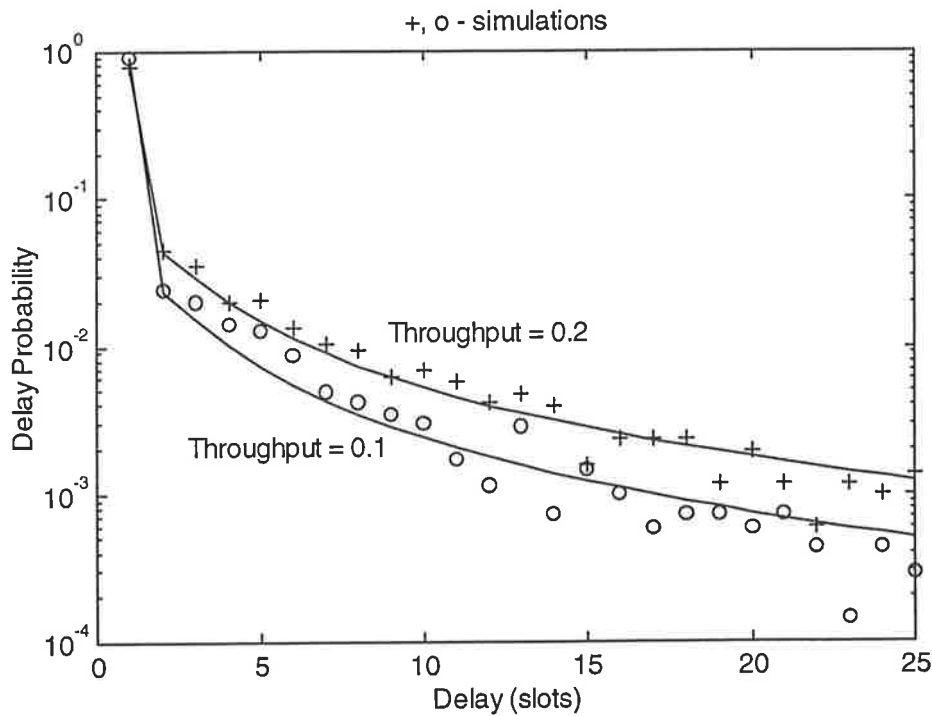


Figure 4.5 Delay Distribution of Binary Exponential Back-off

The delay distribution of S-ALOHA with fixed retransmission probability and binary exponential back-off were compared. The linear PDF of the fixed retransmission probability in the log scale is advantageous in the lower load region where large values of p is possible. In the binary exponential back-off shorter delays are achievable in the lower load but large delay variances are expected in the higher load region.

§4.4 Analyses of the Channel Allocation Queue

Most reservation protocols maintain a channel allocation queue which is necessary when the system is at higher loads where the requests are managed on a given service discipline. The queue can also be exploited in implementing priority schemes for users with different classes. Birth and Death Markov chains (eg. Erlang-C) are commonly used to approximate the performance of channel allocation queues. However, a system like R-TDMA can be suitably analysed using a Discrete Markov Chain since the arrivals actually occur at discrete time intervals. A frame by frame analysis is more appropriate in ATDMA since the slots in the frame are not the same. Unlike in PRMA, the analysis can be done on a slot by slot bases. The other advantage of this analysis is that we can extend it for multislot reservation schemes which will be shown in Chapter 6. The main purpose of providing a more accurate solution for the channel allocation process is because in data systems, the existence of the queue has a more dramatic effect on the delay performance compared to the contention process at higher loads. Another aspect of reservation based protocols is that the channel allocation process differs in many algorithms in which a more general solution is required as in this case. In this section, an approximation of the delay distribution of the queue is presented.

§4.4.1 System Model (*R-TDMA Channel Allocation*)

Our model consist of a fixed number of users and multiple server (timeslots) TDMA system with unlimited buffering capacity. Users are statistically multiplexed exploiting their on-off traffic behaviour (Quasi-random arrivals). Users that find all channels busy in the system are held in the queue with a given service discipline. We also assume that every user is allowed to reserve only one slot in every frame regardless of whether there are more available slots (i.e. voice users). The system model is depicted in Figure 4.6.

When solving the cumulative delay problem, it is necessary to calculate the distribution of the delay in the queue first since the distribution of the transmission delay is derived from the message length distribution. But first we must determine the probability that an arriving user finds exactly q users in the queue. In this case we need to use the Discrete Markov analysis.

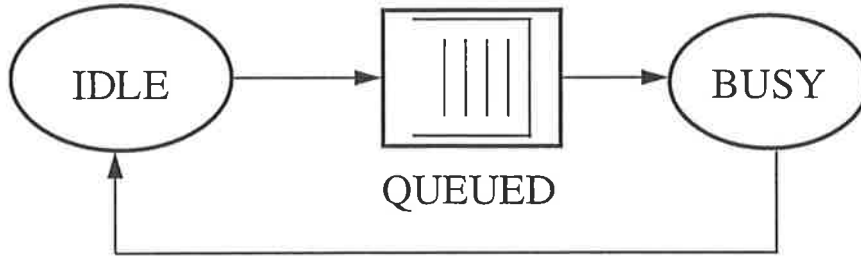


Figure 4.6 TDMA System Model

§4.4.2 Solving the Steady-State Occupancy

Prior to the calculation of the queuing delay distribution the system steady state distribution must be determined. Here we use a Discrete Markov Chain solution. From the model in §4.4.1 the system consist of three states. Therefore it is sufficient to describe the system state using two variables only. Let $\mathbf{P1}(q,r|Q,R)$ be the conditional probability that there are Q users in the queue and R number of users transmitting in frame k , and correspondingly q and r users in frame $k+1$. Then we have

$$\mathbf{P1} = \Pr\{Q_{k+1} = q, R_{k+1} = r \mid Q_k, R_k\} \quad (4.16)$$

Then the probability that there are q and r users in the queue and transmission states is expressed as

$$\pi(q,r) = \sum_{R=0}^N \sum_{Q=0}^{M-R} \mathbf{P1}(q,r \mid Q,R) \pi(Q,R) \quad (4.17)$$

conditioned that

$$\sum_j \sum_i \pi(i,j) = 1 \quad (4.18)$$

From the two expressions, we have to determine the joint one-step transition probabilities $\mathbf{P1}$ from the arrival and departure process in the system. Let T_I be the mean sojourn time (in frames) of a user in the idle state then the probability that a user leaves the idle state in the next frame is

$$\sigma_f = 1 - e^{-1/T_I} \quad (4.19)$$

and similarly the probability of a user leaving the transmission (busy) state is

$$\gamma_f = 1 - e^{-1/T_B} \quad (4.20)$$

Then the probability of x users becoming active in the incoming frame is

$$P_A = \Pr\{X = x \mid R, Q\} = \beta_{in}((M - R - Q), x, \sigma_f) \quad (4.21)$$

where $x \leq (M - R - Q)$. The number of users departing from the transmission state is only dependent on R . In this case our number of departures per frame is a r.v. y expressed as

$$P_D = \Pr\{Y = y \mid R\} = \beta_{in}(R, y, \gamma_f) \quad (4.22)$$

Since the transitions for the single slot reservation scheme is simply a function of the arrivals and departures within the system, then it is more convenient if we construct the arrival/departure probability matrix **AD** for the calculation of the transition probabilities. $AD(x, y)$ corresponding to the probability of x arrivals and y departures. Then the transition probability **P1** is calculated as follows: Firstly, we define a function $diag(M, d)$ which represent a vector of all elements in the diagonal of **M** taken at index, d . The index corresponds to the column position starting from the first column as $d=0$. Negative values apply to the indices to the left of the first column. In the calculation of **P1** we have seven different conditions as follows:

CASE 1 if $(q=Q); (r=R)$ conditioned that $Q=0$ or $R \leq N$

$$Pl(q, r \mid Q, R) = \sum diag(\mathbf{AD}, 0) \quad (4.23a)$$

CASE 2 if $(q=Q); (r>R)$ and $(Q=0)$

$$Pl(q, r \mid Q, R) = \sum diag(\mathbf{AD}, -(r-R)) \quad (4.23b)$$

CASE 3 if $(q=Q); (r<R)$ and $(Q=0)$

$$Pl(q, r \mid Q, R) = \sum diag(\mathbf{AD}, (R-r)) \quad (4.23c)$$

CASE 4 if $(q > Q)$; $(r = R)$ and $(R = N)$

$$Pl(q, r | Q, R) = \sum \text{diag}(\mathbf{AD}, -(q - Q)) \quad (4.23d)$$

CASE 5 if $(q < Q)$; $(r = R)$ and $(R = r)$

$$Pl(q, r | Q, R) = \sum \text{diag}(\mathbf{AD}, (Q - q)) \quad (4.23e)$$

CASE 6 if $(q < Q)$; $(r < R)$; $(R = r)$ and $(q = 0)$

$$Pl(q, r | Q, R) = \sum \text{diag}(\mathbf{AD}, (Q - q) + (R - r)) \quad (4.23f)$$

CASE 7 if $(q > Q)$; $(Q = 0)$; $(R < r)$ and $(r = N)$

$$Pl(q, r | Q, R) = \sum \text{diag}(\mathbf{AD}, -((q - Q) + (r - R))) \quad (4.23g)$$

Otherwise

$$Pl(q, r | Q, R) = 0 \quad (4.23h)$$

The value of π can be determined numerically (iteration) such that the difference between successive iteration is less than 0.001. After calculating π , we now can solve the distribution of the number of users in the queue. Thus an arriving users finds q users in the queue with a probability

$$\pi_q = \Pr\{q \text{ users in queue}\} = \sum_r \pi(q, r) \quad q = 0, 1, 2, \dots \quad (4.24)$$

A matrix based solution is also applicable if a direct relationship between q and r is available. In this case where single-slot per user is assumed, then q is directly related to the number of busy users. Therefore the steady state occupancy distribution is a well known

$$\pi = \pi \mathbf{P} \quad (4.25)$$

where $\mathbf{P}$ is the state transition probability matrix calculated as

$$P(i, j) = \begin{cases} \sum_{k=1}^i \beta_{in}(M - i, j - i + k, \sigma) \beta_{in}(i, k, \gamma) & \text{if } j \geq i \\ \sum_{k=1}^i \beta_{in}(M - i, k, \sigma) \beta_{in}(i, i - j + k, \gamma) & \text{if } j < i \\ 0 & \text{otherwise} \end{cases} \quad (4.26)$$

§4.4.3 Analysis of the Queue (wait and polled)

In the analysis of the queue, we are interested in the delay distribution which will then determine the waiting time before a user will be polled for reservation. It is also applicable in finding the delay distribution of the decrement of the queue for multiple classes (i.e. number of queued users in each class is known). The use of transforms may simplify the solution as used in [Lam77] [RubT89] [Rub79]. However, time domain analysis is aggressively considered here. The solution requires the queue transitions for a given queue position, departures in every frame and then combined with the distribution of the users in the queue as determined in the previous section.

The user's behaviour in terms of its change in positions in the queue depends solely on the number of users ahead in the queue and the number of departures in the particular frame. First, let us consider an arriving users that finds its position in the queue as $z=\{1,2,3..\}$. Given that one or more departures in the transmitting users occur, the queue position is decreased depending on the number of departing users. Thus the transition of the queue position forms a tree structure with $2^{(z-1)}$ possible transitions. A sample tree structure is shown in the next figure. From the figure, the tree can be generated recursively from an initial queue position, z . Correspondingly, a queue transition matrix $\mathbf{Tq}$ can be obtained whose elements are ones for all non-zero entries in the tree with the zeros for all other elements. Thus every row in $\mathbf{Tq}$ is a unique *path* of a user's transition in the queue.

If we let $\mathbf{J}$ be any path in the **tree** then we are interested in determining the probability that path $\mathbf{J}$ is used when a user leaves the queue. Since $\mathbf{J}$ is a row vector, the probability of a transition from $J(k)$ to $J(k+1)$ depends on the number of departures from the busy state. Because the frame is always full when the queue is not empty, then the transition probability from $J(k)$ to $J(k+1)$ is binomially distributed given as

$$Jp(J(k), J(k+1)) = \beta_{in}(N, [J(k) - J(k+1)], \gamma_f) \quad (4.27)$$

Considering the non-zero elements in Jp , the path occurrence probability is calculated as

$$P_j = \prod_{i=1}^{w-1} Jp(i, i+1) \sum_{j=0}^{N_z} \beta_{in}(N, J(w-1) - J(w) + j, \gamma_f) \quad (4.28)$$

4	0			
4	1	0		
4	2	0		
4	2	1	0	
4	3	0		
4	3	1	0	
4	3	2	0	
4	3	2	1	0

Figure 4.7 Tree Structure for $z = 4$ in the Queue Transitions

where $w = \text{length of } \mathbf{Jp}$, and $Nz + J(w) = \text{maximum possible jump in the queue}$. To calculate the delay distribution, we start by calculating the number of jumps, $\mathbf{Qj}$ in each path which is simply the transpose of the sum of the non-zero elements in each row in $\mathbf{Tq}$. So the maximum number of jumps is equal to z where the queue is decreased by one position at a time. We then construct the delay matrix $\mathbf{Dj}$ whose row is the number of jumps in the queue and the column as the queuing delay whose entries are the probabilities of having a queuing delay for the corresponding jump. $\mathbf{Dj}$ is calculated as

$$Dj(j,d) = \begin{cases} \binom{d-1}{j-1} ps^j (1-ps)^{d-j} & \text{for } d \geq j \\ 0 & \text{otherwise} \end{cases} \quad (4.29)$$

where $ps = 1 - \beta \ln(N, 0, \gamma)$. Then and the delay distribution is calculated as

$$D_z = P_j \sum_j Dj(j, d) \quad (4.30)$$

A plot for $\mathbf{Dj}$ is shown in Figure 4.8 which is compared with the simulation method. The plot shows a small discrepancy in the difference between the solution against that of the simulation results.

Another approach to solve $\mathbf{Dj}$ is by having a j -fold self convolution of the delay distribution in each queue position resulting to a Gaussian-like delay distribution. This behaviour can also be shown in the delay distribution from the negative binomial

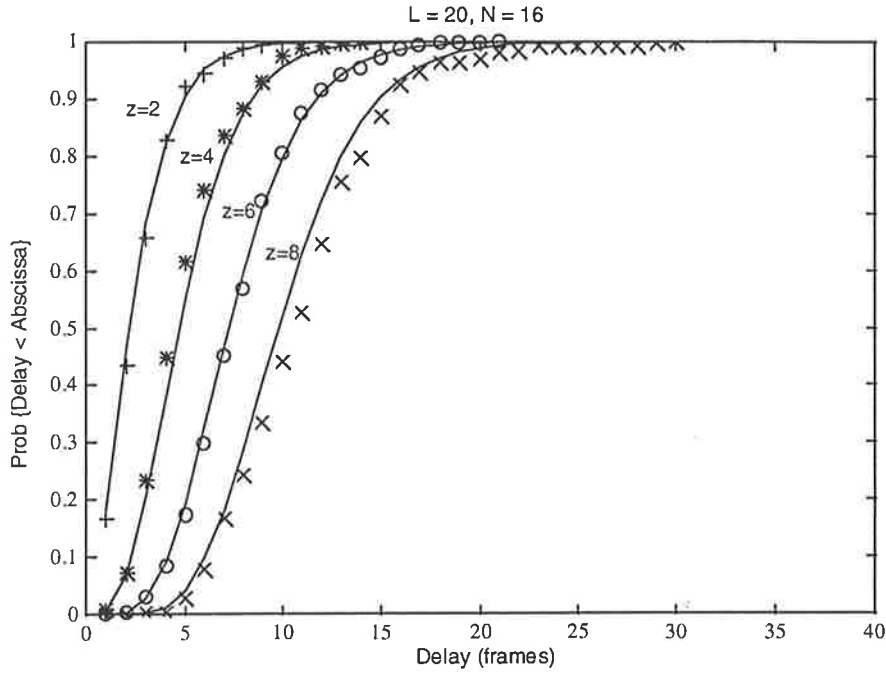


Figure 4.8 Delay Distribution for Different Values of Queue Position

distribution in Eq 4.29. After knowing the procedure in each path, we then average the distribution by considering all values of the arrivals z , with their corresponding probabilities from the result in the Markov Analysis in §4.4.2. When we assume that only one user will depart from the active state ($z+R$), the solution will be simplified. It will be then suitable for systems with very large low departure rates.

§4.5 System Model for R-TDMA Protocols

This section provides some system modelling and analyses for R-TDMA. As an example the ATDMA protocol was selected due to its simplicity. Before looking at the various performance measures, we first describe the ATDMA protocol. ATDMA is a three-stage reservation process. The slots are divided into access slots (R) and information slots (I) (partitioned frame). The I slots are distributed in the frame in order to allow fast access. By allocating appropriate I and R slots the delay caused by the contention and channel allocation process will be minimised. The frame structure of ATDMA is shown in Figure 4.9.

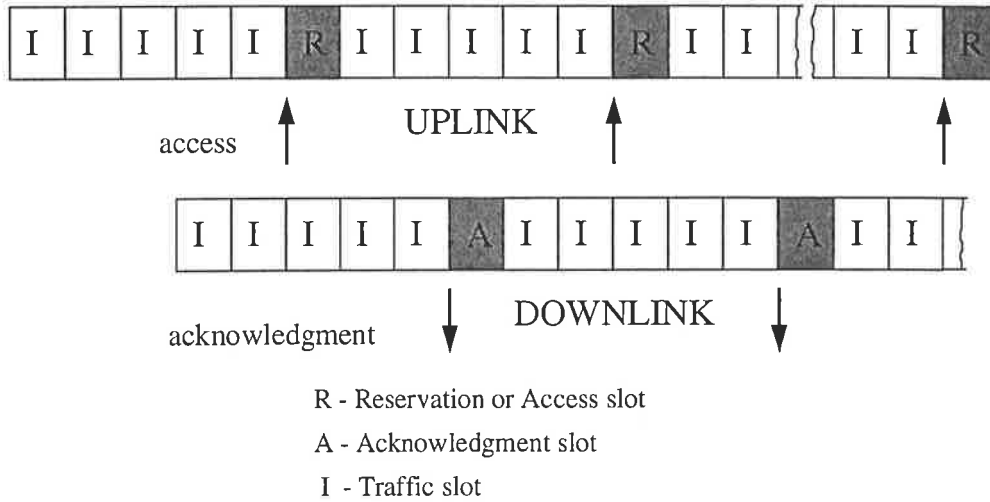


Figure 4.9 ATDMA Frame Structure

As mentioned in [Dunl94] [Dev93] the reason for allocating slots solely for reservation is to guarantee that the mobile terminals can gain access even in high load conditions. It also reduce the load of the downlink channels for the resolution of the conflicts during the contention procedure by having this function done mainly by the acknowledgment (A) slots. This will also ease the problem of providing a prioritisation. Also, ATDMA stability can be easily tuned since the contention process is done mainly on the *R* slots. In the frame structure in Figure 4.9 the uplink and downlink slots are provided with timing advance sufficient for the farthest terminal to receive the access feedback before the occurrence of the next *A* slot. The slots are paired (symmetric) which is essential for voice communications.

In the ATDMA system, every terminal that has either voice packets (VP) or data packets (DP) to transmit will contend for a reservation by sending a reservation packet (RP) immediately in the incoming *R* slot which carries all access information (e.g. terminal identity, service priority, message length for data, etc). Since the access contention procedure is Slotted ALOHA, when two or more terminals contend for the same *R* slot, collision will occur and none of the contending terminals will be successful (unless a capture mechanism is employed). A terminal that experiences a collision will retransmit for reservation in the next *R* slot with permission or retransmission probability, p uniform to all terminals. If the contending terminal cannot successfully access until a threshold expires, the packet in the front of the talkspurt (e.g. case of

voice) will be dropped and the terminal will continue to retransmit until it becomes successful.

After a terminal becomes successful in the contention process (after the receipt of an acknowledgment on the A slots) it will go to the channel allocation state waiting for a slot to be allocated. If an I slot is available a slot will be allocated to that particular terminal on the reception of the acknowledgment in the downlink frame. If no slot is available, it will be held in the channel allocation queue on a first come first serve basis or with a given priority. When a slot is allocated to a terminal, the terminal will continue to reserve that slot until the end of the information packet (VP or DP). Otherwise it will loose its reservation and the slot will be available for the first terminal awaiting in the channel allocation queue. Then at the end of the talkspurt, the terminal will go back to the silence state and repeat the process every time a talkspurt is generated. Our description of ATDMA here slightly differ from that of [Dunl94] due to the exclusion of the Fast Paging Acknowledgment (FPAck) in order to simplify our assumptions. A simplified Markov model of ATDMA is shown in Figure 4.10.

§4.5.1 *S-G Analysis*

The *S-G Analysis* is a useful tool in identifying the throughput variation of a protocol as a function of the offered load or the actual channel load. It predicts the favourable operating region based on the achievable throughput/capacity as a function of the system parameters. It also identifies which parameters the protocol is more sensitive. For example, the effect of channel partitioning and the sensitivity of a reservation protocol to the traffic statistics with the average message length in particular was obtained using this technique. In the analysis of protocols, there are two major assumptions to be considered, the infinite users case (with Poisson Arrivals) and the finite population case (with Quasi-random Arrivals). An infinite user case is always assumed for conservatism and the little difference with that of the finite population case especially when the number of users is quite large (i.e. more than 100).

To analyse the ATDMA protocol, we start with the busy users that are ready to transmit their reservation packets at the R slots. If we have K as a r.v representing the integer number of busy users awaiting for transmission in the incoming R slot, the probability of having exactly k users transmitting given that the users can transmit independently is expressed as

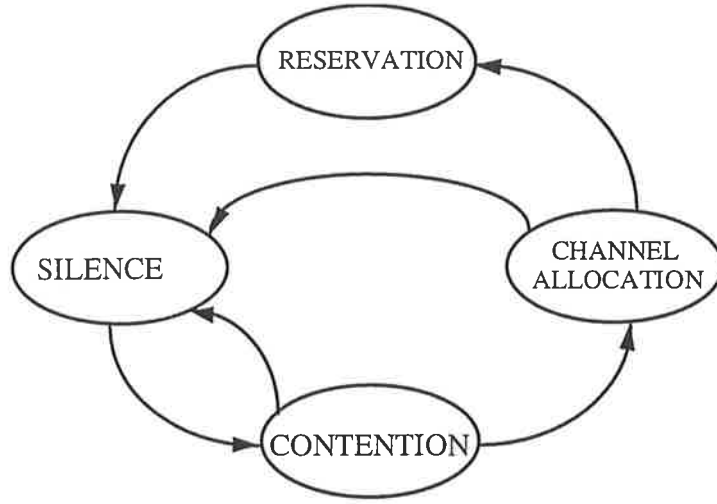


Figure 4.10 ATDMA System

$$P_k = \Pr\{K = k\} = \frac{G_a^k}{k!} e^{-G_a} \quad (4.31)$$

which is known as the Poisson Limit Formula. From the S-ALOHA receiver capture model, it is a conservative assumption that a successful packet transmission happens when exactly one terminal is transmitting in a particular R slot. Therefore, the throughput of the R slots is the probability of exactly one terminal contending.

$$S_a = \Pr\{K = 1\} = G_a e^{-G_a} \quad (4.32)$$

Since every successful user can transmit L packets of information, our normalised throughput equation is the ratio of the number of information packets transmitted in every frame to the total number of slots, N expressed as

$$S = \frac{\min\{L N_a G_a e^{-G_a}, (N - N_a)\}}{N} \leq 1. \quad (4.33a)$$

Similarly, for a finite population with a number of users, M which is much greater than the number of information slots I , we have

$$S = \frac{\min \left\{ L N_a G_a \left(1 - \frac{G_a}{M} \right)^{(M-1)}, (N - N_a) \right\}}{N}, \quad (4.33b)$$

Assuming that every access attempt has a potential of transmitting L packets, the total offered load formula results is expressed as

$$G = \frac{L G_a N_a}{N}. \quad (4.34)$$

For services where a very short queuing delay is required, then the server becomes a blocking system. As the output of an ALOHA system has a Poisson-like interdeparture times, then the resulting throughput will be a result of the contention and channel allocation blocking given as

$$S = L N_a G_a e^{-G_a} \left[1 - \text{ErlB} \left(L N_a G_a e^{-G_a}, (N - N_a) \right) \right] \quad (4.35)$$

where $\text{ErlB}(a, n)$ is the Erlang blocking formula expressed as

$$\text{ErlB}(a, n) = \frac{\frac{a^n}{n!}}{\sum_{j=0}^n \frac{a^j}{j!}}. \quad (4.36)$$

The results of Eqs. 4.31 to 4.36 are shown in Figure B1 to B3 in the Appendix. It shows that the ATDMA protocol is subject for optimisation with the number of R slots as well as the average message length, L as the key parameters.

§4.5.2 Effect of Retransmission Probability

The equations in §4.5.1 generalised the S-G formula of a reservation protocol based on S-ALOHA. However, they exclude the effect of the retransmission probability, p . While it is obvious that the S-ALOHA system is greatly affected by the retransmission probability, an approximate S-G analysis shown here is based on the following model (see Figure 4.11).

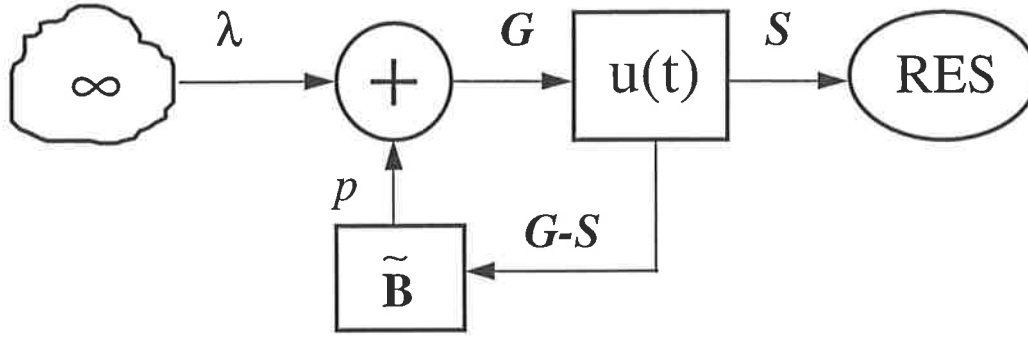


Figure 4.11 Model for S-G Analysis including the Retransmission Probability

The effect of p can only be accounted for if the backlogged users can be estimated. Here we used the constant throughput assumption so that arrival rate λ is equal to the throughput S . Thus we can loosely approximate the backlogged users, $\tilde{B}$ as $(G-S)/p$. Solving the throughput equation we have

$$\begin{aligned}
 S_a &= \Pr \{ 1 \text{ user becoming busy \& no backlogged user retransmits} \} \\
 &\quad + \Pr \{ \text{no user becoming busy \& 1 backlogged user retransmits} \} \\
 S_a &\approx \lambda e^{-\lambda} (1-p)^{\tilde{B}} + \tilde{B} p (1-p)^{\tilde{B}-1} e^{-\lambda}.
 \end{aligned} \tag{4.37}$$

By changing the parameters we arrive at

$$S_a = S_a e^{-S_a} (1-p)^{\left(\frac{G_a - S_a}{p}\right)} + \left(\frac{G_a - S_a}{p}\right) p (1-p)^{\left(\frac{G_a - S_a}{p}\right)-1} e^{-S_a}. \tag{4.38}$$

Since S_a appears at both sides of the equation, a numerical solution is readily available by solving $S_{a(t)} = f(S_{a(t-1)}, G_a, p)$ which converges to S_a for all non-negative values of G_a . Similarly, the throughput formula is given as

$$S = \frac{\min\{L N_a S_a, (N - N_a)\}}{N} \tag{4.39}$$

and the results for the different values of p are plotted in Figure B.4 in Appendix B.

There are two performance limitations in the ATDMA protocol. One is the S-ALOHA contention process which is well understood as having a bi-stable operating characteristics [Git75]. The other is the limited channel availability of a M/M/N configuration. The framing of ATDMA is in fact not favourable to data services since a M/M/1 system is widely accepted as the best configuration for data. Thus it is not surprising that the ATDMA or TDMA in general has an inferior performance compared to other M/M/1 type protocols like the R-ISMA and R-BTMA as ATDMA is optimised for voice transmission characterised by fixed rate and steady traffic.

§4.5.3 Stability of ATDMA

The retransmission probability not only affects the S-G performance of ATDMA but more importantly affects the stability of the system. The stability of ATDMA coincides with the stability of the S-ALOHA contention process which is determined on the R-slot basis. In addition to this problem, the stability of S-ALOHA is decreased if a bursty traffic is supported (see Appendix C) such that the design of ATDMA must incorporate a stability criteria. In this section, the *drift* (rate of increase/decrease in the number of backlogged users) parameter was used and the calculation is as follows.

If τ_R is the distance between two access slots (subframe), then σ_a and γ_a are the transition probabilities from the idle and reservation states at the end of the each subframe given as

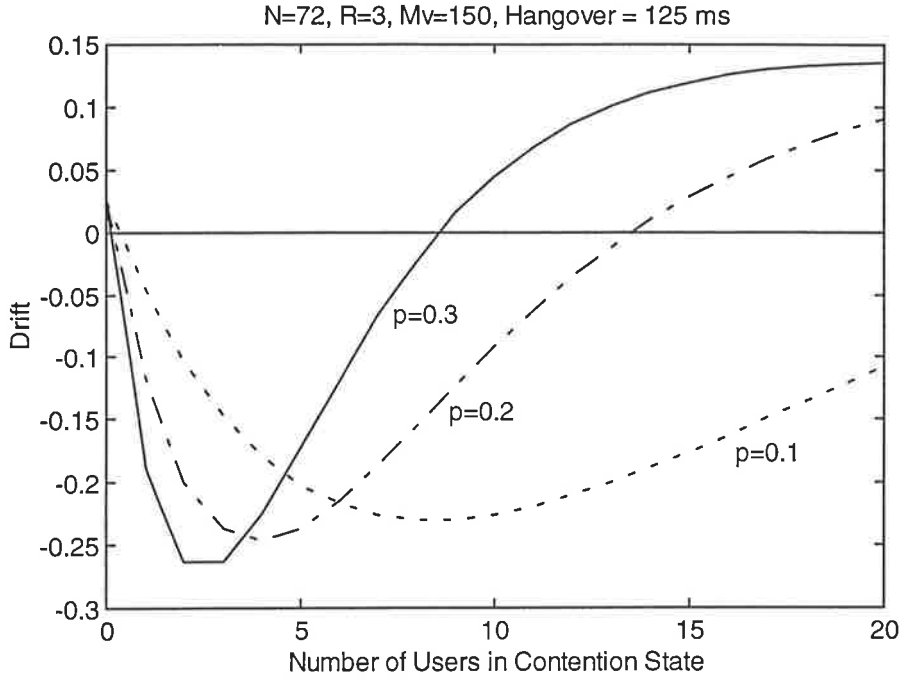
$$\sigma_a = 1 - e^{-\tau_R/T_s} \quad (4.40)$$

$$\gamma_a = 1 - e^{-\tau_R/T_t} \quad (4.41)$$

and the probability of a user being idle is

$$\alpha = \frac{T_s}{T_t + T_s} \quad (4.42)$$

For M users and n backlogged users, we have a maximum of $M-n$ idle users. The number of idle users is only dependent on α and is binomially distributed. Calculating the departure rate from the backlogged state, we have

**Figure 4.12 ATDMA Stability**

$$\begin{aligned}
 \mu(n) &= \Pr\{\text{success} | n \text{ backlogged users}\} \\
 &= \sum_{s=0}^{M-n} \beta_{in}(M-n, s, \alpha) \left[\beta_{in}(s, 1, \sigma_a) (1-p)^n + \beta_{in}(n, 1, p) (1-\sigma_a)^s \right]
 \end{aligned} \tag{4.43}$$

Similarly, the number of users becoming backlogged is

$$\begin{aligned}
 \lambda(n) &= \sum_{s=1}^{M-n} \sum_{j=1}^s j \beta_{in}(s, j, \sigma_a) \beta_{in}(M-n, s, \alpha) \\
 &= \sum_{s=1}^{M-n} [s \sigma_a] \beta_{in}(M-n, s, \alpha)
 \end{aligned} \tag{4.44}$$

resulting to the drift formula for n backlogged users as

$$Drift(n) = \lambda(n) - (\mu(n) + n\gamma_a). \tag{4.45}$$

The last term is for users that depart the backlogged state prior to a successful contention (e.g. after a time-out expires). We test the stability based on typical system parameters. The characteristics of ATDMA is shown in the next figure. The zero crossings (start of unsafe operating point) of the tested parameters of $p=0.3$ is only

around 8 while for $p=0.1$ is beyond 20. However, larger p values correspond to smaller contention delays. Thus a trade-off between the capacity and stability must be considered in the design of ATDMA systems.

§4.5.4 Mean Delay Analysis of ATDMA

The S-G analysis is only concerned with the calculation of maximum capacity or throughput as a function of the offered load. Whilst it is useful in the tuning of ATDMA, the real capacity is measured with respect to the quality of services that has to be maintained. In almost all applications, the delay performance is essential so that the throughput/delay characteristics of ATDMA must be determined. Here, a Transient Fluid Approximation (TFA) for solving the steady-state performance of ATDMA was presented. The main objective in this analysis is to solve the throughput/delay characteristics of the system under some assumed input traffic statistics and channel configuration.

TFA is appropriate when the system exhibits a Markovian property. This criteria is subject to the input traffic statistics and the nature of the processes of the protocol. To start with our analysis, we consider the system model is Figure 4.13. In TFA, we are concerned with calculating the steady state system parameters. For the ATDMA protocol with idle, contention (backlogged), queued and reserve states, the state vector is defined as $\xi = \{m_1, m_2, m_3, m_4\}$ which are the corresponding number of users in each of the states. The steady state vector exist if the system has equilibrium point(s).

TFA involves the calculation of the steady state vector based on 1) initial system state 2) transition probabilities and 3) a particular stable point [Wu94a]. The system state is an imbedded random process which chooses the start of each frame as the imbedded points. Since the method of calculation is numerical (iteration) from the initial state where the system state at the k^{th} imbedded point is only a function of the system state at the $(k-1)^{th}$ imbedded point, the process is Markovian and hence we are dealing with an Imbedded Markov Process. In solving the system state in each imbedded point $\xi(k)$ we need to identify the transition probability matrix from the model. The choice of the imbedded points to be at the beginning of each frame is to enable to formulate the transition probabilities caused by the channel allocation process which takes into account the status (reserved or free) of all slots in each frame.

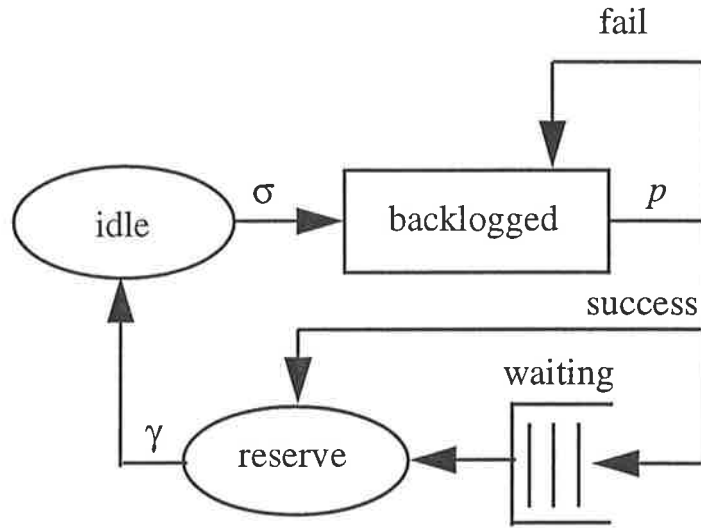


Figure 4.13 ATDMA System Model

Taking into consideration an ATDMA system with Na access slots per frame and Nt traffic slots. Let σ be the transition probability from the idle to the contention state or the packet generation rate of each user, μ be the probability that a user leaves the backlogged state, and γ the probability that a user leaves the reserve state. Then the equilibrium equations are as follows:

$$m_{1(k)} = (1 - \sigma)m_{1(k-1)} + \gamma m_{4(k-1)} \quad (4.46)$$

$$m_{2(k)} = m_{2(k-1)} + \sigma m_{1(k-1)} - \mu_{(k-1)} \quad (4.47)$$

where

$$\mu'_{(k)} = \left[p m_{2(k-1)} (1 - p)^{m_{2(k-1)} - 1} + \beta \text{in}(m_{1(k-1)}, 1, \sigma) (1 - p)^{m_{2(k-1)}} \right] \quad (4.48a)$$

$$\mu_{(k)} = \min \{ m_{2(k-1)} + \sigma m_{1(k-1)}, \mu'_{(k)} Na \} \quad (4.48b)$$

Finding the transition rates for the queue and the reserve states is not so trivial since the queue and the channel occupancy are dependent. So we need to determine the sum of these two states as m_{34} . From an initial state vector of 3-state system (combine queue and reserve states), $\xi'(0)$ the m_1 , m_2 and m_{34} can be determined. Then m_{34} can be decomposed into m_3 and m_4 from an assumed service discipline of the queue (eg. FCFS single slot reservation). In this case, $m_4 \leq N$ such that a slight difference in the transition

rate in Eq.4.46 is expected. Since the output of the S-ALOHA contention process has a Poisson-like interarrival time, then we can model the queue and the reserve state as a M/M/c/p queue and the delay can be solved using Little's result. In the iteration, the decomposition of the queue is as follows:

Let P_c = channel occupancy distribution of the queue, then we can solve the number of transmitting users as

$$m'_{4(k-1)} = \sum_{j=0}^{N_t-1} j P_c(j) + N_t \left(\sum_{j=N_t}^M P_c(j) \right) \quad (4.49)$$

$$m'_{3(k-1)} = \sum_{q=0}^{M-N_t} q P_c(q + N_t) \quad (4.50)$$

Normalising, the mean number of transmitting users in the current imbedded point is

$$m_{4(k)} = m_{34(k-1)} \frac{m'_{4(k-1)}}{m'_{4(k-1)} + m'_{3(k-1)}} \quad (4.51)$$

then m_{34} becomes

$$m_{34(k)} = m_{34(k-1)} - \gamma m_{4(k-1)} + \mu_{(k)} \quad (4.52)$$

where P_c is calculated using the quasi-random input and delayed users model [Coop72]

$$P_c(j) = \begin{cases} \binom{M}{j} a'^j P_c(0) & \text{if } j \leq N_t \\ \frac{M! a'^j}{(M-j)! N_t! N_t^{j-N_t}} P_c(0) & \text{if } j > N_t \end{cases} \quad (4.53)$$

where

$$P_c(0) = \left[\sum_{j=0}^{N_t} \binom{M}{j} a'^j + \sum_{j=N_t}^M \frac{M! a'^j}{(M-j)! N_t! N_t^{j-N_t}} \right]^{-1} \quad (4.54)$$

and

$$a' = \frac{1}{\gamma \left[\frac{1}{\sigma} + T_{backlogged} \right]} \quad (4.55)$$

$$a' = \frac{1}{\gamma \left[\frac{1}{\sigma} + \frac{m_{2(k-1)}}{\sigma m_{1(k-1)}} \right]} \quad (4.56)$$

After the transition rates are determined, the throughput and delay are simply calculated as

$$\text{Throughput, } S = \frac{m_{1(\infty)} \sigma L}{N} \quad (4.57)$$

$$\text{Access Delay, } D_a = \frac{m_{2(\infty)} + m_{3(\infty)}}{\sigma m_{1(\infty)}} \quad (4.58)$$

$$\text{Message Delay, } D_m = \frac{M - m_{1(\infty)}}{\sigma m_{1(\infty)}} = \frac{M}{\sigma m_{1(\infty)}} - \frac{1}{\sigma} \quad (4.59)$$

and a sample plot is shown in Figure 4.14. The plot shows that the contention (S-ALOHA) delay is quite steady at increasing load but the channel allocation delay grows at throughput in the region above 0.8. This is typical in M/M/n/p queue.

§4.6 Summary

In this chapter, analysis methods for R-TDMA were presented. ATDMA performance was evaluated so that the important parameters can be determined in the design of R-TDMA. In the first part of this chapter, R-TDMA system models were identified and some analyses methods were discussed. In §4.3, an approximate model based on *successive trials* for Slotted Random Access protocols was derived as random access protocols will play a major role in the design of R-TDMA. A solution for S-ALOHA with fixed retransmission probability and binary exponential back-off were presented that exhibit a good agreement with the simulations. Then in §4.4, a Discrete Markov Analysis is suggested to model the channel allocation process with users having high transition probabilities as opposed to the traditional Birth and Death Markov chains.

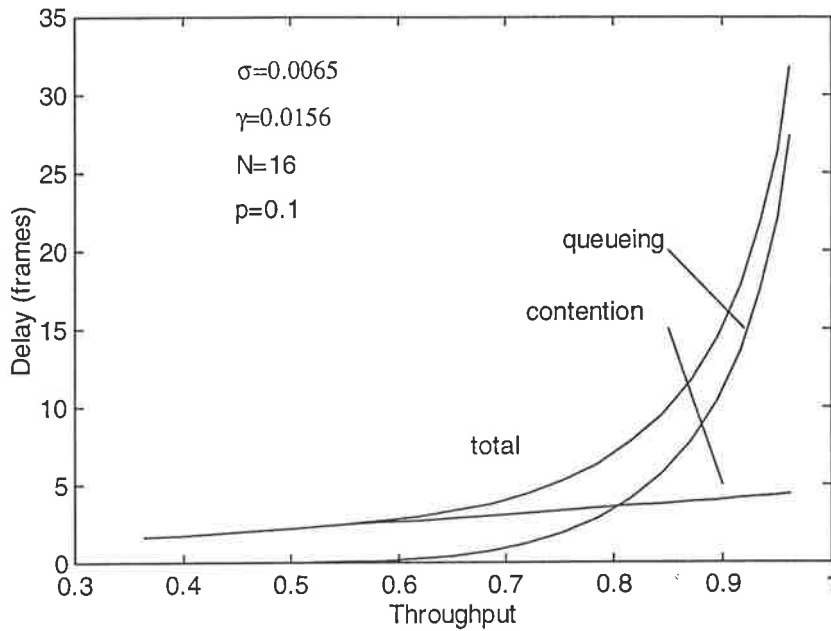


Figure 4.14 ATDMA Performance Using TFA

A sample plot of an ATDMA system performance using the TFA. In this method, the message or total delay is the sum of the contention and queuing delays. The iteration is calculated based on 50 users and a single access slot per frame ($N_a=1$).

A simplistic time domain analysis based on a binary tree transition for the queuing delay distribution is also presented which is again sustained by simulation results. Later in §4.5 an analysis of ATDMA was presented which included the S-G analysis for the different frame partitioning and the effect of the retransmission probability. Then the stability based on the drift parameter was briefly introduced and in the later part the mean throughput/delay analysis based on TFA was applied to ATDMA. The unique feature of the TFA method in ATDMA is the inclusion of the analysis of the queue in order to obtain higher accuracy.

Chapter 5

Reservation-TDMA Protocols for WPC

Design and Performance Analysis

This chapter proposed some channel access methods and protocols for WPC. The schemes to improve the reservation protocols are divided between the access mechanisms and the resource allocation procedure. An optimisation for the ATDMA protocol is also presented which includes both the frame structure as well as controlling the traffic statistics. Prioritisation is also given with much importance both in the channel access and in the channel allocation to ease the problem of accommodating multimedia services in the wireless link. As a first step in evaluating the performance of a packet access protocol for WPC, packet voice traffic is firstly considered. This is due to the strict delay requirement of voice and circuit-switched oriented services. For example in the packet voice system, a terminal must establish a connection within 32 milliseconds [Goodm89] right from the start of the talkspurt arrival. This has also to take into account the delays incurred in the packetisation of the first speech packet, i.e. sampling, quantisation, coding and the latencies in the vocoder. It should also be noted that in the real application, other factors may affect the quality of voice transmission like the level of background noise and the characteristics of the microphone employed at the voice terminal. In the first part of this chapter, the Reservation-TDMA access mechanism for WPC is described

§5.1 WPC with Reservation-TDMA Multiaccess Protocol

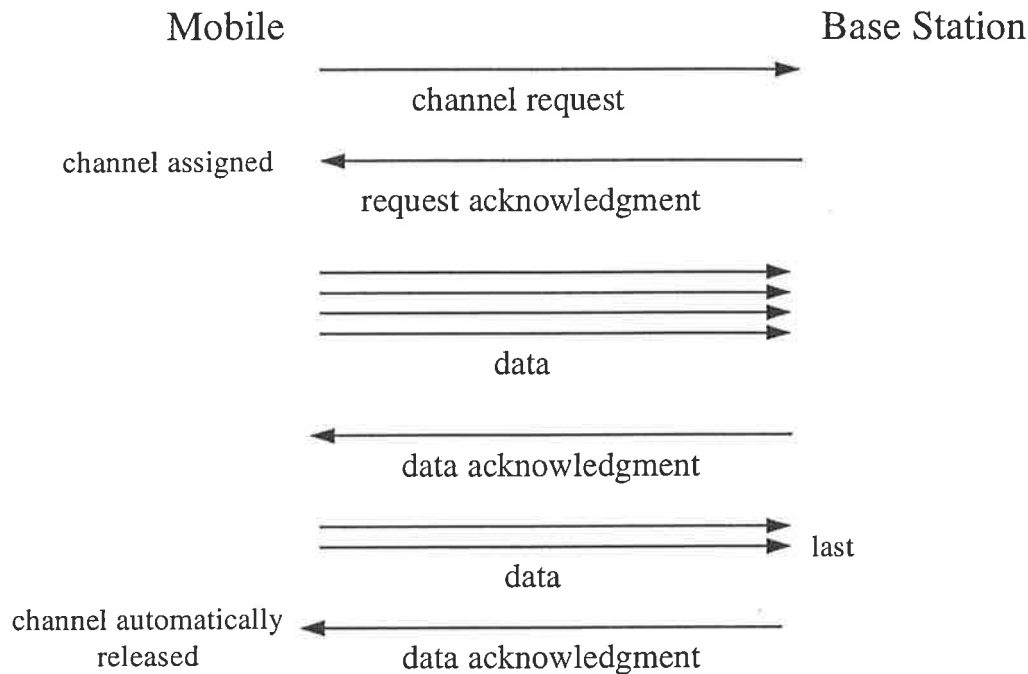
The concept of Reservation-TDMA is to provide an effective way of multiplexing various services in the wireless environment based on packet-switched TDMA technology. To achieve this, the radio channel barrier must be overcome which means, the upper layer and middle layer protocols must compensate for the characteristics of the underlying physical channel. With the channel structures described in Chapter 4, the unit of transmission for R-TDMA is a *burst* which consists of one slot duration. The length of a burst is optimised based on the physical layer characteristics of the radio channel (i.e. not too long to penalised the synchronisation and not too short to reduce

the number of information bits). Every burst (slot) is protected by a training sequence and depending on the length of the burst, it can be divided into subfields where each field consists of its own training sequence and an information field. A delimiter and a guard time is also used to indicate the end of a burst.

Usually a burst contains only a few tens or hundreds of bits sufficient to carry a short signalling information. So there is a need to increase the transmission units in order to transmit some data and other signalling information. The best option is to group the bursts into a larger information unit referred to as a radio block. The radio block not only increases the information unit size but also introduce its own error protection mechanism by employing a burst interleaving and an error check (block error check sequence) in addition to the forward error correction that is almost a requirement in the radio channel. Usually, the size of the block as well as the frame period are obtained from the speech codec sampling rate and speech frame size. This is due to the strict delay requirement of voice service and the steady nature of its traffic (continuous stream).

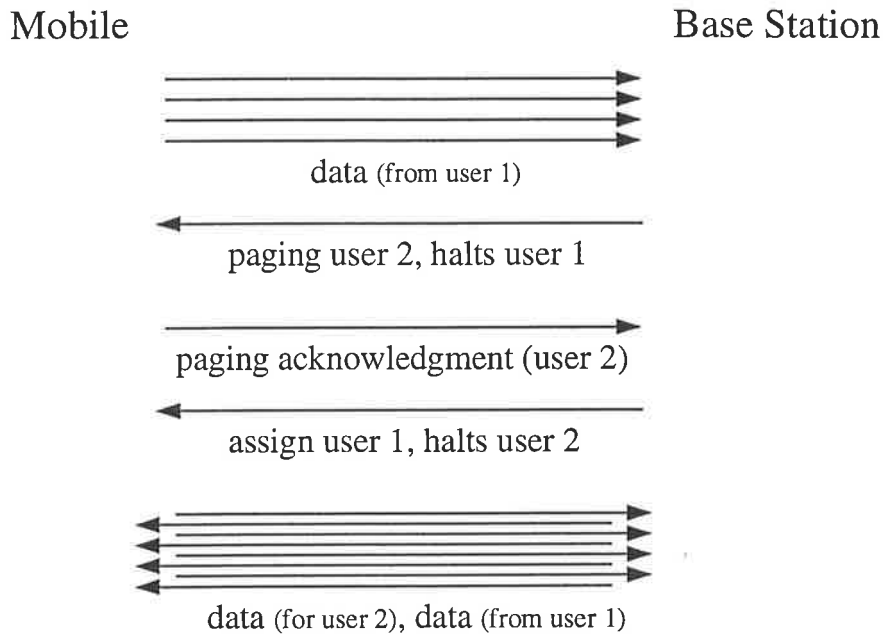
Fast access and fast acknowledgment is one requirement of R-TDMA in order to support packet switching. To achieve this, the request or reservation packet (RP) and the acknowledgment packet (AckP) sizes must fit into one or two bursts length (preferably one) so that the request or acknowledgment can occur in at least one frame period. To reduce the latency, a user will transmit immediately in the first incoming slot of the assigned timeslot after an acknowledgment is received. This means, the first incoming slot must be the start of the block in contrast to a fixed block arrangement of a fixed multiframe sequence.

At the multiaccess layer of R-TDMA, the basic uplink signalling messages are composed of a channel request message, uplink paging acknowledgment and an uplink transmission acknowledgment. The basic downlink signalling messages are channel request acknowledgment, channel assignment message, paging message, transmission acknowledgment and broadcast of channel and multiaccess parameters. An example of a signalling involved in an uplink transmission of an acknowledged data transfer in R-TDMA is as follows:



The example above indicates three important fast signalling messages. Other signalling required for some transmission abnormalities are not shown. The channel request and the request acknowledgment requires a common channel accessible by all users while the data acknowledgment can be transmitted on a common channel or on the downlink pair channel called an associated channel like in [GSM 5.01]. If a common channel is used for the data acknowledgment, a user has to compete with the rest of the users so that is preferable to assign an associated channel for this purpose. Thus an uplink user has to constantly monitor the downlink channel pair in order for the user to identify specific signalling. Other information concerned on the physical layer are to be transmitted on the associated channel. These signalling messages are scheduled as a form of a stealing frame if in case the downlink channel is used for downlink data transfer.

The use of an associated channel becomes necessary in the uplink channel because of the difficulty in acknowledging the downlink transmissions. The other possibility of uplink acknowledgment is via a random access together with the channel requests. Knowing that a random access is reserved for channel request due to its lower throughput, this approach is discouraged. However, the use of an associated channel in the uplink requires some user scheduling. In this case, the downlink transmission must regularly identify the user currently allowed to transmit in the uplink (acknowledgment



of one user or data from another user) in which the use of temporary assignment becomes necessary. An example of this scenario is depicted in the illustration above where user 1 and user 2 are scheduled by the base station in favour for signalling messages.

§5.1.1 Logical Channel Structure

In order for the protocol to work, some logical channels were identified. First is a broadcast channel (BCH) in the downlink to identify the TDMA frame structure, location of the logical channels as well as common information with regards to the location, frequency, resources, etc. The broadcast channel is omitted in the frame structure in Chapter 4 because a radio channel may or may not consist of a broadcast channel. In this way, a user has to monitor only one radio channel which consists of a BCH then move to another radio channel once the broadcast information is received. Secondly, the performance analysis is always concentrated on the uplink due to the difficulty in detecting the busy users as well as the scheduling of user reservation.

Another logical channel is the access or reservation channel (RCH) used for channel request in the uplink. The number of access channels and their locations are identified by the broadcast channel. Next is the Permanent Acknowledgment Channel (PACH) in the downlink used to acknowledge the access requests as well as assign the

uplink free traffic/associated channels (TACH) to the users requesting them. The TACH is mainly used for the transmission of user information and user specific signalling. The TACH is dynamically used for traffic and associated signalling information. Uplink acknowledgments are also transmitted on the TACH by the use of user scheduling previously described. Lastly, a paging channel (PCH) is required in the downlink to transmit the recipient for the downlink transmission in each PACH. In this way, every user can identify the which timeslot to listen in every frame. The RCH, PACH and the TACH are located respectively in the *R-slots*, *A-slots* and *I-slots* of the TDMA frame structure.

§5.1.2 R-TDMA Support for Voice Traffic

Packetised speech with on-off pattern can be easily accommodated in the R-TDMA because of the abundance of RCH hence providing a fast access mechanism. A voice user performs a reservation process in every talkspurt in the uplink. It starts by transmitting an RP on the first incoming RCH. Once successful, an acknowledgment is scheduled in the next incoming PACH together with a slot assignment if a free slot TACH is available or it will be held in a queue until a free TACH is available. When a user is assigned with a TACH, it will start transmitting right at the first incoming slot which is also the start of the first block. For voice traffic, only one user is allowed to reserve the TACH for the transmission of information messages. However, it can be interrupted by the transmission of signalling messages or acknowledgments from another user listening to the downlink pair of the particular TACH. This happens because the downlink pair could be assigned to another user. The reservation of the TACH expires automatically upon the transmission of the last speech block. In the downlink direction, the users listen to the PCH all the time in order to identify which slot are the intended downlink messages transmitted. Once the user is paged, it listens immediately to the assigned slot in the paging signal and the base station immediately starts transmitting the speech frames (polling mechanism). At the end of the talkspurt, the channel becomes free automatically (i.e. no acknowledgment for voice). In this way, an uplink channel owned by a certain users has a downlink pair channel assigned to another user. This is the main difference of the circuit-switched TDMA and the packet switched R-TDMA.

§5.1.3 R-TDMA Support for Data Traffic

The transmission of data is similar to that of the voice traffic except that the data user can actually reserved specific amount of blocks as identified by its message size stored in the buffer. This is because a data user does not request for a channel unless it has stored some information that are ready-to-transmit. In this case, it is possible to assign one or more users per TACH and schedule their transmission depending on the requests. Unlike the voice traffic, this scheme requires a reservation policy that can be changed dynamically by the base station. This is done by assigning a channel immediately after the channel request but schedules the transmission at a predefined period. Secondly, periodic data traffic can be transmitted properly by scheduling the transmission. This scheme can also enhance the prioritisation of heterogeneous users. And lastly, the other requirement for data is to provide a periodic acknowledgment of the transmitted blocks in order to guarantee an error delivery by retransmitting the erroneous blocks. For the downlink transmission, data is easily scheduled by the base station.

§5.2 R-TDMA Performance with Packet-Voice Traffic

Here we present an approximation of the performance of voice-only RTDMA. We use the ATDMA frame structure due to its simplicity. Some analyses are found in [Dev93], [Mitro93] while in [Dun193] the performance evaluation mainly rely on simulations. The analysis in [Mitro93] used a discrete Markov chain. Our approximate analysis considered here decoupled the contention phase to the channel allocation phase. The reason is to identify the system performance attributed by each process. Moreover, the contention process of a voice traffic is stationary since voice packets (VP) are dropped if they are not transmitted in a short waiting time threshold. This refers to VP's in front of the talkspurts that experience a period equivalent to the delay threshold both in the contention and channel allocation processes. As seen in Figure 4.10 some users may return to the silent state from the contention or channel allocation states. From the frame structure of ATDMA an optimal combination of R and I slots is necessary [Dev93] and is also identified in the S-G analysis in §4.5.1. In this section a simple method to determine the near optimal channel structure (R/I slots) is presented. The analysis is based on a targeted number of users that can be supported as a function of the percentage of speech packet dropped [Goodm90]. The novel idea in this

optimisation is to include the speech hangover parameter in order to achieve the best system performance. This is based on the idea where sufficient buffering (hangover in speech) can improve the performance of reservation based protocols as the temporal speech parameters varies. In the calculation of the packet dropping probability or percentage of speech packets dropped, the delay distribution in each of the contention and channel allocation phase are important because a packet is dropped only after the tolerable delay expires. A rough approximation method for ATDMA capacity with voice traffic is shown in the Appendix D.

§5.2.1 Contention Process

The approximation of the packet dropping probability due to contention can be derived from the behaviour of the S-ALOHA contention process. From §4.3 the equation for a non-capture case with fixed number of voice users, M_v for the delay distribution is given in Eq.4.10a and Eq.4.10b. Similarly, the talkspurt delay distribution is

$$P_{cd}(k) = \Pr\{\text{delay} = k\} = \begin{cases} (1 - G/M_v)^{M_v-1} & \text{for } k = 1 \\ \left[1 - (1 - G/M_v)^{M_v-1}\right] [P_{tr}(k-1)] & \text{for } k \geq 2 \end{cases} \quad (5.1)$$

where P_{tr} is given in Eq. 4.9b and G is the aggregate load for all users with throughput S_c in the relation

$$S_c = G e^{-G} \quad (5.2)$$

and the constant throughput which is a function of the users activity is given as

$$S_c = \frac{M_v}{R (T_t + T_s)} \quad (5.3)$$

Then the equivalent number of frames that are passed prior to the success is

$$n_f = \sum_{k=1}^{\infty} k \sum_{j=(k-1)N_a}^{kN_a-1} P_{cd}(j) \quad (5.4)$$

where n_f is the number of frames passed and N_a is the number of R slots and also the number of A slots. Usually, a one-frame delay threshold is used so that the packet dropping probability is the ratio of n_f to the VP length given as:

$$Pdc = n_f \frac{\tau}{T_t} \quad (5.5a)$$

similarly we can calculate the mean number of dropped packets from the $E[r]$ as

$$Pdc = \frac{E[r]}{N_a} \frac{\tau}{T_t} \quad (5.5b)$$

From the average number of retransmissions, we can obtain the percent of packet dropping from the retransmission probability p and the number of R slots per frame. The main reason for choosing S-ALOHA for the contention process is its very short delay at lower load region. Since it is expected that for very small voice packet dropping probability, the load is very low (around 0.1), the problem of stability is not so serious in voice ATDMA.

§5.2.2 Channel Allocation Process

The channel allocation packet dropping probability is the probability that an arriving packet will experience a delay more than the threshold (one frame duration). In the case where the waiting time threshold is zero, the packet dropping probability will be the same as the packet delay probability. Otherwise, the probability that the waiting time, w is greater than t is defined from the Erlang C model (Eqs. 5.6 - 5.8).

$$P(w > t) = P(w > 0) \cdot P(w > t | w > 0) \quad (5.6)$$

$$P(w > 0) = \text{ErlC}(Ao, N) \quad (5.7)$$

$$P(w > t | w > 0) = e^{(\rho - 1) N \mu t} \quad (5.8)$$

where $\rho = Ao/N$, μ is $1/T_t$ and the Erlang C formula is [Coop72]

$$ErlC(Ao, N) = \frac{Ao^N / [(N-1)!(N-Ao)]}{\sum_{k=0}^{N-1} \left(Ao^k / k! \right) + Ao^N / [(N-1)!(N-Ao)]} \quad 0 \leq Ao < N \quad (5.9)$$

The probability of delay takes into account that a good exponential fit from the talkspurt length distribution exist. Therefore this is a valid assumption. In the case where the number of users are quite small (say ATDMA with 16 or 32 slots), it is more advisable to use the Engset Delay model. The equations are as follows:

$$P(w > t) = c \sum_{j=0}^{M-N-1} \frac{g(t)^j}{j!} e^{-g(t)} \quad (5.10)$$

$$g(t) = \frac{N\mu}{\gamma} + N\mu t \quad (5.11)$$

$$c = P_o(M-1) \frac{(M-1)!}{N!} a^N \left(\frac{a}{N} \right)^{M-N-1} e^{N\mu/\gamma} \quad (5.12)$$

where

$$P_o(x) = \left[\sum_{k=0}^N \binom{x}{k} a^k + \sum_{k=N+1}^x \frac{x!}{(x-k)! N! N^{(k-N)}} a^k \right]^{-1}$$

and $a = \gamma/\mu$ or T_f/T_s . After evaluating the waiting time, the probability of j packets drop per talkspurt of the channel allocation queue for a given waiting time threshold τ is expressed as

$$P_{dch}(j) = P(j\tau < w \leq (j+1)\tau). \quad (5.13)$$

Then the average number of packets dropped in every talkspurt is the expectation of P_{dch} from where we can calculate the packet dropping rate.

$$Pda = \Pr\{\text{packet drop}\} = \frac{\tau}{T_t} \sum_{j=1}^{\infty} j \cdot P_{dch}(j) \quad (5.14)$$

By adding the dropping rate of the contention and channel allocation states, the optimal frame structure can be obtain (see Figure 5.1 and 5.2).

§5.2.3 Results based on the model

The combination of R and I slots in a frame at different values of the speech hangover for a packet dropping of 2 percent, $p=0.3$ and 72-slot ATDMA frame is shown in Table 5.1. From the results in the optimisation, it was found that the variation of the load has a more dramatic effect on the channel allocation queue compared to the S-ALOHA contention process. It was also shown that even if the number of slots were increased, the resulting packet dropping rate was just slightly below the one percent mark. This suggests that some improvements in the contention process is required. One way to solve this problem is to impose a *reservation time-out* or equivalently vary the hangover period. By imposing a reservation timeout, the effective access rate will be reduced thereby reducing the packet dropping rate. For example, in the system optimised in Table 5.1 (see below) , 4 R -slots are required to carry the maximum number of users at a 2 percent packet dropping rate in which a hangover of 100 ms is required. If only 140 users are in the system, the minimum dropping rate is only 0.78 percent at a hangover period of 250 ms. This would mean that the user has a time-out of 150 ms from the last hangover frame before it loses its reservation.

Another concern from the results is that the optimal frame structure requires a hangover of 100 ms which corresponds to an average talkspurt duration of 0.9 seconds. At 10 Kbps speech codec, this is equivalent to approx. 1.1 Kbytes of speech. If we accommodate data in the system, most data services have an average message size of less than 1.1 Kbytes. This would mean that the data access rate is faster than that of packet voice. Therefore, the optimisation does not necessarily be for voice service only.

Table 5.1 Optimal tuning of Voice ATDMA

Hangover (ms)	Maximum Users			Number of R-slots		
	<i>Erlang</i>	<i>Engset</i>	<i>Engset*</i>	<i>Erlang</i>	<i>Engset</i>	<i>Engset*</i>
50	155	165	135	6	6	11
75	160	170	160	4	4	5
100	160	170	160	3	3	4
125	160	170	160	3	3	4
150	155	165	160	3	3	3
200	155	160	155	2	3	3
250	150	160	150	2	2	3
375	140	150	140	2	2	3
500	130	140	130	2	2	3

* Calculated using Engset Model including the effect of the transmission efficiency, χ .

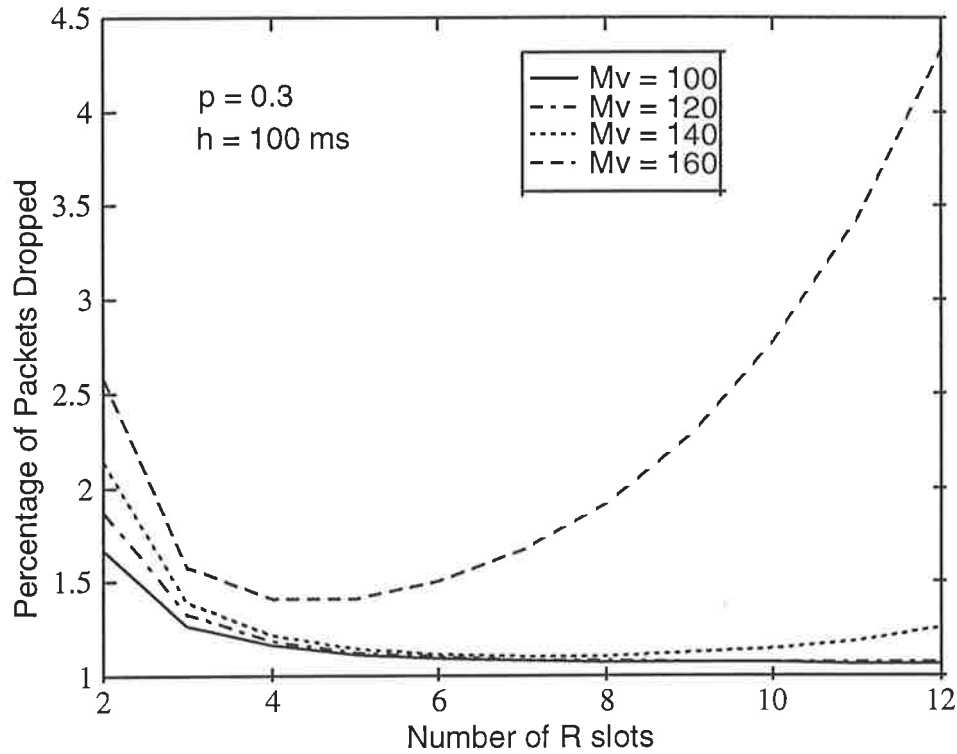


Figure 5.1 Frame Optimisation of ATDMA for Voice Traffic

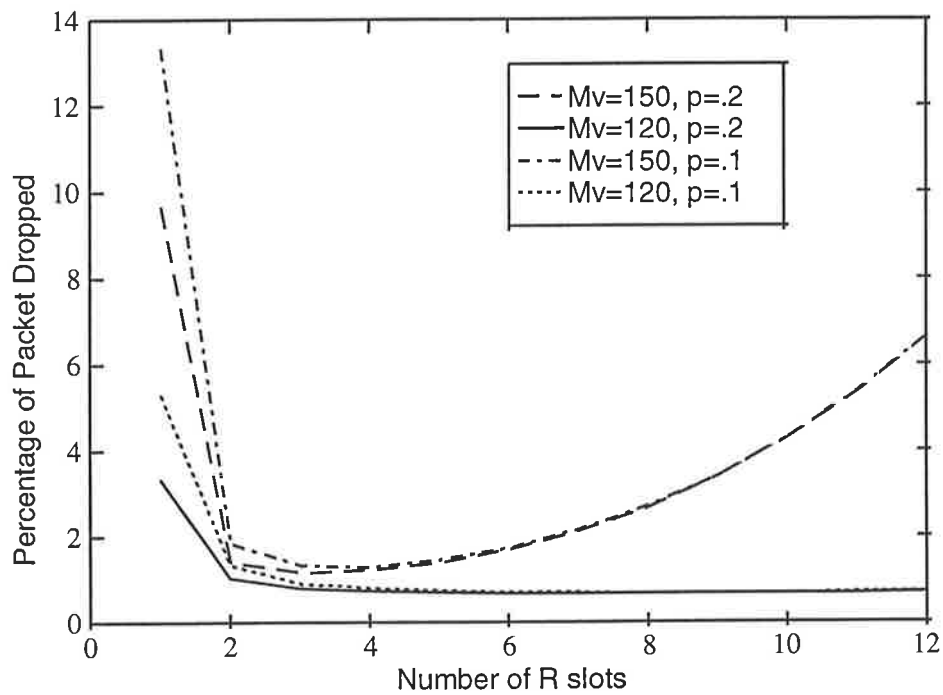


Figure 5.2 ATDMA Performance at Various Retransmission Probability

The figure shows an optimal value of R when the maximum number of users is required. The retransmission probability has little effect on the optimisation. (hangover = 250 ms)

§5.3 Reservation Policy for Data users

The main difference between data from voice is that data users can request and reserved specific amount of resources to schedule its transmission. For a single slot reservation scheme, there are two main reservation policies according slot assignment. One is the use of a channel or slot allocation queue. In this case, the users that have made their request via the request channel are served on a defined service discipline. First Come First Served (FCFS) and the Shortest Message First (SMF) service disciplines are commonly used. Variants of these two disciplines are also possible.

The other reservation policy is the *immediate assignment* describe in §5.1 where no queuing is involved. If a new user finds no available channel, the base station will assign the user to a reserved channel. In the reserved channel, there could be one or more users already reserving so that the new user will compete with the exiting users. This happens when the system is in the near overload or overload region where a form of sub-multiplexing is required. It is also possible that users will be reassigned to a different channel or slot with less load. In this scheme, the transfer delay of a message only consists of a transmission delay (no queuing). Thus a scheduling policy is required.

§5.3.1 Immediate Assignment Allocation Scheme

There is no difference between the different reservation policies when the system is underload (utilisation < 0.6) since users will be served on a FCFS basis. It is when the load is high that the reservation policies will take effect. When looking at the scheduling of multiple users in one particular channel, the fairness criteria must be defined. In this case, fairness would refer to the relationship between the transmission time and the message size. So the objective is to achieve a linear message size versus transmission time characteristics. Both the FCFS and the SMF policies cannot achieve this because in FCFS, users with small messages might wait to users with very long messages while in SMF, the users with very large messages may not be given a time to transmit. To achieve a balance between the two, an *incremental round-robin reservation* is proposed here.

The *incremental round-robin reservation* is a form of a controlled reservation. To start with its description, a channel could have a number of assigned users Mc . At the start of the reservation cycle, each user is allowed to reserved Lc blocks or bursts.

Therefore in each user can transmit L_c blocks or bursts every round. To achieve fairness and conserve the channel, the base station must know the message size of each user at the start of its transmission so that it will have a knowledge of the remaining message length after each transmission (explicit demand). Having this information, L_c is chosen to be the smaller than the smallest message of all the users ($1 < L_c < \min\{L(t)\}$). If $L_c = 1$, the system behaves like a TDMA. However, it is advisable to use $L_c = \min\{L(t)\}$ so that at each round, at least one user leaves the system allowing users with smaller messages to achieve shorter transmission delays. A new user joining the channel will start its first transmission at the end of the current round. Upon its arrival, the L_c may or may not be updated based on the new user's message size.

§5.3.2 *Performance Comparison*

The performance of R-TDMA with data traffic was evaluated using simulations. In an overloaded system, each channel must carry the maximum number of users. An exponential packet size in slot units is used with an average of 100 slots. It is assumed that a new user joining the round robin waits until all the users have transmitted in their turn. This creates a latency since the packet size of the new user could be smaller than the rest of the users assigned to the channel. The scheme is compared with the FCFS service discipline. The results are plotted in Figure 5.3. The figure clearly demonstrates the level of fairness achieved by the incremental round-robin reservation as compared to the FCFS scheme. The linear relationship between the packet size and delay determines the effectiveness of the proposed scheme.

The performance is once again evaluated in the high load region so that the number of users assigned in the channel can vary in time as the load fluctuates. This time, an exponential idle time is used as well as an exponential packet size distribution. To increase the fairness of the reservation, the reservation period L_c , is updated upon the arrival of a new user in order to reduce the latency. Using the same traffic parameters the results are plotted in Figure 5.4. The plot exhibits some scattering of the delay which is due to the variation of the channel load. However, the maximum expected delay is always within the $L \leq D \leq LMc$. The fairness is measured in terms of the average *relative difference* which is the ratio of the difference between actual transmission delay and the expected delay over the expected delay. The expected delay is taken from the measured samples using the first order or linear fit..

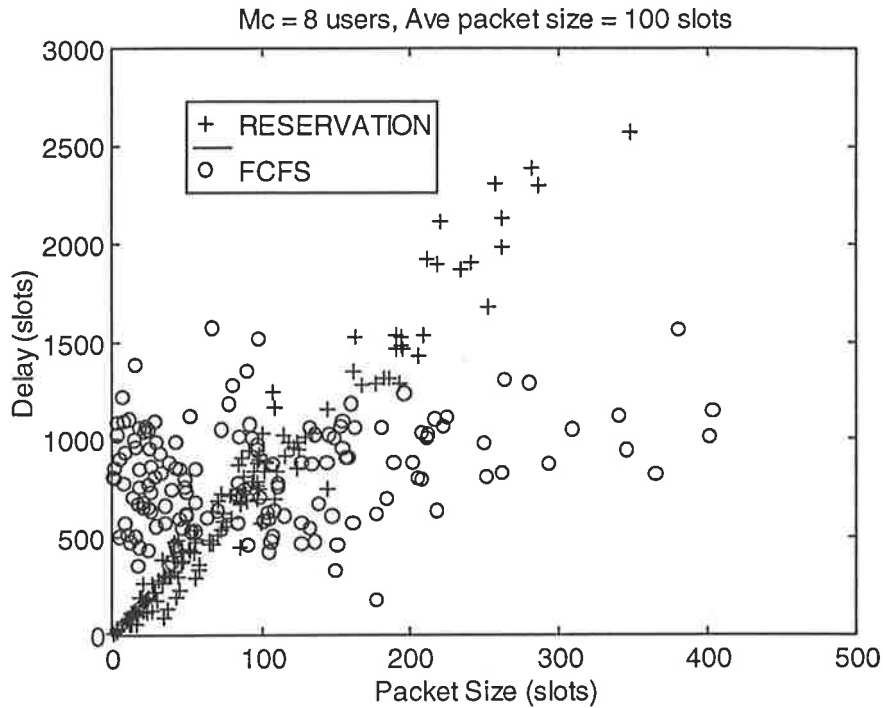


Figure 5.3 Comparison of FCFS and Reservation Schemes at Overload Region

Relative Difference = 0.13 and 3.69 for reservation and FCFS respectively. A fixed of 8 users were assigned on one channel and the average message size was 100 slots. The packets were exponentially distributed. The reservation scheme shows a high degree of fairness.

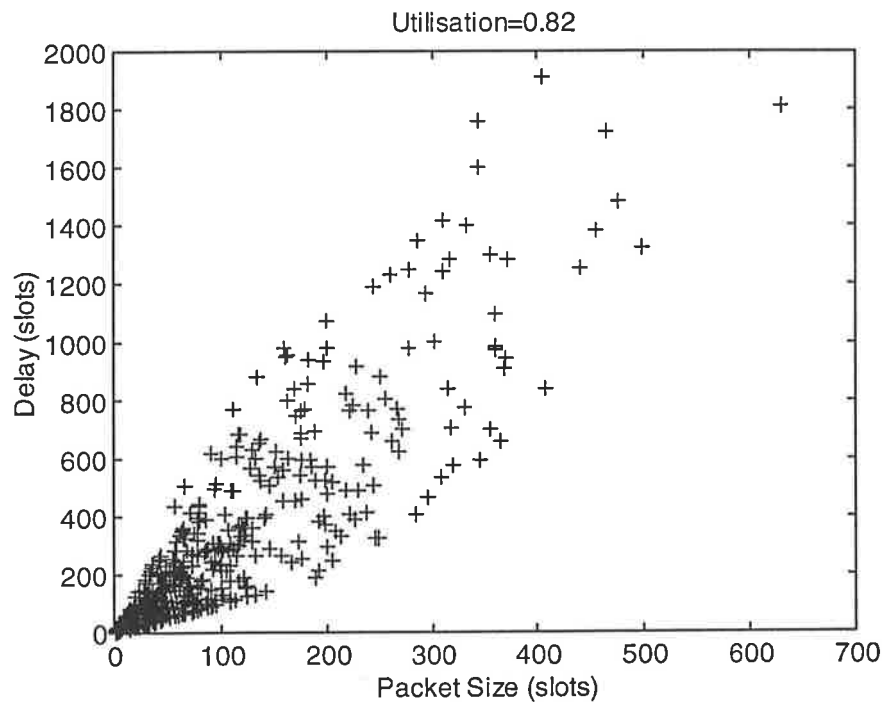


Figure 5.4 Delay Performance and Fairness of Data Scheduling at Higher Load

The simulation used a binary state source model. The reservation parameter was updated upon the arrival of a new user. Measured Relative Difference = 0.4.

§5.4 Enhancements to the R-TDMA protocols

In §5.2 the performance of ATDMA for voice traffic was considered and some optimisation results were obtained. The results were based on very conservative assumptions. This section shows some improvements on the protocol mainly on the access mechanism. Moreover, we consider some improvements to the S-ALOHA within the physical layer like the random access capture, topology and coding.

§5.4.1 Effect of Capture and Forward Error Correction

In the cellular environment, sometimes a receiver can successfully receive a packet if more than one users transmit on a particular slot. This phenomena is referred to as the packet capture. The importance of packet capture is not only for the S-G performance but more importantly in the stability since an increase in the load near the maximum throughput region can significantly stress out the delay rather than the throughput itself. This leads to a significant increase in the number of backlogged users so that a method to increase the capture probability is required. The capture effect is a result of two conditions. One is the near-far condition whereby mobile terminals that are closer to the base station can be received first which then the base station receiver synchronisation can lock-on to the first receiving packet and thus increasing its probability of capture. This also depends on the packet arrival time window in the burst structure of the physical layer. The second condition is due to the instantaneous power level of mobile terminals due to the propagation characteristics (fading). This causes some terminals which can be received with more power level than others and thus increase the capture probability. Therefore, a modification of the original S-G formula is required to include the effect of capture. Then the S-G formula for a capture case is

$$S = \frac{\min \left\{ L N_a \left[e^{-G_a} \sum_{k=1}^{\infty} \frac{G_a^k}{k!} C_p^k \right], (N - N_a) \right\}}{N} \quad (5.15)$$

where : C_p^k is the probability of one packet being captured for k users transmitting. From the Poisson arrival process, the higher values of k is already irrelevant. A very conservative capture model is used in [Ham95] where a capture probability of 1 and 2/3 are used for one and two transmitting users respectively with zeros for all the rest. Here

we use a capture model which is observed from the results in [QiWyr94] given as

$$C_P^k = \begin{cases} 1 & \text{if } k = 1 \\ .6 & \text{if } k = 2 \\ .2 & \text{if } k = 3 \\ 0 & \text{otherwise} \end{cases} \quad (5.16)$$

Then the S-G performance of a capture case are shown Figure B.5 in the Appendix B. When compared to the non-capture case, the improvement over the non-capture case was noticeable. Under this capture model, we assumed that there is only a single receiver in which at most only one packet can be captured in the event that multiple packets are transmitted.

If a capture model is used, it is already assumed that the decreased probability of capture is due to severe bit errors. But sometimes, the bit errors are caused by the channel fading characteristics. The S-ALOHA is known to have decreased throughput with fading channels [Dav80]. Therefore the random access mechanism in ATDMA must consider both the error correcting code and the channel quality. If we have P_e as the probability of bit error, e as the number of correctible errors by the Forward Error Correcting (FEC) code and b number of bits in the random access burst, we have

$$S = \frac{\min \left\{ L N_a \left[G_a e^{-G_a} \sum_{n=0}^e B_{in}(b, n, P_e) \right], (N - N_a) \right\}}{N}. \quad (5.17)$$

If we associate the probability of bit error to the number of colliding users, we need to express the bit error probability as P_e^k . Then the S-G equation becomes

$$S = \frac{\min \left\{ L N_a \left[e^{-G_a} \sum_{k=1}^{\infty} \sum_{n=0}^e B_{in}(b, n, P_e^k) C_P^k \frac{G_a^k}{k!} \right], (N - N_a) \right\}}{N} \quad (5.18)$$

which generalised the performance of ATDMA with all of its sensitive parameters. In this way we can provide a bond between the MAC protocol and the attributes of the channel. The determination of P_e^k is outside the scope of the study of protocols but rather related to the physical layer (i.e. impact of modulation, power control, channel fading, etc.). Also note that we only consider the effect of transmission errors for the random access burst. This is because after the user can successfully contend for reservation, the power control and timing advance can already be activated by both the terminals and base station to mitigate the errors in the information bursts. Therefore there is a need for a better coding scheme in the random access burst. The error control procedure for the information packets also vary in many implementations.

§5.4.2 Capture and Antenna Beam Overlap

It should be noted that the capture effect can be intentionally incorporated in the design. For example, the use of longer burst window and different transmission power can enhance the capture effect. The other way to increase the throughput is to allow simultaneous capture of different packets. Hence we need to employ several independent receivers. The independence of such receivers maybe obtained by diversity technique (eg. space diversity). Our topological model is shown in Figure 5.5. From the topology, we can assume that a certain packet can be received by any of the receivers with equal probability which is defined from the capture model. Then the probability that a packet is received by at least one receiver is defined as

$$\Pr\{\text{success} | k \text{ packets transmitted}\} = 1 - \Pr\{\text{not received by any receiver}\}$$

and the equations are as follows:

$$\begin{aligned} &= 1 - \beta_{in}\left(Nr, 0, \frac{C_p^k}{k}\right) \\ &= 1 - \left[1 - \frac{C_p^k}{k}\right]^{Nr} \end{aligned} \tag{5.19}$$

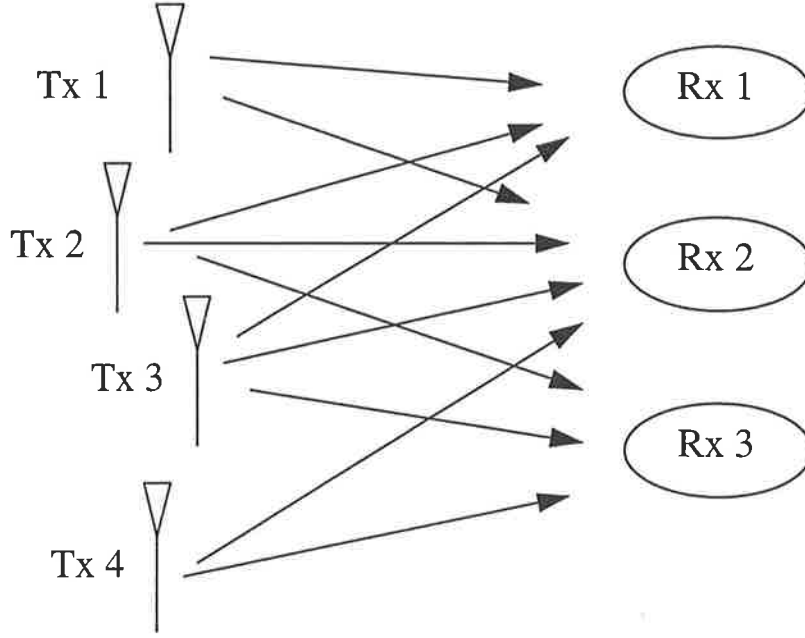


Figure 5.5 Topological Model for Multiple Packet Capture

A number of mobile stations (Tx) are performing random access in a group of receivers (Rx). A mobile station can be heard by a receiver if it is within the coverage. This scenario requires receivers with good capture capability.

Then the throughput of the contention process becomes

$$S_a = \sum_{k=1}^{k_{\max}} \frac{G_a^k}{(k-1)!} e^{-G_a} \left[1 - \left[1 - \frac{C_p^k}{k} \right]^{Nr} \right] \quad (5.20)$$

where $k_{\max}$ is the maximum number of transmission from the capture model. We plot this result using the capture model previously defined and it shows that the performance is far better than the single receiver case in all other configurations in Appendix B.

The overlaps in adjacent sectors and cells in the cellular networks is inevitable. Whilst this will enhance the capacity of the system [Eklun86] [Wat95], it has a negative effect to the performance of the random access mechanism. Load partitioning in ALOHA networks is known to have better performance than the load sharing. However, we will examine the system when capture capability is present. The spatial overlap in wireless networks may be a result of several antennae/receivers using the same frequency whose beams are spatially scattered vertically or horizontally to provide a receiver diversity. To generalised this scenario, it is more convenient to quantify the amount of overlap in terms of a parameter F_s as the ratio of the overlap area to the sum

of the area within the coverage. If G is the total load in the sector, then the load per antenna in a 2-antennae case is

$$G_1 = \frac{G}{2}(1 + Fs) \quad (5.21)$$

and the probability of a successful attempt in the non-overlap area given that k other users transmit is

$$P_{su} = \Pr\{\text{successful attempt}\} = \sum_{k=0}^{k_{\max}} \frac{G_1^k}{k!} e^{-G_1} \left[\frac{C_p^{k+1}}{k+1} \right] \quad (5.22)$$

However, when the user is in the shared area, it competes with all other users in the sector and the probability of success is

$$P_{so} = \Pr\{\text{successful attempt}\} = \sum_{k=0}^{k_{\max}} \frac{G^k}{k!} e^{-G} \left[1 - \left[1 - \frac{C_p^{k+1}}{k+1} \right]^2 \right]. \quad (5.23)$$

Then the total probability of success becomes

$$P_{st} = P_{so}Fs + P_{su}(1 - Fs) \quad (5.24)$$

The results of the 2-antenna case with capture is shown in Figure 5.6. The figure shows that any overlap can degrade the success probability of the S-ALOHA system. By employing two antennae, the performance was improved and a graceful degradation of the success rate is shown.

§5.4.3 R-TDMA with Dynamic Frame Configuration

The main problem with ATDMA is its fixed framing structure which limits the flexibility of the protocol for mixed traffic. The flexibility can also be increased by having an adaptive retransmission probability adjusted to the load. In the case of voice, fixed frame configuration is acceptable because the traffic statistics is uniform for all users. However, optimal throughput cannot be achieved for data users because the traffic statistics vary in time. To address this problem, we introduce an ATDMA with configurable frame structure.

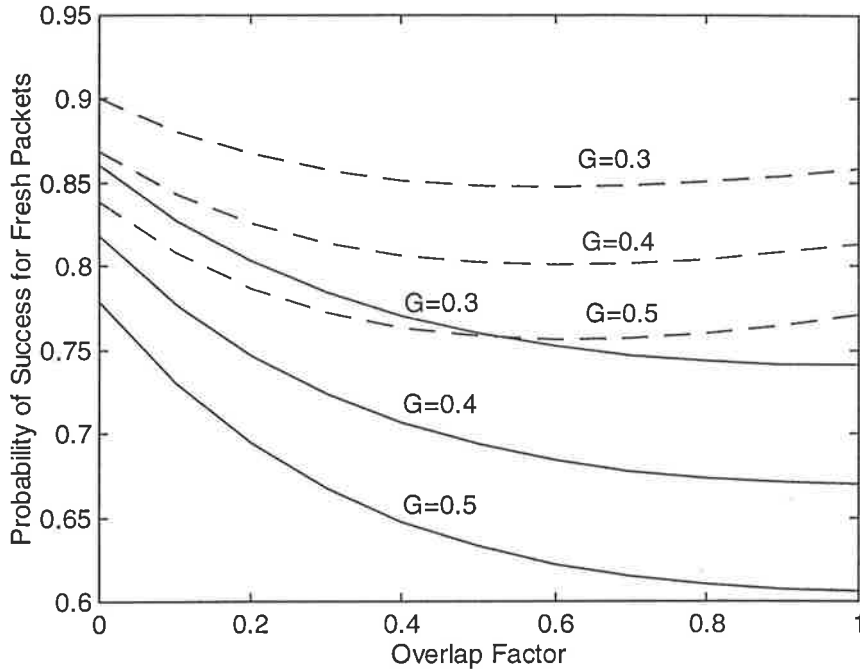


Figure 5.6 Effect of Overlap on the Performance of Random Access Protocols

(The dashed lines represent the overlap performance with capture. It is shown that the success probability has minimum values in the region of 0.5 to 0.6 of overlap factor. The capture model in §5.4.1 is used)

The protocol consists of N slots TDMA frame where a number of slots are allocated for control and the rest for traffic. Initially there are $Na_0 = \{1, 2, 3, \dots\}$ slots in the frame optimised for a 100 percent voice traffic load as in ATDMA. When mixed traffic is supported, Na can be increased $\{Na \geq Na_0\}$ so that the frame structure is optimised while the voice QoS is maintained. The frame optimisation is handled by a QoS control like a Pseudo-Bayesian algorithm for the contention process. The frame structure follows a definite arrangement. For example, in a 32-slot frame, the sequence will be as follows: **TS0, TS16, TS8, TS24, TS4, TS12, TS20, TS28**. If TS0 and TS16 are Na_0 then they are always reserved for control purposes. The uplink control slots are used as access slots and the corresponding downlink is used to acknowledgment. The last slot, TS31 is mainly used for paging the location of Na 's and correspondingly an acknowledgment and other control purposes in the uplink. The control mechanism must be capable of optimising the system performance because any increase in the number of control slots affects both the performance of voice contention and channel allocation as well as that of data delay. Thus the resource allocation has to divide the frame into the control slots and traffic slots for both voice and data while maintaining their respective

QoS. However, the most important problem to be addressed in the case of bursty traffic is the one of stability. Long term variation of the arrival process (i.e. change in proportion of each traffic class) makes fixed frame configuration protocol prone to instability. Even if an adaptive retransmission probability is employed, it is not a guarantee that stability can be achieved.

To model the system, a 16-slot ATDMA is tested with 2 fixed access slots located in TS1 and TS9 using simulations. Two other access slots are used on demand basis located in TS5 and TS13. A two state MMPP data traffic source is used with arrival rates $\lambda_1=1/24$, $\lambda_2=1/90$, sojourn time $t_1=2000$ and $t_2=5000$, and average message size 16 and 64 all in slot units respectively. The random access mechanism is a S-ALOHA with Pseudo Bayesian collision resolution algorithm [see Ref [BerGal92]]. The retransmission probability is broadcasted by the downlink channel periodically set by the broadcast interval parameter. The Pseudo-bayesian estimator takes the average arrival rate for the approximation of the backlogged users, $\tilde{B}$. Together with the broadcast of the retransmission probability are the locations of the access slots. The retransmission probability is selected as

$$\text{Parameters : } \begin{cases} p = 0.3, & Na = Na_o = \{1,9\}, & \lambda = 8\lambda_s & \text{if } \tilde{B} \leq 3 \\ p = 1/\tilde{B}, & Na = \{1,9,5\}, & \lambda = \frac{16}{3}\lambda_s & \text{if } 3 < \tilde{B} \leq 6. \\ p = \max\{0.05, 1/\tilde{B}\}, & Na = \{1,9,5,3\}, & \lambda = 4\lambda_s & \text{if } \tilde{B} > 6 \end{cases} \quad (5.25)$$

where λ_s is the average arrival rate of the MMPP source from Eq. 3.3. and λ is the effective arrival rate used in the Pseudo-bayesian estimator.

The simulation used an exponential messages and the broadcast intervals from 10 frames to 150 frames. The contention delay and waiting delay were measured and are tabulated in Table 5.2. The results clearly showed the effect of the update interval to the contention delay. Consequently, at very long update intervals (i.e. 150 frames), the system becomes unstable. This clearly shows the unsuitability of the fixed frame ATDMA for bursty traffic.

**Table 5.2 Results of R-TDMA/DFC**

broadcast interval (frames)	contention delay (slots)	Waiting delay (slots)
10	67.45	179.98
20	88.86	207.97
30	90.74	198.88
40	125.04	220.53
50	140.70	192.02
75	162.15	190.55
100	235.60	223.80
150	infinite (unstable)	N/A

§5.4.4 Integrated Voice/Data ATDMA Protocol ^[Buot95b]

Data service access is a prime importance that must be considered in the design of future wireless access protocols. The task of providing a packet based access mechanism for voice are due to two reasons: 1) to exploit effectively the user's activity (on-off pattern) of voice traffic and 2) to provide an easy and effective method of integrating mixed services. Traditionally, in the integration of voice and data, there are two implementations that are commonly used, the non-preemptive scheme and the preemptive resume scheme. The non-preemptive scheme is necessary if the data services have strict QoS criteria in terms of delay. However, it is not attractive when the number of channels are few since lower multiplexing gain is expected for voice and data resulting to a poor channel utilisation. Thus the second case is sometimes necessary but the QoS performance for data is not guaranteed (preemption can dramatically increase the transmission delay). Moreover, both strategies for integrating data to voice will in some extent (especially the non-preemptive scheme) affect the performance of the voice service. But the main advantage of having a channel allocation phase prior to the reservation in ATDMA protocol is the ease of implementation of the various techniques to control the QoS in voice/data integration.

By virtue of a reservation protocol, a user that has reserved a channel has the right to own the channel unless its reservation has expired. To avoid confusions we refer the data access as a *temporary reservation* since the voice users can preempt them in case where no spare channels are available. Under this strategy, a data user that is preempted has to go back to the channel allocation state and wait again for the resumption of transmission whenever it is allocated with a channel. The selection of which data user will be preempted in the event that two or more data users are currently transmitting depends on the base station controller. As a good assumption, the data user

whose slot can provide the shortest access delay for the incoming voice user must be preempted.

The other problem in the integration is in the contention process since contention between voice and data will surely affect the performance of voice service. To alleviate the problem, we will use the unreserved slots for the access procedure of data users (data users can wait for a tolerable delay). In this way, the data users are mainly limited to the spare capacity of the voice service. This scheme classifies the slots as either voice slots or data slots (unreserved) where data users are limited only to the data slots for both information transfer and contention. In this way, the access slots are limited for voice users only and thus guarantee a QoS maintenance. This generic protocol that is proposed is based on the assumption that the base station (central control) has full knowledge of all the users and slots status. This procedure is named as *Frame Lookahead Technique*, (FLT) in [Buot95b] because of the base station's ability to employ a *frame look-up* and the users to update their knowledge of the status of the slots from a limited status broadcast. To start with the protocol procedure, let us assume a N -slot ATDMA frame with R access and acknowledgment slots. We have a M_v and M_d voice and data users respectively. M_v is greater than $N-R$ such that the data users are sometimes locked out in the event that the number of active voice users are equal or exceeding N .

In FLT (see Figure 5.7), every data terminal that has packets to send, will listen to all downlink control slots (A slots). If the A slot has nothing to resolve or acknowledge, it will be idle and will be wasted. To utilise it, the base station will broadcast the location of the free slots in the uplink and optionally, the permission probability (pd) for the access contention. This information will be stored by the data terminal as a *lookahead vector*. By identifying a match between the lookahead vector and the slot number of the incoming slot, the data terminal will attempt to access in the incoming slot based on the received permission probability (see Figure 5.8).

If the attempt is successful the data terminal will be allocated by a slot. The slot number will be transmitted by the base station upon acknowledgment. Otherwise the data terminal has to wait for a slot to become available on a first come first serve basis. If the contention is not successful, it will attempt to contend on the next free slot from the lookahead vector. If there is no free slot available, it has to listen to the next lookahead broadcast and repeat the process until a successful contention will occur.

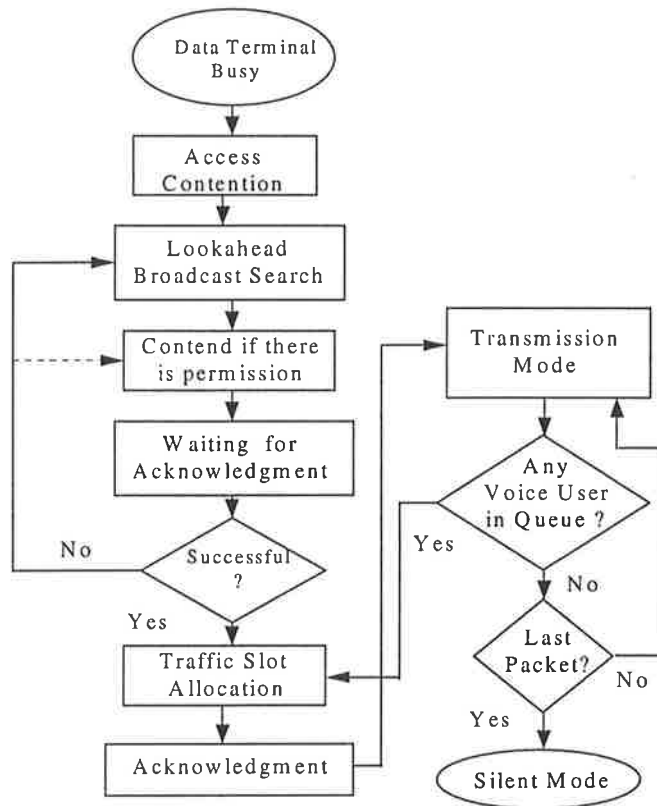


Figure 5.7 Data Access Algorithm using Frame Lookahead Technique

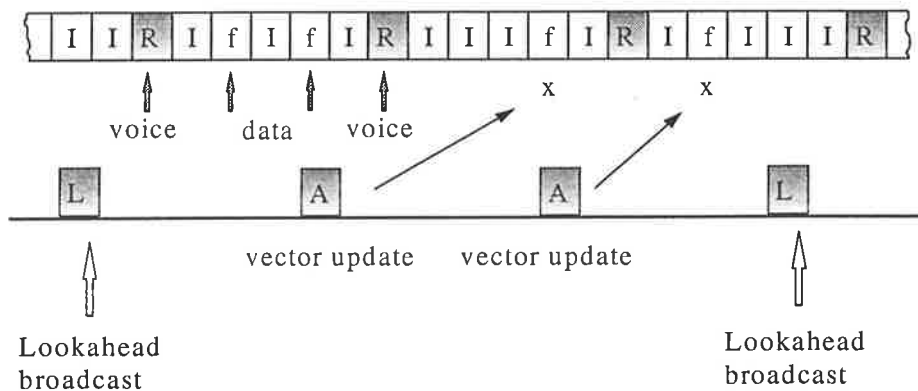


Figure 5.8 Frame Lookahead Technique for Voice/Data ATDMA

It should be noted that the lookahead vector does not necessarily represent all the free slots in the frame. Some slots maybe reserved for incoming voice terminals or being reserved for voice terminals already on queue (during congestion). Also, if the lookahead broadcast interval is long enough, some free slots may not be broadcasted. Another requirement for the protocol to work is that every slot allocation must be realised by all data terminals in order to update their lookahead vector. This is not a problem since the channel allocation and acknowledgment will be performed by the A

slots. Data preemption happens when a voice terminal is waiting for a slot in the queue and the slot number used by a data terminal is nearest to the next A slot. When a data terminal is preempted, it will be held at the head of the data queue. Timeout for data is not implemented. In order to reduce the overhead caused by the access contention, the throughput of the random access (S-ALOHA) must be maximised. Alternatively, a Pseudo-bayesian algorithm for contention resolution can be employed to obtain a good estimate of the number of users in the contention state. Also, fixed retransmission probability is possible at low load region.

The algorithm was evaluated using simulations. A TDMA frame structure similar to RACE 2084 microcellular ATDMA is used. A frame consist of 72 slots with a frame duration of 10ms. The number of access slots vary from 6 to 9. The packet access delay threshold for voice is 10ms which is equivalent to one cycle in the TDMA frame. The voice retransmission probability after a collision was 0.2 and was fixed to all users. From the speech model, the mean listening silence and gaps were 2.22 sec and 0.178 sec with a rate of 0.27 and 0.307 respectively. The mean talkspurt length is 0.5611 sec. This is taken from the 50 ms hangover period. The data packet arrival for each data terminal was assumed to be an independent Poisson process with data length as exponentially distributed. A model of queued users [see Chapter 3] was used in order to have a more realistic load. We assume that the data terminals have an infinite buffering capacity in order to obtain the effect of access delay during congestion. In order to determine the robustness of the protocol, small source worst case traffic values are used (0.05-0.1) equivalent to a maximum of $M_d = 200$ data terminals for $M_v=140$ voice terminals multiplexed in the 72 slots TDMA frame.

For the fixed retransmission probability, the result is plotted in Figure 5.9. It clearly shows that system is greatly influenced by the S-ALOHA contention process. This is shown by the bistable behaviour of the system. For lower system load, the system has very low access delay. The drop of the throughput results when the contention process is overloaded. For the Pseudo-bayesian contention resolution, the system is very stable with an increasing throughput with loading. The results (see Figure 5.10) shows that throughput of 0.8 is attainable for an access rate of 0.5 access/second per terminal at delay of less than a second. Higher values of throughput is expected at lower access rates and higher source traffic. The throughput values were calculated based on the data slots only.

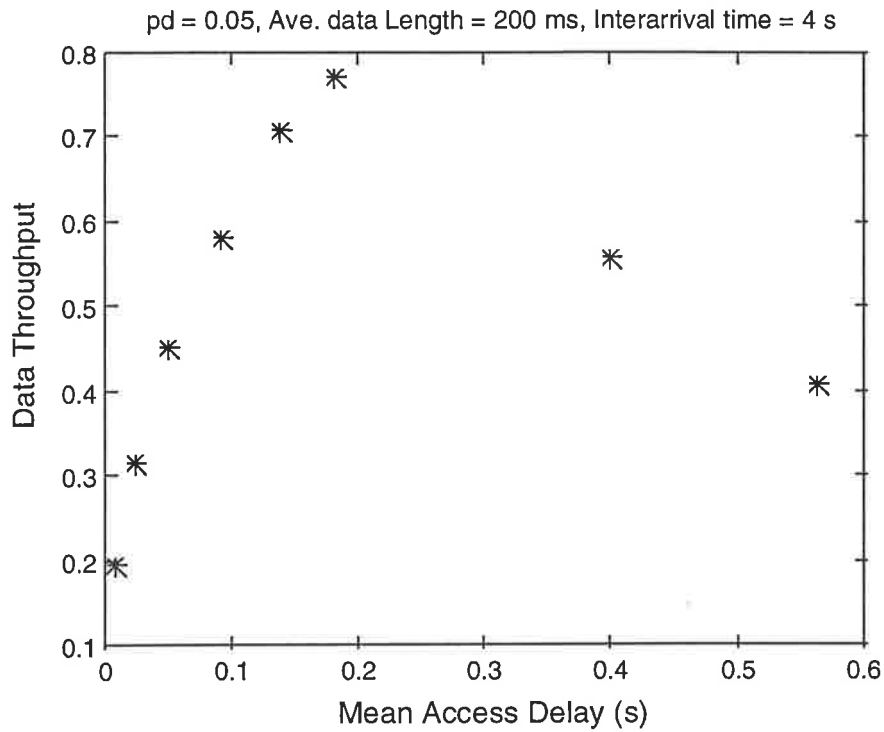


Figure 5.9 Mean Throughput/Delay for fixed retransmission probability

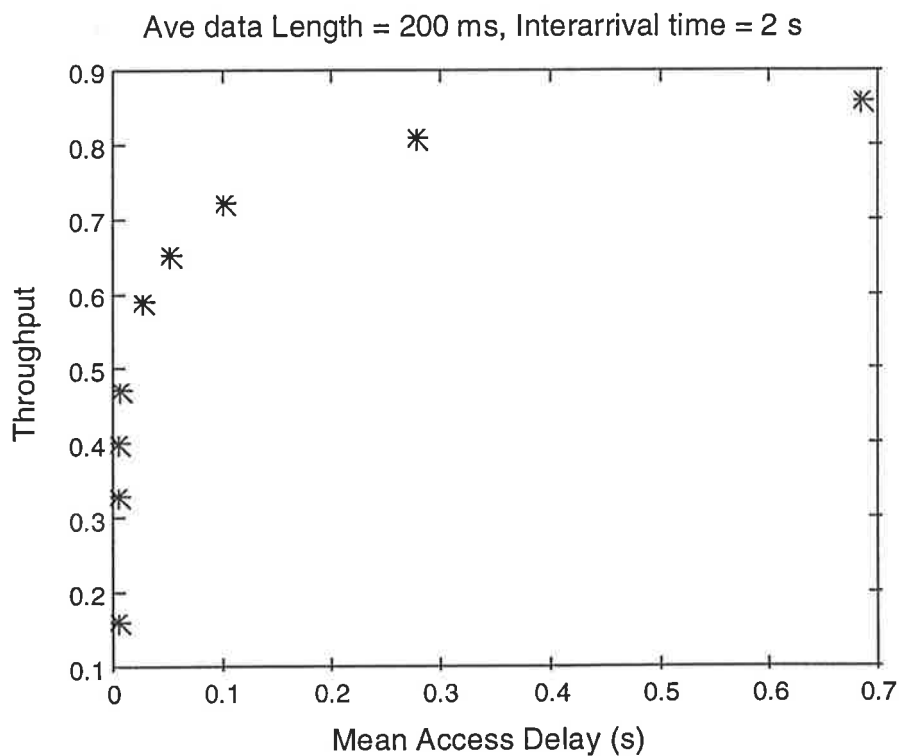


Figure 5.10 Throughput/Delay using Pseudo-Bayesian Algorithm

The pseudo-bayesian algorithm in the contention process achieves better performance compared to the fixed retransmission probability

§5.5 Multipriority Channel Access

Providing a prioritisation for reservation protocols can be employed during the channel request or during the channel allocation phase. However, for two-stage reservation protocols (e.g. PRMA), prioritisation can only be done in the channel access phase. This motivates to provide a prioritised contention for random access protocols. Accordingly, there are existing random access algorithms with priority (see Ref [ChU95][LiP93][Pkaz89]) but are not necessarily suitable in the WPC environment. Stack algorithm is known to be suitable for mobile channels [NiVP95]. Here, a multipriority stack algorithm which provides a good rejection to lower class users at all region of the cumulative delay distribution is proposed. In [Wu94b], a prioritisation scheme for S-ALOHA was achieved by using different retransmission probabilities for each traffic class (e.g. higher retransmission probability for voice over data). However, random retransmission is not effective in providing prioritisation. This argument is supported based on the following discussion.

Consider a S-ALOHA system with two classes of users with a population of M_1 and M_2 with their corresponding retransmission probabilities, p_1 and p_2 respectively. If S_1 and S_2 are the corresponding throughputs, we have G_1 and G_2 as their respective channel loads. From [Kli75] we have

$$G_1 = \frac{S_1}{\left(1 - \frac{G_1}{M_1}\right)^{M_1-1} \left(1 - \frac{G_2}{M_2}\right)^{M_2}} \quad (5.26)$$

$$G_2 = \frac{S_2}{\left(1 - \frac{G_1}{M_1}\right)^{M_1} \left(1 - \frac{G_2}{M_2}\right)^{M_2-1}} \quad (5.27)$$

The two classes experience the same probability of success at their first attempt if both M_1 and M_2 are large which means the first attempt is dependent directly to the total channel load. This behaviour cannot be avoided in random access protocols unless sufficient feedback regarding the status of the channel is provided. Regardless of whatever retransmission probability it may be, the channel load always approach to G if the required throughput is S . The only effect of lower value of p_2 is to delay the retransmission of low priority users to give way for higher priority users. However, when p_2 is small, it is likely that the class 2 users will maintain a large number of

backlogged users along with G_2 . Eventually lower values of p_2 will mean longer delays after a collision. Although this scheme may not be effective in providing prioritisation, the stability of the system can be increased by lowering the retransmission probability of the lower class users.

By extending the approximations in §4.3 to a two class system of infinite users with given arrival rates per class, the total load G was calculated based on the total throughput and the delay distribution was calculated by using p_1 or p_2 instead of p in Eq. 4.9b. The results were then plotted in Figure 5.11 to 5.13. The accuracy of the approximation was sustained by a simulation for the delay distribution of each class. Then effect of varying p_2 is noted showing that the use of smaller p_2 only increases the average delay of class 2 users. More importantly at low load region where higher retransmission probability is preferable due to the less probability of three or more users colliding, the use of smaller p_2 does not necessarily improve the system performance at all. Thus, in the next subsection a multiclass contention scheme based on the stack algorithm which can provided better prioritisation was proposed.

§5.5.1 *Stack Algorithm*

Most TDMA frame structures are partitioned type where special slots (access slots) are allocated solely for random access. In this case, the access slots are staggered in the frame to provided sufficient time interval between reservation attempts. This can be implemented by providing a timing advance for the uplink channels greater than the sum of the round trip propagation delay and processing time. This configuration allows a guaranteed feedback before the next reservation attempt arrives. For power saving requirement, mobile terminals does not monitor previous feedback information (Collision/No-Collision) prior to the arrival of its information packet. This is also essential to gain stability and prevent deadlock condition if any of the users will receive feedback errors. It should be noted that feedback errors or lost feedback will always occur in the random access procedure due to the propagation characteristics and receiver losing synchronisation. Lastly, the binary or ternary feedback is easy to implement as they don't require long information field in the reservation packet. It should be noted that the information field in the reservation packet is very short since a large portion of the access burst is allocated for timing advance and training sequence (i.e. GSM burst structure). In this case stack algorithm fits into the scene.

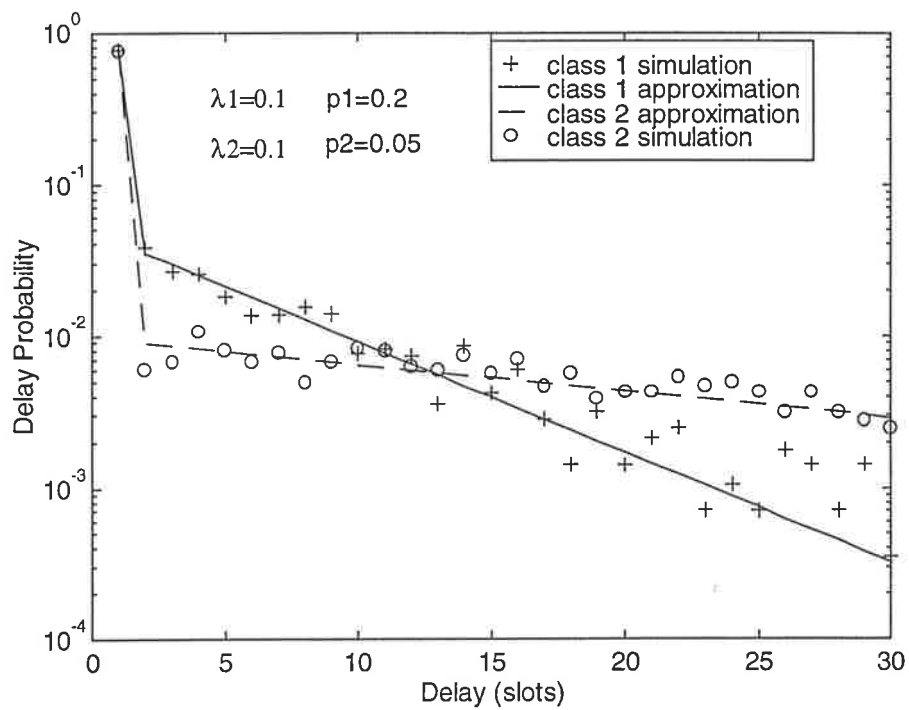


Figure 5.11 S-ALOHA with two classes of different retransmission probabilities

(Note: The model for Slotted Random Access in §4.3 is suitable for the delay calculation of S-ALOHA with two classes of users)

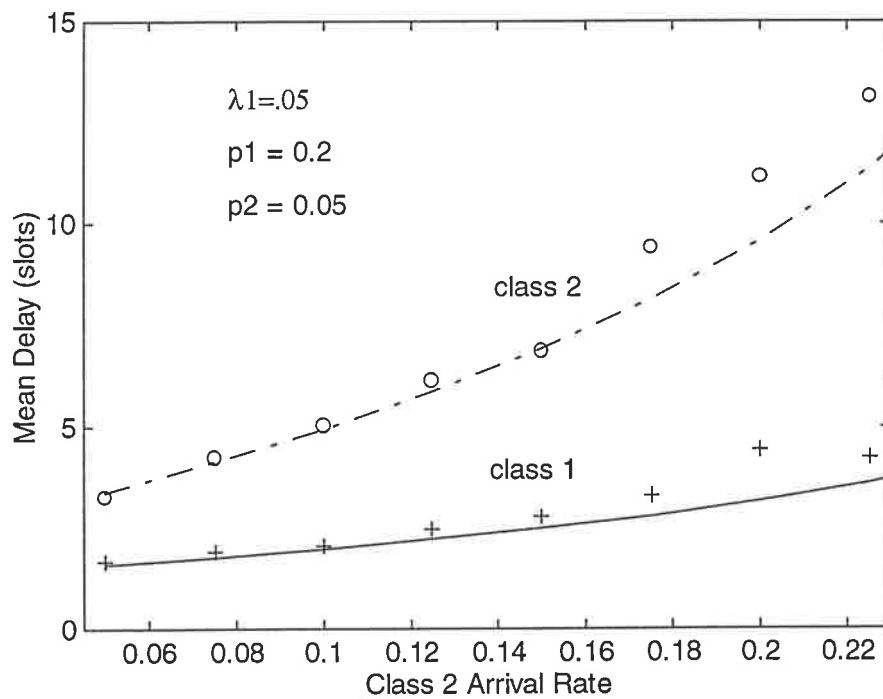


Figure 5.12 Two-class S-ALOHA with Fixed Class 1 Load

The ticks are the simulation results. Discrepancies at higher loads are due to the bi-stable characteristics of S-ALOHA as it flip-flop into two stable regions

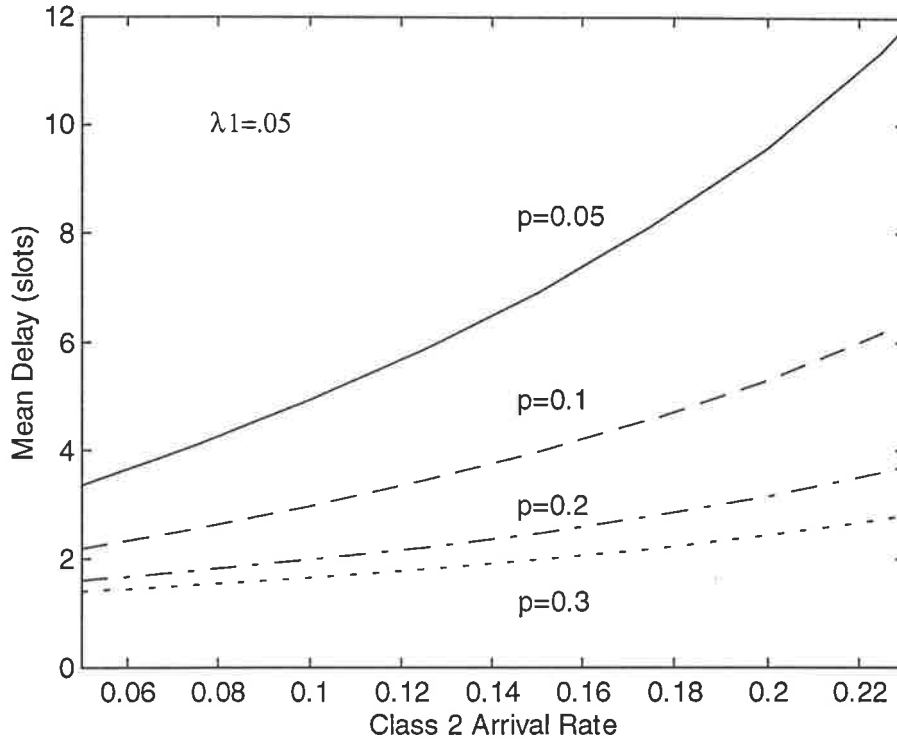


Figure 5.13 Effect of Different Retransmission Probability on Class 2

The plots are taken only at low delay equilibrium point. Note that the use of small values of p_2 only increase the delay of class 2 instead of maintaining a good rejection of class two users and give favour to class 1 users.

The analysis and description of Stack Algorithm is accurately demonstrated in [Tysb85] for the non-blocked type and in [Cap79] for the blocked type or *tree algorithm*. The main difference between stack and ALOHA type algorithms is the use of manipulating counters for each user in order to process the feedback information (collision or no-collision). The counters are updated in every feedback information received as broadcasted by the central control (base station) in every acknowledgment slots (A slots). For a generalised nonblocked stack algorithm, the counters are updated as follows:

- 1] Upon the arrival of a packet, the counter is set to a certain level $c_{i,0}$. A user transmits its packet on the next slot if $c_i=1$.
- 2] If the level at the previous slot is $c_{i-1}=1$ and the current feedback is *NC* then the packet leaves the system.
- 3] If the level at the previous slot is $c_{i-1}=1$ and the current feedback is *C* then the level goes to $c=m$ with probability P_m where $m \in \{1, 2, \dots, m_{max}\}$.

- 4] If the level at the previous slot is $c_{t-1} > 1$ and the current feedback is NC then the level is decremented by 1.
- 5] If the level at the previous slot is $c_{t-1} > 1$ and the current feedback is C the level is incremented by a $m_{max} - 1$.

§5.5.2 Approach to Stack Prioritisation

From the procedure in §5.3.1, it is described that the stack algorithm has a Last Come First Served service discipline. Although the contention process must be lightly loaded in a Reservation protocol, when traffic is bursty the access mechanism could experience congestion or loads close to the maximum limits. In this case, the stack could experience some periods where the effective service rate is lower than the arrival rate. In this case, users at the back of the stack will experience very long service delay. These instants will cause performance degradation in terms of cumulative delay. Thus our main objective in the design of a multiclass stack algorithm is to avoid the worse condition where high priority users are *tagged* at the back of the stack and experience intolerable delays. In the single class stack algorithm the most important parameter of the algorithm is the splitting parameter P_m in the event that there is a collision and the stack level is 1. Heuristically, we can develop a multiclass algorithm exploiting the splitting parameter. But first, we have to analyse the single class stack algorithm and see the possibility of developing a multiclass algorithm while maintaining good performance. First, we consider a stack algorithm with only one class of users (single priority).

At low load conditions, the stack is empty most of the time as the users transmit their packets with high probability of success. Assuming a Poisson source with an aggregate load of λ packets/slot. Then the probability of j packets that arrive in a particular slot is

$$P_{SS}(j, \lambda) = \frac{\lambda^j}{j!} \exp^{-\lambda} \quad (5.28)$$

If the λ is below 0.30, then the probability of j arrivals is negligible for $j \geq 2$. If a collision occur in any of the slots given that the stack is empty is most likely caused by two packets. The number of packets involved in the first collision following an idle or success when the stack is empty is called the *multiplicity*, z . At low load region,

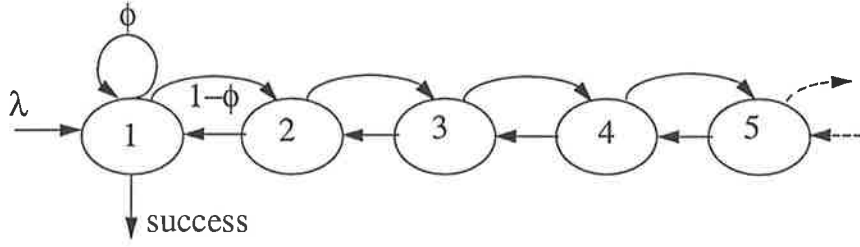


Figure 5.14 Transition Diagram of a Stack Algorithm with $m=1$

$\Pr\{z=2 \mid z \geq 2\} \cong 0.9$. By limiting the multiplicity to $z=2$, the analysis of the stack can be simplified.

Like in many literature [Tsyb85][Vved94] we use here the notion of a *session*. A session is period in between two moments of an empty stack [Tsyb85]. Thus in this case, all sessions start with multiplicity of two. We then need to calculate the transmission probabilities at the beginning of the session which are either success, idle or collision. Using ϕ as the splitting parameter right after the collision (see Figure 5.14), we have

$$P_{succ} = \beta_{in}(2,0,\phi) P_{ss}(1,\lambda) + \beta_{in}(2,1,\phi) P_{ss}(0,\lambda) \quad (5.29)$$

$$P_{idle} = P_{ss}(0,\lambda) \beta_{in}(2,0,\phi) \quad (5.30)$$

$$P_{fail} = P_{ss}(j > 1, \lambda) + P_{ss}(1, \lambda) \beta_{in}(2,1,\phi) + \beta_{in}(2,2,\phi) \quad (5.31)$$

This approximation of the probability of success holds through all instants during the session since the backlogged packets at the back of the stack has no influence on the transmission probability of the packets at the head of the stack. The previous equations can also approximate the maximum stable load with the different values of ϕ . The maximum load is λ_{max} which occurs at $\max\{\lambda\} < P_{succ}$. The value of ϕ must be 0.5 which is the best splitting for two users. Unless the probability of three users colliding is high then the probability of success decreases rapidly in some region with high ϕ . At very low values of ϕ , the throughput will be attributed mainly by the fresh packets and similarly, both extreme values of ϕ decreases the throughput. Values of ϕ above 0.5 are applicable only when capture mechanism is considered [Vved94].

The use of $\phi = 0.5$ at low load region is appropriate due to two reason. Firstly, the multiplicity is often two and secondly, the variation of ϕ at lower loads has little effect on the performance of a stack algorithm. However, at higher loads, this parameter must be tuned to achieve higher throughput and good delay performance. Extending the approximation described earlier, the probability of success right after a collision is the sum of the probabilities at all multiplicities multiplied by their corresponding weights as:

$$P_{succ} = \sum_{z \geq 2} [B_{in}(z, 0, \phi)P_{ss}(1, \lambda) + B_{in}(z, 1, \phi)P_{ss}(0, \lambda)] \hat{P}_{ss}(z, \lambda) \quad (5.32)$$

where

$$\hat{P}_{ss}(z, \lambda) = \frac{P_{ss}(z, \lambda)}{\sum_{j \geq 2} P_{ss}(j, \lambda)}.$$

From the above equation, it is suggested that ϕ is less than 0.5 to maximise the probability of success. This agreed with the result in [Tsyb85] in which a maximum throughput of approximately 0.4 is achieved at values less than 0.5. If we consider the impact of very low values of ϕ , it will also result to successive collisions since the splitting is not optimal. However, the backlogged packets after the collision in the case of multiplicity greater than two can be further divided by having different increments right after the first collision. In fact a throughput of 0.4 can only be achieved with more than one increment value, say $m = \{2, 3, 4, \dots\}$. Thus we can propose an algorithm, where the use of many initial levels right after the collision for different user classes will increase the performance of the stack algorithm as it further splits the collided users according to their class (see Figure 5.15). The previous equation also proved that the throughput of the stack algorithm does not vanish to zero even if the load is greater than the maximum throughput. Accordingly, the suggested value of ϕ can be found by differentiating P_{succ} (see Appendix E) resulting to

$$\phi = \frac{V}{1 + V} \quad (5.33)$$

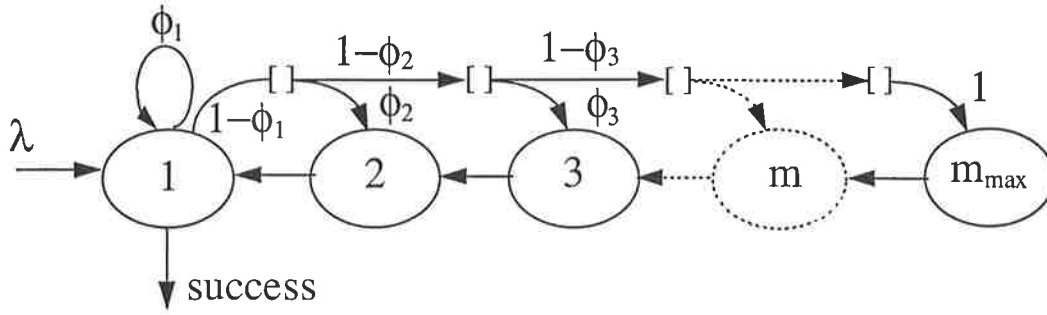


Figure 5.15 Multiple Splitting Model for Stack Algorithm

where

$$V = \frac{(1-\lambda) \sum_{z \geq 2} z \hat{P}_{ss}(z, \lambda)}{\sum_{z \geq 2} z(z-1) \hat{P}_{ss}(z, \lambda)}. \quad (5.34)$$

The result of Eq. 5.32 is plotted in Figure 5.16 showing that more initial levels are required in order to mitigate the problem of successive collision at higher loads as the optimal splitting parameter is less than or equal to 0.5 at any load. The other important parameters in the stack algorithm is the value of the successive splitting parameters in the case where $m > 2$. This is suggested if the value of ϕ is less than 0.5 since we need to further split the users that does not retransmit immediately. In this case lower priority users can back-off faster by using different splitting parameters. Our model is based on Figure 5.15. For a 3 initial-level stack, we have

$$\phi_1 = \frac{V}{1+V} \quad (5.35)$$

$$\phi_2 = \frac{\phi_1}{1-\phi_1} \quad (5.36)$$

and

$$\phi_3 = 1 \quad (5.37)$$

The recommended values for the splitting parameters are plotted in Figure 5.17. The results demonstrated the existence of an optimal splitting parameter for the different arrival rates.

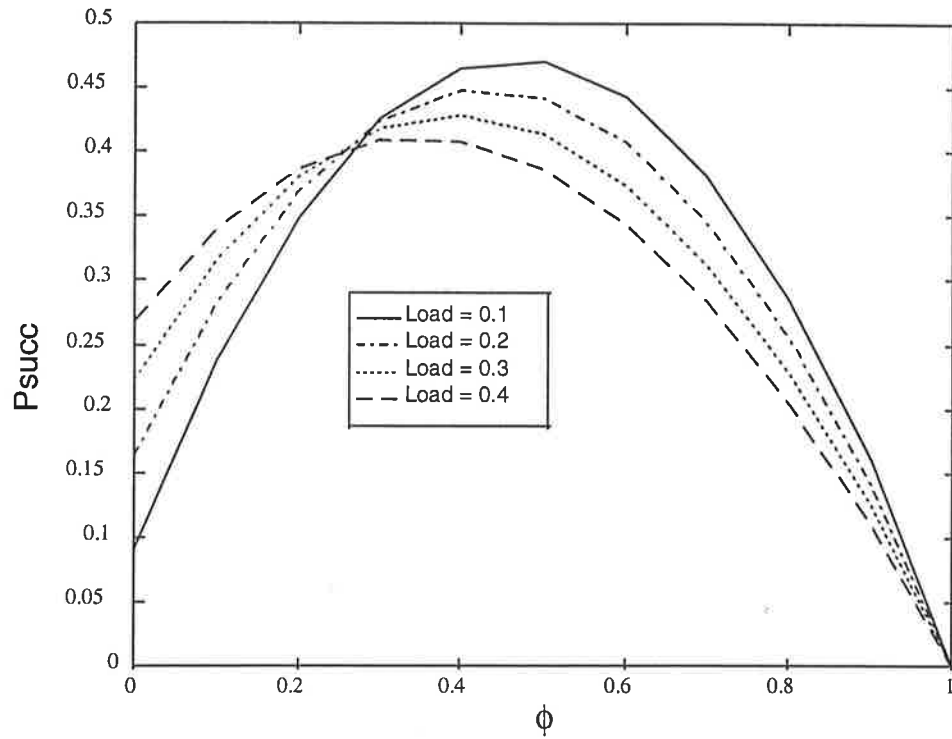


Figure 5.16 Plot of the Probability of Success right after a Collision

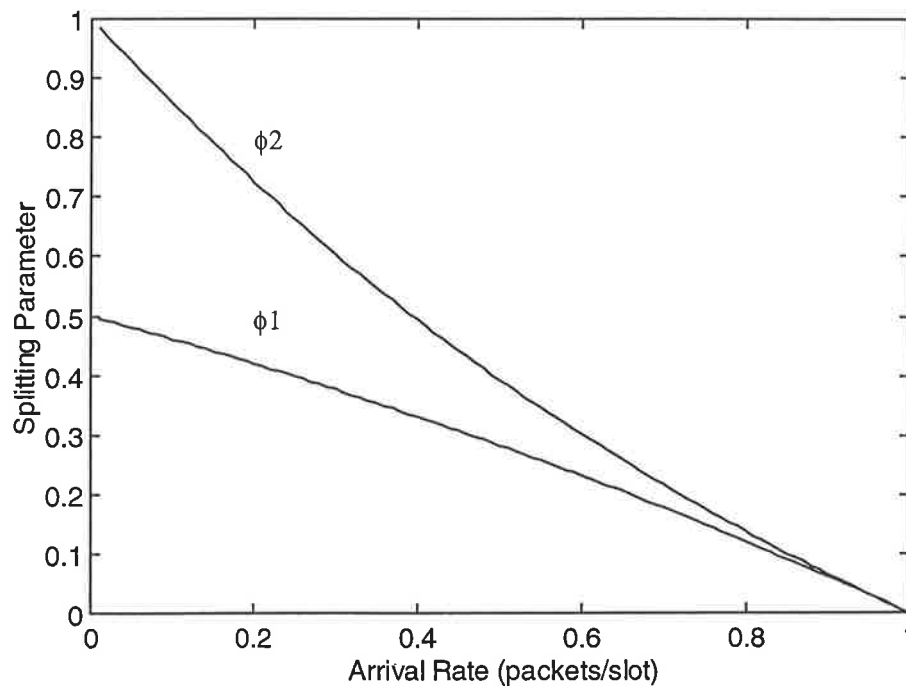


Figure 5.17 Recommended ϕ_1, ϕ_2 against λ for Optimal Splitting

§5.5.3 Multipriority Stack Algorithm ^[Buot96a]

Based on the discussions in the preceding section, we can develop a multiclass stack algorithm where the initial counter values and counter increment/decrement will depend on the class of the user. The idea here is to let the lower priority users to back-off faster in order to give way to higher priority users. The algorithm is as follows:

- 1] Upon the arrival of a packet the counter is set to $c_k=1$ where k is the user class. Thus an arriving packet transmit immediately on the incoming slot regardless of the current feedback (*nonblocked stack algorithm or immediate first transmission*).
- 2] If $c_{k,t-1}=1$ and the current feedback $f_t=NC$ (*No-Collision*), the user leaves the system.
- 3] If $c_{k,t-1}=1$; the current feedback $f_t=C$ (*Collision*) ; $k = 1$; the counter is updated as

$$C_{k,t} = \begin{cases} 1 & w.p. \quad \phi \\ 2 & w.p. \quad 1-\phi \end{cases} \quad (5.38)$$

- 4] If $c_{k,t-1}=1$; the current feedback $f_t=C$; $k > 1$; then the counter is updated as

$$C_{k,t} = \begin{cases} 1 & w.p. \quad \alpha \\ 2k-1 & w.p. \quad (1-\alpha)\phi \\ 2k & w.p. \quad (1-\alpha)(1-\phi) \end{cases} \quad (5.39)$$

- 5] If $c_{k,t-1} > 1$; the current feedback $f_t=NC$; then $c_{k,t} = c_{k,t-1} - 1$.
- 6] If $c_{k,t-1} > 1$; the current feedback $f_t=C$; then $c_{k,t} = c_{k,t-1} + k$.

An example of a collision resolution of a two-class system is shown in Figure 5.18. Each user is denoted by u_k^i for the i^{th} user of class k . The illustration describe the little effect of class 2 users to class 1 users under a period of many successive collisions as the class 2 users back-off faster than the class 1 users. This algorithm is similar to the two-class algorithm devised by [Stav91]. However, we use a different increment value and multiple initial levels dependent on the class of the user in order to avoid the problem where high priority users are **tagged** at the back of the stack during the periods of congestion. This proposed algorithm can achieve stable throughputs for high priority users even if the average service rate is lower than the total arrival rate. Therefore different stable operating points for each class will result. The effectiveness of the algorithm is shown by the rejection of the lower priority users at all points of the cumulative delay (see Figure 5.19). More results are shown in the Appendix F.

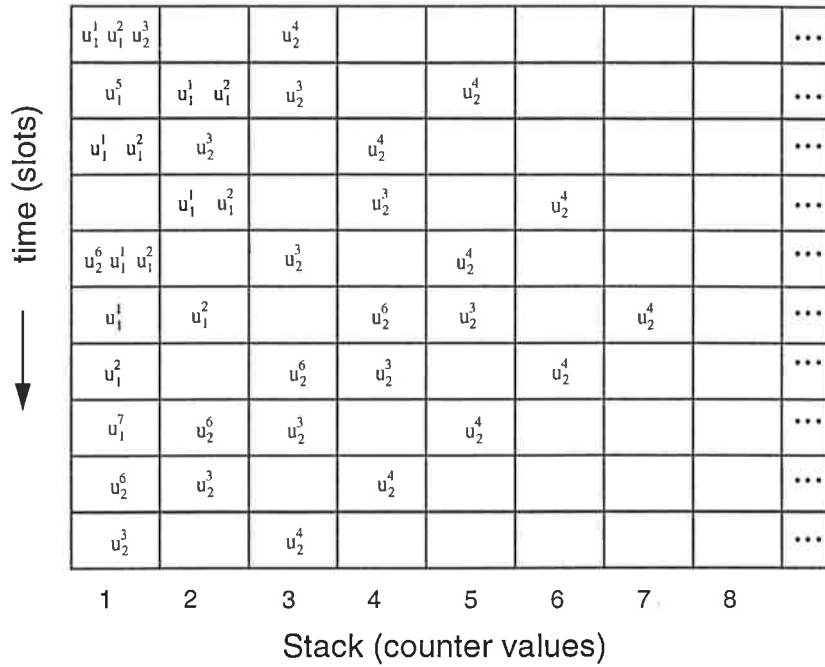


Figure 5.18 Example of a Two-class Collision Resolution Procedure

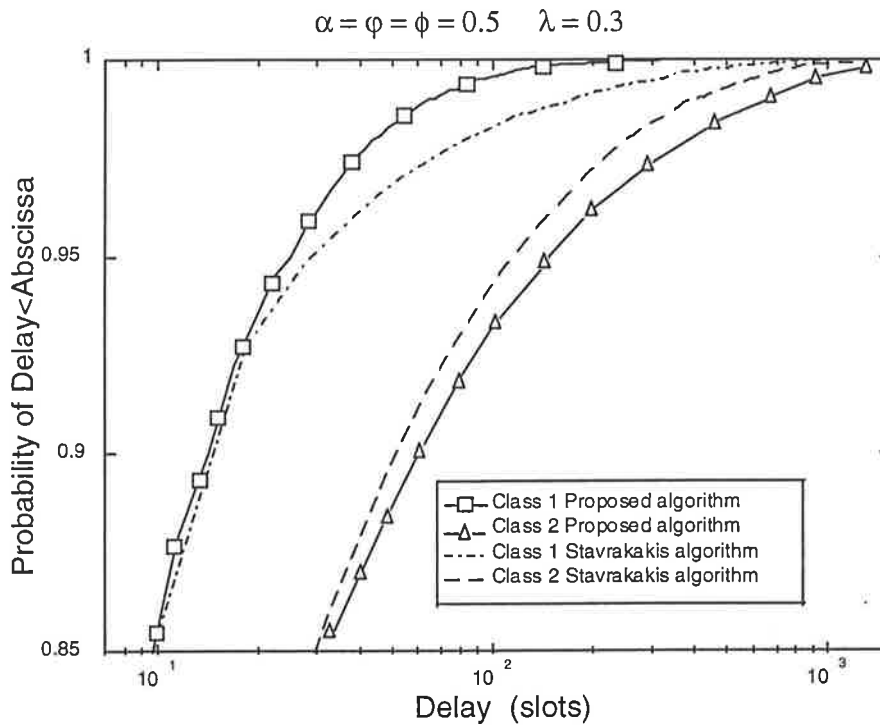


Figure 5.19 Performance of a Two-class Stack

The rejection of class 2 users are shown by the large difference in the cumulative delay at all region while the Stavrakakis algorithm did not provide good rejection at higher delay percentile. The parameters used were not optimal but can be determined based on Eq. 5.35 and Eq. 5.36.

§5.6 ATDMA with Stack Algorithm

To provide a higher stability for reservation protocols, a more stable collision resolution algorithms (CRA) is required. While this area is well understood, there are algorithms that can provide more stable S-G performance than the *Fixed Retransmission Probability S-ALOHA* like the *Pseudo-bayesian Algorithm*, *Binary Exponential Back-off Algorithm* and the *Stack/Tree Algorithm*. The *Pseudo-bayesian* algorithm is good if the user are monitoring the system most of the time so that they can continuously update the estimate of the best retransmission probability while the *Binary Exponential Back-off Algorithm* has been famous for its ease of implementation because the users need only to record its own retransmissions. However, the aforementioned CRAs are not so attractive due to their large delay variance at higher load at the same time, stability is not guaranteed. Here, we introduce another version of R-TDMA which is an ATDMA system with a stack collision resolution algorithm.

§5.6.1 Analysis of ATDMA with Stack CRA using TFA

The main objective in this section is to evaluate the performance of a reservation protocol with a Stack CRA. A TFA analysis of a ATDMA/TREE system as a tool for system performance was then developed. In TFA, both S-G and throughput/delay performance can be obtained but the analysis is not so trivial compared to the original ATDMA because of the stack. We use here a blocked stack⁺ CRA or sometimes called Tree Algorithm [Cap79]. In this protocol, the behaviour is the same as that of the original ATDMA except that in the event of a collision in the channel access, the CRA used is as follows:

1. Upon arrival, a user must check if a collision resolution is currently in progress or if all backlogged users have finished transmitting from the last session. An alternative implementation is for the base station to broadcast an idle signal in the paired *Ack* slots.

⁺ Blocked stack is sometimes called a tree algorithm in which a DFT scheme is used. The DFT scheme becomes effective since the terminal can determine the status of the channel from the feedback channel. A terminal can only transmit once an idle channel is identified i.e. the broadcast of two successive idle slots or through the status flag appended in a downlink packet.

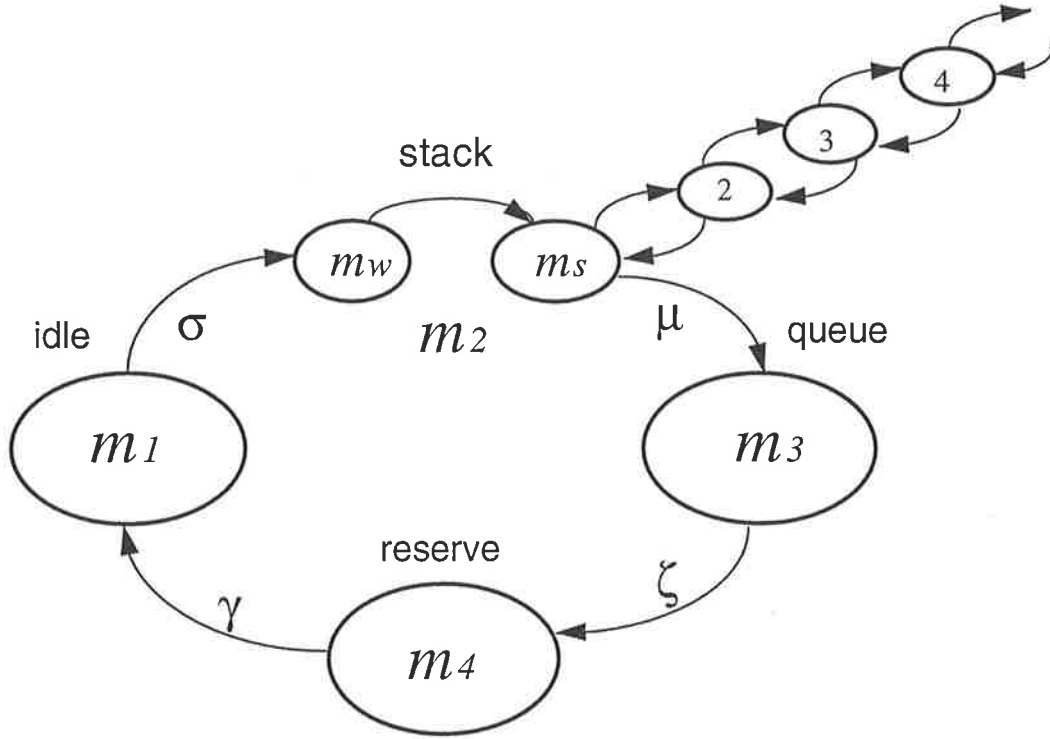


Figure 5.20 Transition Diagram of Reservation Protocol with Tree/Stack CRA

2. In the event of collision, a user retransmits with a probability ϕ or it will opt to wait at stack level 2.
3. Users in the stack has to monitor the feedback in the paired *Ack* slots of all preceding contentions. A success means a decrement in the stack by 1 while an increment of 1 for collisions.
4. Users with stack level 1 transmit on the next contention slots. All the rest are backlogged and listen to the feedback channel.

The state transition diagram of the system is shown in Figure 5.20 where m_1 , m_2 , m_3 and m_4 are the number of users in the idle, contention, queue and reserve states. Then transition rate matrix is constructed as

$$P(\xi(k)) = \begin{bmatrix} (1-\sigma)m_1 & m_1\sigma & 0 & 0 \\ 0 & -\mu(m_2, m_1)^{k-1} & \mu(m_2, m_1)^{k-1} & 0 \\ 0 & 0 & -\zeta(m_3, m_4)^{k-1} & \zeta(m_3, m_4)^{k-1} \\ m_4\gamma & 0 & 0 & (1-\gamma)m_4 \end{bmatrix}. \quad (5.40)$$

Since the values of σ , γ and ζ are already determined in the solution for ATDMA in §4.5.4, we have to solve the service rate of the stack first. The solution for the stack throughput presented here is due to J. Massey and also found in [RomSi90]. To define the service rate of a stack, we have to determine the length of the collision resolution interval (CRI) when a given number of users collide on the start session. If B_n is the length of the CRI in slots with n users are involved in the start of the collision, the service rate μ , is defined as $n/(1+B_n)$ where $B_n=1$ if $n=\{0,1\}$. Thus our problem is when the CRI involves more than 2 users. It should be noted that a Tree or Stack Algorithm uses the splitting technique to resolve the collisions. Therefore users involved in a collision will be divided into retransmitting users and users that are pushed into the stack. Let $Q_i(n)$ as the probability of exactly i users retransmit when n users collide, $Q_i(n)$ is binomially distributed expressed as

$$Q_i(n) = \binom{n}{i} p^i (1-p)^{n-i} \quad (5.41)$$

Therefore, when i users retransmit, $(n-i)$ users are pushed back to the stack. Thus the conditional probability of the length of the CRI given i users retransmit $B_{n/i} = 1 + B_i + B_{n-i}$ for all $n \geq 2$. From [RomSi90], B_n is calculated recursively as

$$B_n = \frac{1 + \sum_{i=0}^{n-1} [Q_i(n) + Q_{n-i}(n)] B_i}{1 - Q_0(n) - Q_n(n)} \quad (5.42)$$

where $B_0 = B_1 = 1$.

From the transition diagram in Figure 5.20, the contending users are divided among the waiting users, m_w and the users in the stack, m_s . However, it is impossible to calculate the average throughput if we split the two backlogged components because the solution of the stack only considers the number of colliding users at the start of the CRI. Therefore, we have to calculate the average throughput at each imbedded point by assuming that the sum of m_w and m_s is always constant within all imbedded points of the CRI when the system reaches equilibrium (see Figure 5.21). Based on this assumption, the throughput of the stack is calculated based only on the distribution of m_2 .

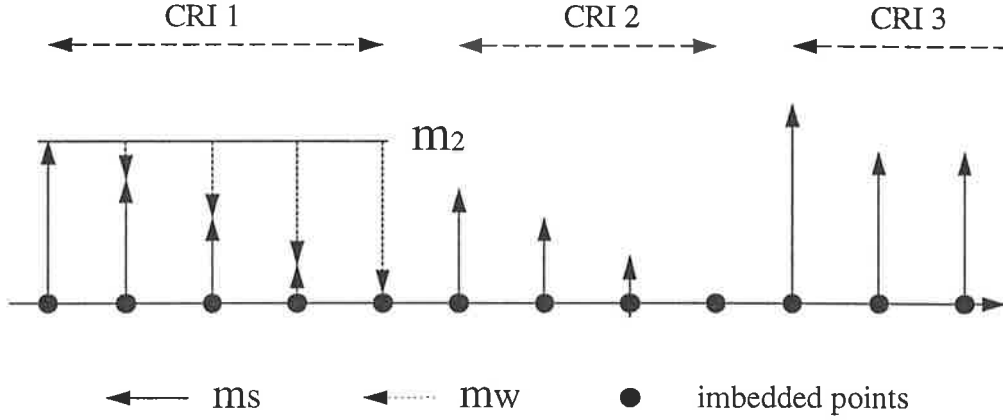


Figure 5.21 Relationship between the backlogged users in the Stack.

Note the constant m_2 assumption during a CRI. m_2 is also a random variable

To evaluate the average throughput of the stack we have:

$$\mu_{(k)} = \min \left\{ m_{2(k-1)}, Na \sum_{c=0}^{m_1} \left[\frac{c}{B(c)+1} \right] \beta_{in} \left(m_{1(k-1)}, c, \frac{m_{2(k-1)}}{m_{1(k-1)}} \right) \right\} \quad (5.43)$$

where Na is the number of access slots per frame. The expressions for the other transition probabilities are found in §4.5.4. A sample plot of the throughput-delay characteristics of ATDMA with tree algorithm is shown in the next figure. A comparison with the original ATDMA is shown where the main advantage of using the tree CRA is the stable throughput (steady curve) in the overload region. It is evident that there is no significant difference if the load of the access mechanism is low (i.e $G \leq 0.1$) but the nature of data traffic is the large degree of variation of the traffic statistics with which the need for a stable CRA is essential (refer to §5.4.3).

§5.6.2 Voice/Data Prioritised Stack ATDMA

We have evaluated the performance of a multiclass stack algorithm in §5.3 but not in a realistic environment. This sub-section is devoted to the performance of the algorithm under voice and data traffic where voice is taken with higher priority over data. This prioritisation is considered in many studies but it is not clear on what level of constraints the contention (access) delay may attribute to different services. Yet it is detrimental to voice quality, no justification has been made on why data services are taken with lower priority.

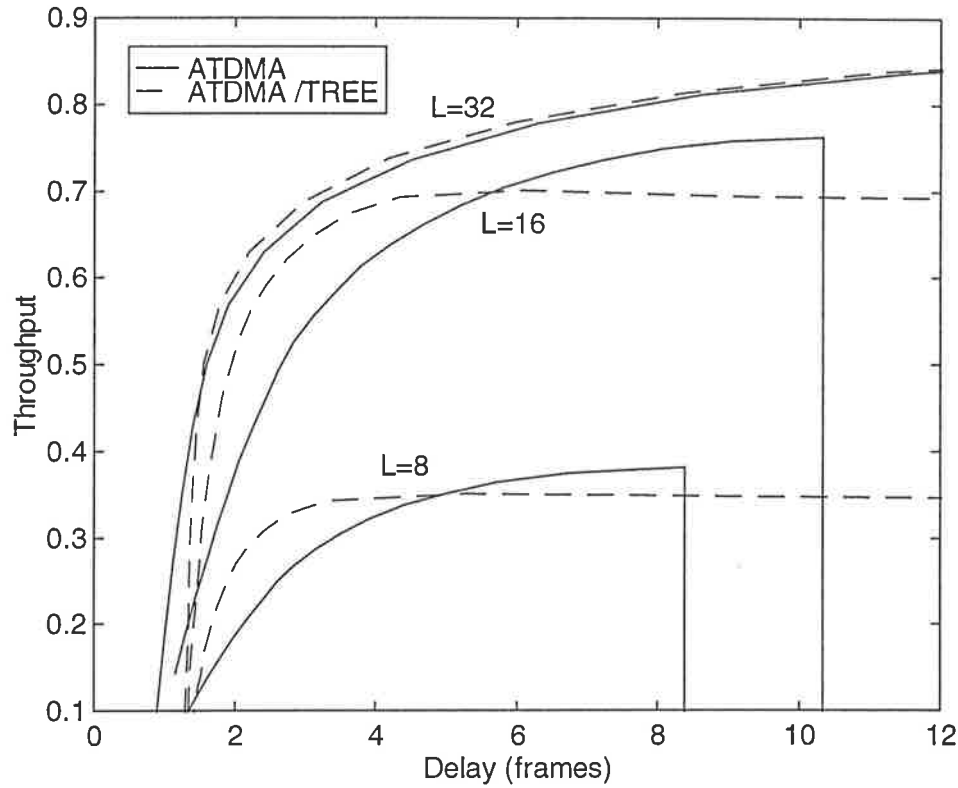


Figure 5.22 Comparison of ATDMA with ATDMA/TREE using TFA

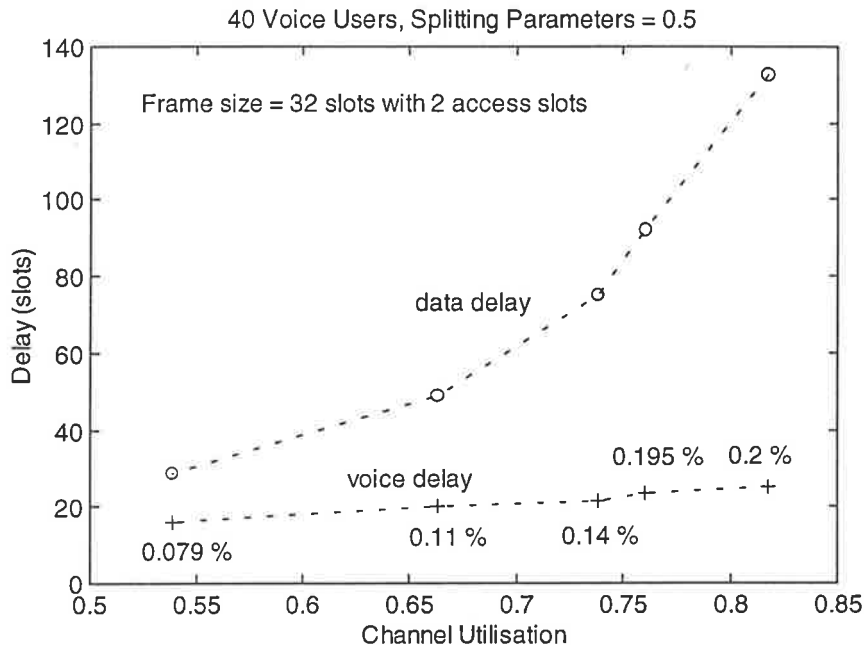
Parameters are $N=16$, $N_a=2$. The ATDMA retransmission probability, $p=0.1$, and the stack splitting parameter, $\phi=0.5$. The graph shows a stable throughput for the stack CRA.

But as shown in the proceeding chapters, data packet transmission can be expedited by using more resources (i.e. multislot reservation). Thus the delay budget at the contention phase can be relaxed in contrast to voice service.

We use a simulation model to determine the contention delay. We adopt a slot by slot simulation with fixed number of voice users alternating between active and idle states. The data messages arrive in a Poisson process from a single traffic generator. Data messages are exponentially distributed. The model consists of four modes for voice users namely silent, contention, channel allocation and reservation with three modes for data (excluding the silent mode). The CRA used is a two-priority stack algorithm. From the voice traffic statistics, a 45 percent utilisation is used if no data traffic is considered. The main objective is to enable the voice contention delay to remain at acceptable level even if the data load is increased.

Table 3 Simulation Parameters

Parameters	Values
Number of slots/frame, N	32
Number of Voice Users, M_v	40
Speech hangover	125 ms
Access slots	8, 16
Ave. data length, L	32 slots
Ave. talkspurt duration	1.103 sec
Ave. silent duration	1.956 sec
frame size	10 ms

**Figure 5.23 Performance of a Voice/Data System using Prioritised Stack CRA**

The corresponding packet dropping rate as a result contention for voice are shown in the figure. These values are much lower than the minimum requirement of one percent.

The simulation parameters are shown in Table 5.3 and the result is plotted in Figure 5.23. The effectivity of the two priority stack algorithm is demonstrated whereby the effect of increased data load does not result to significant increase in voice contention delay. Hence, the congestion is felt only by the data users. However, the access delay of 120 slots for data users at approx. 0.8 throughput is equivalent only to 40ms which is a very insignificant amount of delay in most data services. The steady access delay for voice also proved the effectiveness of the proposed multiclass stack algorithm

§5.7 Random Access and the Polling Solution

The design of adaptive protocols has been considered in some literature. Examples of such family of protocols are the URN Protocol, Split Reservation Upon Collision (SRUC) and Mini-slotted Alternating Priorities (MSAP). Details in the development of adaptive protocols can be found in [Tob80] together with their corresponding relative performance. The underlying concept in developing such protocols is to ensure that they will perform well at various traffic statistics as reservation technique alone is not sufficient to enable the protocol to be adaptive.

The stability of Local Area Networks (LANs) is achieved by its mechanism of scheduling the transmission of each node in the network. In the token ring protocol, the right of transmission is passed from one node to another once every node has finished transmitting all its stored information. For the star network, users are polled in a predetermined fashion to avoid any conflict during a transmission. The combination of *polling* and *reservation* can achieve a very good performance at the high load region. This is because polling can achieve a very high throughput when almost all users in the system are in the active state. In mobile data systems, the use of polling in the MAC protocols has already been adopted. However, it is mainly used in the channel allocation broadcast in order to uniquely identify the successful user during the random access process. Here, we use the polling mechanism as a back-up for the contention process to increase the stability and improve the system performance in the high load and near overload region (see Figure 5.24). Hence, an adaptive/hybrid protocol called **SCARP** [Buot96c] which exploits the frame structure of a reservation protocol to accommodate a polling mechanism was proposed.

§5.7.1 Protocol Description

In the wireless environment, a substitute for random access is *polling*. Polling ensures stability but it enduces large access delay even if the system is lightly loaded. Conversely, random access algorithms are prawn to instability or unfairness problems but attains short access delay in the low load region. To mention, S-ALOHA is unstable in some region but it attains a certain level of fairness. In contrast, stack algorithm is very stable but it exhibits unfairness due to its LCFS counter discipline. In this section a marriage between random access and polling is investigated. The random access will provide short access delay during light loads while polling will enhance the access

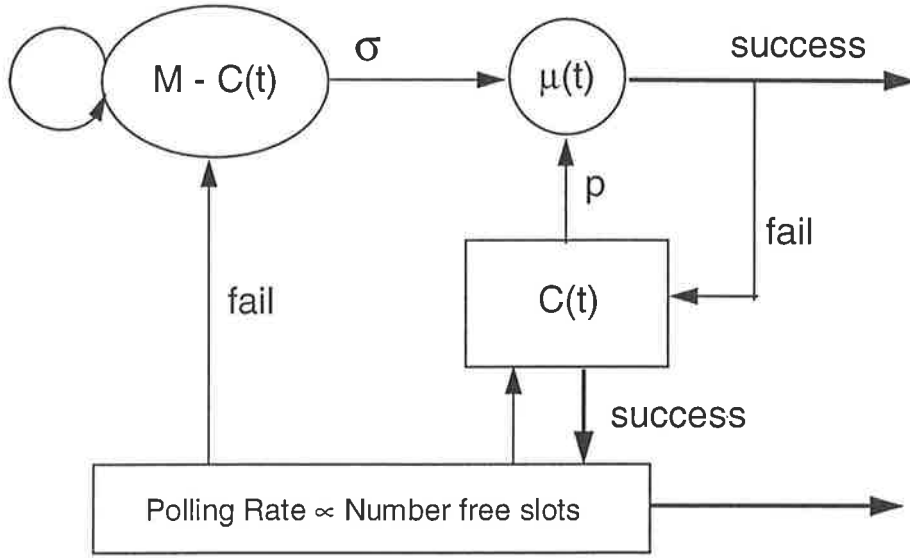


Figure 5.24 Model for combined S-ALOHA Random Access and Polling

performance at higher loads. The use of combined polling and random access was also demonstrated in [Lu94] and later in [LiMer94] all for integrated voice/data wireless system. In those protocols, the use of polling was mainly for high priority users. Also, polling is sometimes necessary during instants of large queued users so that the base station can sequentially poll each user to transmit a synchronisation burst to maintain synchronisation and power control. In this section, we combine polling, random access and reservation to develop an *adaptive reservation protocol*.

The proposed protocol **SCARP**, stands for *Silence-Contention-Acknowledgment-Reservation-with-Polling*. This protocol is a variant of Advanced TDMA in which the addition of a polling state has a two-fold advantage. First is its higher stability which provide fast recovery of the S-ALOHA when it operates in the high delay stable operating point or in the unstable region. In this case, a higher retransmission probability parameter can be used in the S-ALOHA thereby improving its performance in the underload region. Secondly, the frame structure does not have to be optimised according to the traffic statistics since the polling mechanism enhances its adaptability to various traffic types as it provides extra capacity to the access rate. So the SCARP protocol can manage to maintain higher throughput even if the average message length is relatively short.

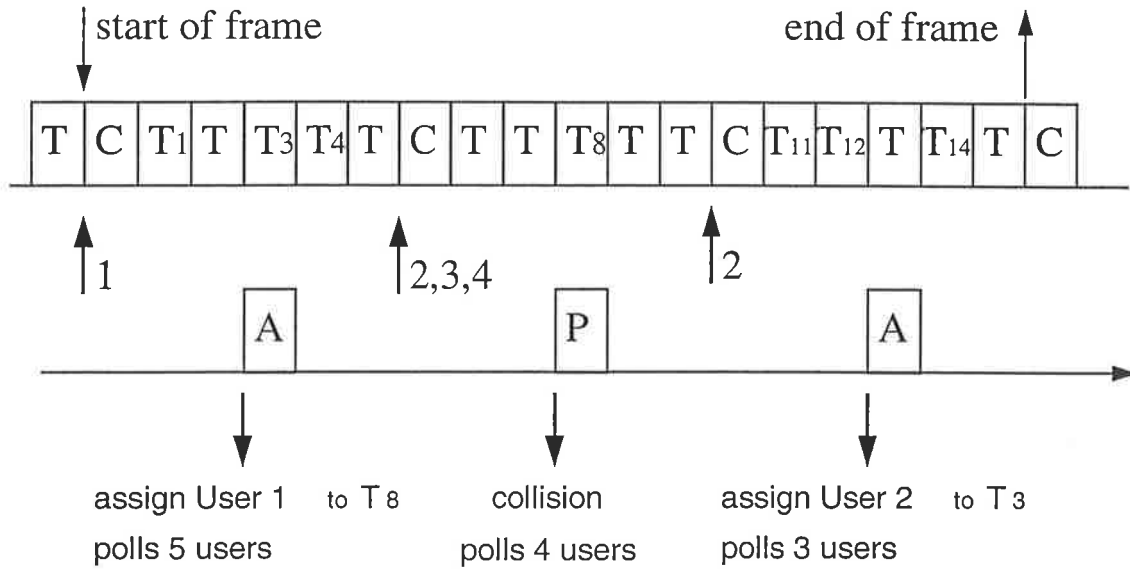


Figure 5.25 Sample Time Chart of Polling and Reservation Protocol

The figure demonstrates the behaviour of a combined random access and polling. At the start of the frame, user 1 executes a random access in the control slot (C) and then acknowledged in the first incoming A slot. At this time, the base station identifies 5 free traffic slots (T), therefore it polls a maximum of 5 users. In the next access slot, 3 users contend and a collision occurred. This time the base station can still poll 4 users. After which user 2 succeeded and the base station polls 3 users. Thus the polling rate depends on the number of free slots identified.

The channel structure of Advanced TDMA is used with the exception of the Fast Paging Acknowledgment slot. In an N -slot TDMA frame, random access (R) slots are allocated for random access in the uplink. These R slots are distributed in the frame to minimise the latency and to provide enough time for acknowledgment. As explained in [Dun194], every R slot is paired with an A slot for acknowledgment and slot allocation. The random access feedback requires only a small portion of the information field. Therefore most of the information field in the A slots are mainly used for slot allocation purposes. Thus, it is possible to provide multiple acknowledgment in one A slot. Also, a time shift is provided for the uplink and downlink greater than the round trip propagation delay + processing time for the contention process.

The principle behind the SCARP protocol is shown in Figure 5.25 above. The main idea of the SCARP is to provide a polling mechanism once there are free slots in the frame. The random access mechanism is mainly used over the polling mechanism and the proportion of their throughput depends on the access rate.

§5.7.2 State Transition Cycle

In the protocol, we assume that data information are generated by multiple users registered to a single base station. During the initial transmission, the terminal listens to one of the radio channels of the nearest base station and synchronise in order to contend together with the existing users. In the random access, the terminal identify itself as a new user and is subject for authentication. Upon authentication, a terminal identity, TI is allocated to every terminal accommodated by the base station. The TI must include a base station identifier in order to avoid confusion with the neighbouring cells.

The terminal activity is assumed to be alternating between *idle* and *active* modes. A terminal is said to be idle if it has no packet in its buffer. We describe the protocol cycle by starting with a user in the idle mode (S state). When it becomes active, it goes to the contention state (C) and attempts for contention in the first incoming R slot. Unsuccessful users retransmit in the next R slot with a probability p .

After a successful contention, the terminal goes to the acknowledgment state (A) and waits for a slot to be allocated. When a slot is available during the successful contention, the terminal is assigned immediately with a slot in the first A slot. Otherwise it waits on a FCFS basis. When a slot is allocated, the terminal moves to the reservation state (R) and start transmitting its packets. After all the packets in its buffer are transmitted, the terminal loose its reservation and goes to the polling state (P). At the polling state a terminal waits for the acknowledgment if all the packets are transmitted correctly. If all packets are successfully transmitted, the terminal goes back to the S state. Terminals in the silent and contention states are polled with lower priorities than the terminals in the polling state. The polling of the terminals in the silent and contention states is done in a cyclic fashion. This allows terminals in the contention states to have two access mechanisms (polling and random access). If the packet error probability is low, the rate of polling for the contending terminals is high thereby reducing the average access delay. The process is shown in Figure 5.26.

§5.7.3 Throughput/Delay Approximation

Here we present an approximation based on TFA. Since the polling mechanism involves a common process to both the A and P states. The Markov model of the state transitions are simplified in the Figure 5.27. The exclusion of the polling state is due to two reasons.

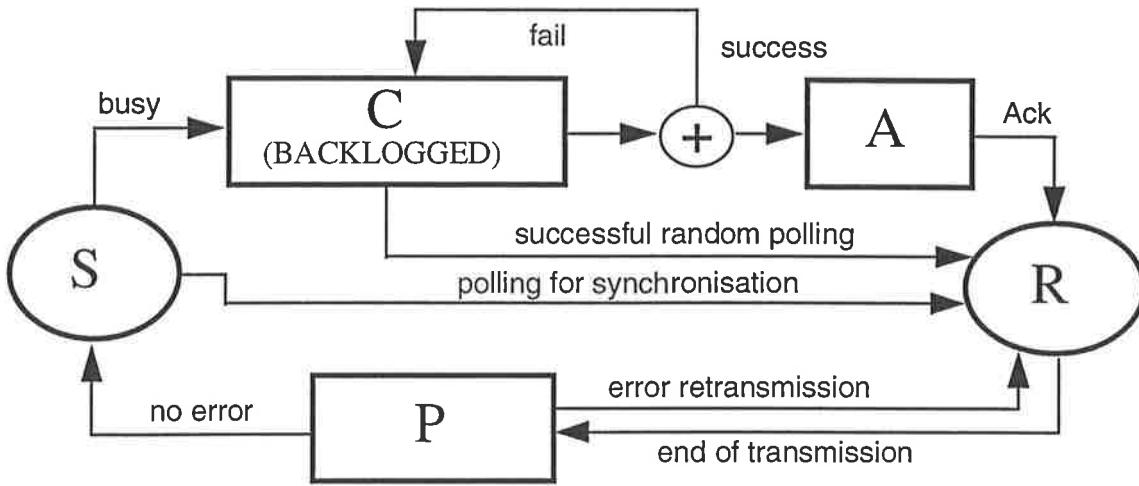


Figure 5.26 State Transition Cycle

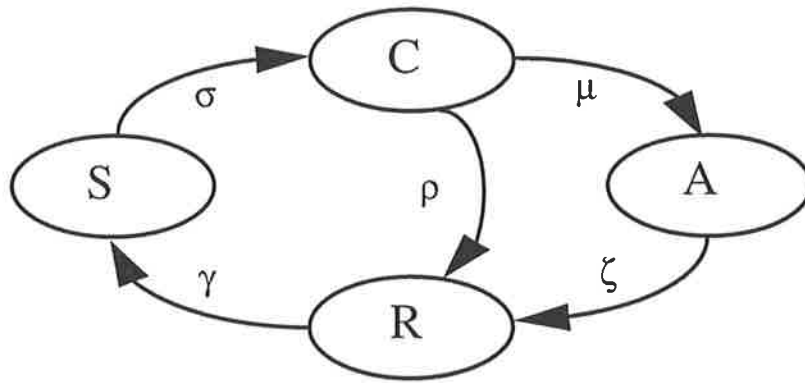


Figure 5.27 Simplified Markov Model

One is due to the assumption of a noiseless channel. The other is the small probability that a user in the polling state will become active and attempt a random access then become successful. The TFA of the SCARP protocol is the same as that of ATDMA with the addition of the polling mechanism. Denoting the number of users in the idle, backlogged, queue and reserved states as m_1 , m_2 , m_3 and m_4 , we have:

$$\begin{aligned} \mu'_{(k)} = & m_{2(k-1)} p (1-p)^{m_{2(k-1)}-1} \left(1 - \frac{\sigma}{Na}\right)^{m_{1(k-1)}} \\ & + m_{1(k-1)} \frac{\sigma}{Na} \left(1 - \frac{\sigma}{Na}\right)^{m_{1(k-1)}-1} (1-p)^{m_{2(k-1)}} \end{aligned} \quad (5.44)$$

and

$$\mu_{(k)} = \min \left\{ \left(m_{2(k-1)} + \sigma m_{1(k-1)} \right), \mu'_{(k)} Na \right\} \quad (5.45)$$

The probability of successful polling is the probability that the polled user is backlogged and it did not successfully contend in the current slot. Then the rate of successful polling, ρ is the product of the polling rate and the probability of successful polling given as

$$\rho' = \Pr\{\text{success}\} \text{ Polling Rate}$$

$$\rho'(k) = Na \left[\frac{m_{2(k-1)} - \mu(k)/Na + \sigma m_{1(k-1)}}{m_{2(k-1)} + m_{1(k-1)}} \right] \left[Nt - m_{4(k-1)} \left(1 - \frac{\gamma}{Na} \right) \right] \quad (5.46)$$

As in ATDMA, the system states are calculated as follows:

$$m_{1(k)} = (1 - \sigma)m_{1(k-1)} + \gamma m_{4(k-1)} \quad (5.47)$$

$$m_{2(k)} = m_{2(k-1)} + \sigma m_{1(k-1)} - \mu(k) - \rho(k) \quad (5.48)$$

$$m_{34(k)} = m_{34(k-1)} - \gamma m_{4(k-1)} + \mu(k) + \rho(k) \quad (5.49)$$

The throughput and delay are calculated similar to that of ATDMA in §4.5.4. The throughput-delay characteristics of SCARP is plotted in next figure. It shows that even for small values of the average message size, L , the protocol still exhibit a steady throughput/delay.

§5.7.4 Calculation based on the Polling cycle

Another approach in calculating the polling process is based on the polling cycle time. The calculation of ρ assumed a *roll-call* polling sequence. Therefore we are interested in calculating the probability that a polled user is busy and has not successfully contended in the R slots it has passed. First, the length of the polling cycle has to be calculated as

$$t_{p(k)} = \left\lceil \frac{m_{1(k-1)} + m_{2(k-1)}}{Nt - m_{4(k-1)} + m_{4(k-1)}\gamma} \right\rceil \quad \text{access slots} \quad (5.50)$$

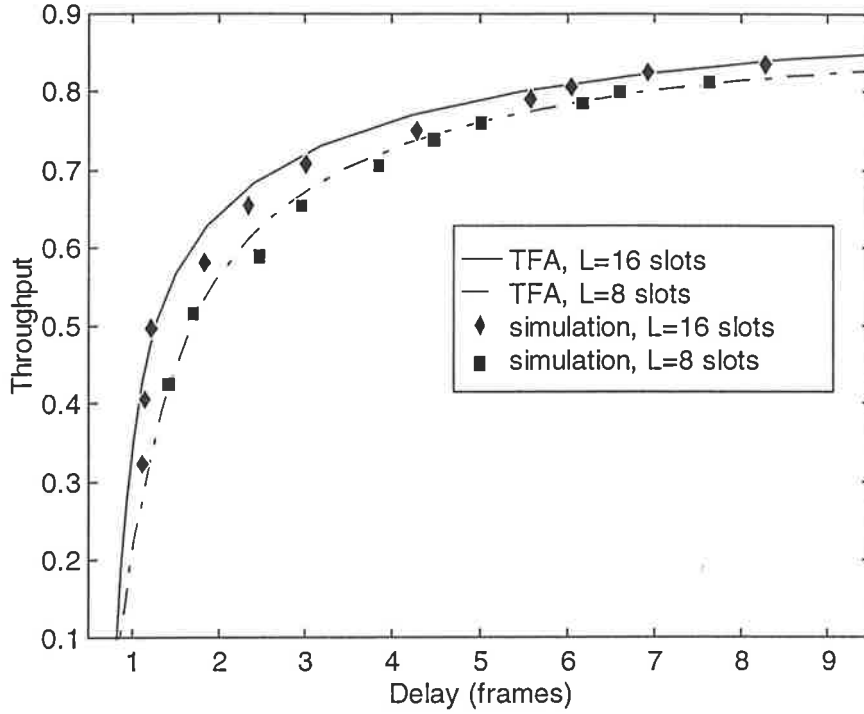


Figure 5.28 SCARP Protocol Performance using TFA and Simulation Methods
Discrepancies are due the bistable behaviour of S-ALOHA Parameters: $N=16$, $N_a=2$, $p=0.1$

The probability that a particular user will be successful in any access slot while it is active is

$$\rho_{s(k)} = \frac{\mu(k)}{m_{2(k-1)} + m_{1(k-1)}\sigma} \quad (5.51)$$

Thus the probability that a user becomes successful in the t^{th} access slot after it becomes active is geometrically distributed expressed as.

$$P_s(t)_k = \rho_{s(k)}(1 - \rho_{s(k)})^{t-1} \quad (5.52)$$

where $t = \{1, 2, 3, \dots, t_p\}$. Then the probability that the user is successful before it is polled U_p , is expressed in (5.53) and (5.54) and the probability of successful polling is in (5.55).

$$Fp(t)_k = 1 - \exp\left(-\frac{t}{T_s}\right) \quad (5.53)$$

$$U_p(t_p)_k = \frac{\sum_{z=1}^{t_p} \sum_{m=1}^{t_p-z} P_s(m)_k Fp(z)_k}{\sum_x Fp(x)_k} \quad (5.54)$$

$$\Pr\{\text{user polled}\} = \Pr\{\text{user busy}\} \Pr\left\{\begin{array}{l} \text{user fails contention} \\ \text{until } t_p \text{ duration} \end{array}\right\} \Pr\left\{\begin{array}{l} \text{user fails contention} \\ \text{in the current access slot} \end{array}\right\}$$

$$\rho_{f(k)} = Fp(t_p)_k (1 - U_p(t_p)_k) \left(1 - \left[\frac{\mu(k)}{m_{2(k-1)} + m_{1(k-1)}\sigma} \right] \right) \quad (5.55)$$

§5.7.5 Stability

We compare the stability of the SCARP to that of the ATDMA protocol using the drift parameter (expected increase/decrease in the number of backlogged users). The calculation limits only to the mean values of the drift based on a constant throughput assumption. For a given throughput, the arrival rate can be calculated as $\lambda = \frac{S N}{L N_a}$.

From then our mean number of free slots is $N(1 - S[1 + \gamma])$. Since the SCARP requires the mean number of idle users, we use the equivalent source model where $m_1 = \lambda N_a / \sigma$. From here we can calculate the throughput of the backlogged state as

$$\rho(m_2) = \frac{m_2}{m_2 + m_1} [N(1 - S(1 + \gamma))] \quad (5.56)$$

$$\mu(n) = \lambda e^{-\lambda} (1 - p)^{m_2} + e^{-\lambda} m_2 p (1 - p)^{\max\{0, m_2 - 1\}} \quad (5.57)$$

for the polling and random access respectively. The drift for SCARP is $\lambda - \rho(m_2) - \mu(m_2)$. A sample comparison of the stability of SCARP to that of ATDMA is shown in Figure 5.29. It shows a large difference of the drift characteristics of SCARP against ATDMA where the SCARP protocol has negative drift for all large values of the number of backlogged users.

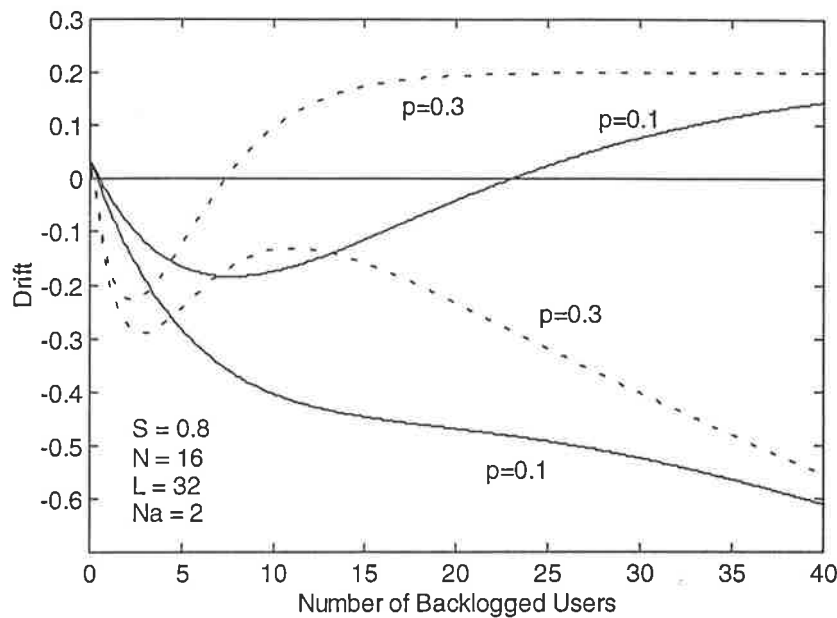


Figure 5.29 Drift as Function of Backlogged Users for SCARP and ATDMA

The variation of the retransmission probability is known to affect the stability of the ATDMA. However, the SCARP protocol can use higher values of p in order to reduce the access delay since the problem of stability is already handled by the random polling mechanism.

§5.7.6 Simulation Model

In the simulation, a noiseless channel is used with 16 slots per TDMA frame. Two access slots were allocated, sufficient to provide fast access with minimum overhead with the used traffic statistics. The process starts with all users in the silent state. Since a single arrival model is used for the message generation, a user is allowed to wait indefinitely if it is either in the contention or acknowledgment states. In the case of polling, a sequential polling discipline is used. Users just being polled and users just finished transmitting are held at the end of the polling sequence. Since a noiseless channel is assumed, only users in the silent and contention states are polled.

The advantage of using combined polling and random access is clearly shown in the comparison with SCARP and ATDMA. From the results, the ATDMA is unstable when the access rate is high caused by the reduction of the average message length. Also, when the operating point of the Slotted ALOHA is already close to the maximum throughput (0.36), the tendency that the contention process to flip-flop between the two stable regions cannot be avoided causing a large access delay. Even when the access rate is relatively low, the SCARP protocol still exhibit superior delay performance. Polling compensates the contention process when the S-ALOHA operated in the high

delay region as it pulls back the operating point to the low delay region. Under the SCARP protocol, it is also possible to use higher retransmission probabilities since the contention process can already avoid the unstable state. This enable the system to obtain shorter access delays. The results are plotted in Appendix G.

§5.7.7 Base Station Polling Control

In the plot of the percentage of polling to the in the access mechanism demonstrate that the polling mechanism is only effective at higher loads. That is, when the average number of backlogged users is large. Thus to increase the success rate of the base station and avoiding unnecessary polling when the Random Access mechanism is lightly loaded, the base station has to estimate the number of backlogged users and only initiate a polling mechanism once the estimated number of backlogged users suggest a higher successful polling rate. The control mechanism can be implemented using a Pseudo-Bayesian algorithm as shown in Figure 5.30.

§5.8 Integrated-SCARP Protocol

The use of polling and random access is primarily for the integration of real time and delay tolerant services (i.e. voice with random access and polling for data). However, the SCARP protocol can operate in a homogeneous user environment as shown in the previous section. The concept of a third generation terminal is to enable to achieve connectivity across different environment (WLAN, cellular, PCN, etc.). The current approach to achieve such connectivity is to employ multimode handsets to support various protocols for the different wireless systems. However, a more efficient scheme is to design a protocol that suits to the two different wireless environments. Here, we introduce the I-SCARP protocol to support voice and data users.

In the frame structure of SCARP (which is ATDMA), we have R slots (uplink) paired with A slots (downlink) and I slots. Users that would like to attach to a particular central control (base station) select a radio channel and send a reservation packet (RP) on the first incoming R slot indicating whether it operates as a LAN terminal or as a PCN terminal. Upon registration/attachment to the central control, a terminal is assigned with a terminal identity (TI) then joins the rest of the active terminals. For a LAN terminal with ready-to-send data packets (DP), it waits for a broadcast of the

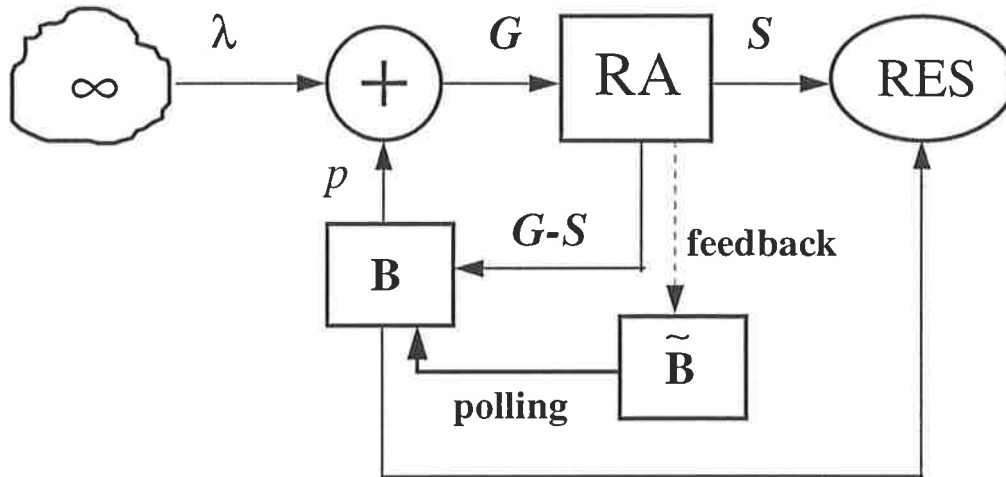
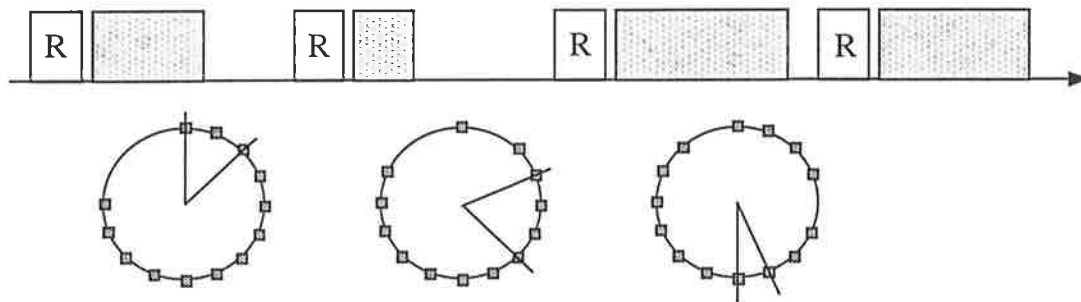


Figure 5.30 Polling Control Mechanism

Figure 5.31 Polling in TDMA used in the I-SCARP for LAN terminals
(Shaded areas correspond to busy slots)

polling packet (PP) with its TI included in the address field. When a terminal with DP's is polled it transmit immediately to the upcoming I slot indicated in the assignment field. The broadcast of the PP's are scheduled at every A slot. When a LAN terminal will send a voice packet (VP), it contends for a reservation the same as that of the ATDMA. The same procedure applies to that of the PCN terminals with all traffic types (voice or data). All LAN terminals attached to the central control are limited to the polling mechanism for the transmission of data packets. Figure 5.31 shows the polling mechanism with success and new arrivals departing and joining the ring and the size of the polling window is proportional to the number of free slots in every frame. The performance of this protocol depends on the traffic statistics of the applications supported. Unfortunately, it is not shown in this thesis.

§5.9 Summary of Chapter 5

In this chapter, design improvements for R-TDMA are proposed. Firstly, the performance of ATDMA is evaluated for voice traffic. The voice-only system is considered as a basis for the frame optimisation. The maximum capacity is evaluated using different values of speech hangover. It is found that significant improvements can be obtained by properly choosing a hangover value. The second improvement can be obtained by an appropriate FEC scheme and capture capability. It shows that a more powerful FEC is required for the reservation packet. The use of multiple receivers is also advantageous as the performance of the random access is affected by any overlap of the coverage from multiple antennae with the same radio channel. However, the throughput degradation can be compensated by strong capture.

The third improvement is the use of prioritisation at the random access. The S-ALOHA with retransmission priorities is investigated. Further improvement is achieved by using a prioritised stack algorithm. The algorithm has been tested for three-level priority. Later, the scheme is tested in voice/data protocol. The last among the design improvements is the use of hybrid access which is the combination of polling and S-ALOHA. The protocol (SCARP) is adaptive and stable and is capable of handling steady, periodic and bursty traffic (I-SCARP). Analysis is also presented showing a good agreement with the simulation results.

Chapter 6

Multimedia Access Protocol

The previous chapter was concerned with the design and performance of reservation based TDMA protocols and the focus of the techniques were on the access mechanism. In this chapter we digress to the channel allocation mechanism and propose some methods to enhance the flexibility so that a wide range of input traffic (multimedia) can be supported by R-TDMA. The objective of a channel allocation mechanism is to match the QoS requirements of each user to the available resources in the multiaccess environment.

Because multimedia refers to a mix of services with higher degree of dissimilarities, the multiaccess protocol must be flexible enough to adapt with the various requirements. In R-TDMA, the use of multislot reservation is almost essential for multimedia support. This would mean rate adaptability for synchronous (connection oriented) traffic using multirate (MR) and for asynchronous (connectionless) traffic using variable bit rate (VBR) transmission. The MR transmission is supported with bit rates multiple of the speed of a single slot up to the maximum speed equivalent to the maximum number of slots per TDMA frame. For the VBR transmission, the reservation technique can be fully exploited using reservation policies like best effort, threshold transmission, prioritisation, etc.

The resource allocation problem in multimedia systems must also incorporate the physical layer characteristics due to the scarcity and the quality variability of the radio channel. In the later part of this chapter, the association of the physical layer to the multiaccess layer is taken into account in order to attain some improvements of the QoS of the multimedia services. In particular, the technique of using variable coding rate and multislot reservation as well and variable source coding for video with respect to the channel load and multislot capability are investigated in this chapter.

The heart of a multimedia multiaccess is a central resource allocator which manages the available resources to a number of competing applications (see Figure 6.1). Each user transmit through an appropriate Service Access Point (SAP) in the multiaccess layer. The physical resources are organised based on the available

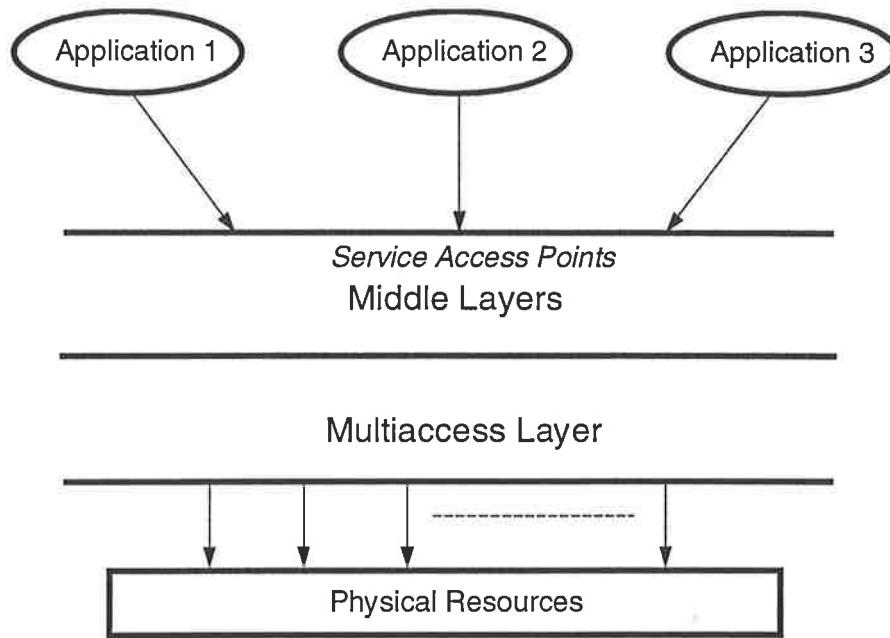


Figure 6.1 Association of Multiaccess Layer in Multimedia Transport

information from the multiaccess layer and the characteristics of the radio channel. To deliver a service with a defined QoS a user has to negotiate with the multiaccess layer for some resources. Then the multiaccess layer assign some resources (capacity) to the user. In the event that the required capacity is not granted, the multiaccess layer will negotiate with the physical layer regarding the quality/speed trade-off to be assigned to the particular user. In this way a greater flexibility in terms of QoS maintenance can be achieved.

§6.1 Multislot Reservation for Multimedia R-TDMA

The Multislot reservation scheme is also called as Generalised TDMA [RomSi90] whereby users are allowed to own one or more slots per frame. It has also been described in many reservation protocols like the R-ALOHA, PRMA, ATDMA and others as part of the protocol's capability. However, in the above mentioned protocols multislot reservation has not been carefully considered as a means of achieving the various QoS requirements of multimedia services. The next subsections consider the different multislot reservation schemes in R-TDMA protocols. Then an approximate analysis is presented based on Markov models.

§6.1.1 Multislot Allocation Schemes and Fairness Criteria

The problem of fairness in the allocation of the spare channels in the multiaccess protocols always exist since most often the central control has no knowledge with the users' demands. The initial solution to this problem is the assignment of priorities to different classes of users so that a portion of the available resource can be allocated to each priority. Eventually, a problem arise when two or more users belonging to the same class compete for some resources. In the context of multislot reservation, fairness can be related to an appropriationment problem since the user demands are proportional to the size of the messages stored in their corresponding buffers contending for limited channels [IbaK88]. However, traditional appropriationment solutions are not applicable since the system is dynamic. In our case, we define the fairness, Φ in terms of the delay from

$$\Phi = \max \left\{ \frac{x_j}{L_j} \right\} - \min \left\{ \frac{x_j}{L_j} \right\} \quad \forall j \quad (6.1)$$

where x_j and L_j are the number of channels reserved and the message length respectively by user j . The objective is minimising the fairness function subject to $\sum x_j = N$ and $x_j = \{0, 1, 2, \dots, x_{max}\}$. Since L_j/x_j is equivalent to the transmission delay, D_j , we have

$$\Phi = \text{Minimise} \quad \left\{ \max\{D_j\} - \min\{D_j\} \right\} \quad (6.2)$$

subject to $1 \leq D_j \leq L_j$

However, the messages does not occur at the same time so that we only take the mean delay as the reference for the fairness. Instead, differences from the mean delay are taken and consequently modifying the function as

$$\Phi \propto f(D) = \sum |D_i - \bar{D}| \quad \forall i \quad (6.3)$$

By averaging the square of the differences, we arrive naturally with the variance formula. Thus the fairness is equivalent to minimising the variance of the delay. To achieve this criterion, we will use a longest service time first (LSTF) multislot allocation policy for the spare channels. Prior to the discussion regarding the multislot algorithms, we first identify the resource allocation schemes required in a reservation protocol.

There are three main types of multislot reservation policies that can be adapted in the resource allocation of multimedia traffic. They are as follows:

- Threshold type Channel Request
- Best Effort Channel Allocation
- Multirate Channel Allocation

The *threshold type* request is the simplest multislot reservation policy wherein the busy users requests for additional slots if the reserved slots are not sufficient to meet the message delay requirement predefined based on the QoS criteria. The threshold parameter determines the request status of each user. It is usually calculated based on the remaining packets stored on the user's buffer and the number of slots currently reserved by the particular user. The second policy (*best effort*) is a centrally controlled scheme whereby the resource allocator requires the necessary information of each user to optimise the available resources. Both the best effort and the threshold type policies are suitable only for asynchronous transmission because the additional slots are allocated in terms of their availability. The third policy is for *multirate* transmission. It is suitable for synchronous transmission. In this scheme the users requests for a number of slots defined in the service criteria. The user will start transmitting only when the required number of slots are allocated otherwise it will wait for an allowable queuing delay until the slots are allocated. The multirate policy is characterised in terms of its delay and blocking parameters. The multirate policy can be combined with the best effort and threshold-type policies.

§6.1.2 Approximate Analysis using Birth and Death Markov Chains

Birth and Death Markov models are very popular in analysing TDMA systems. Here we will use this tool to model the multislot system in obtaining the throughput delay characteristics. The crucial part in the modelling is finding the transition rates based on a chosen multislot reservation policy. For simplicity and generality, we will use a threshold type policy and calculate the performance using the rigorous time domain analysis. There are two parameters that has to be identified in solving the performance of multislot R-TDMA protocols. They are the user distribution and the

channel occupancy distribution. These two parameters can be represented by two variables that are dependent and subject to the reservation policy.

A Birth and Death Markov chain approach was used to evaluate the performance of this system (see Figure 6.2). For a Poisson arrival process, it is quite trivial that $\lambda_k = \lambda \forall k$. But the problem is in the value of μ_k which depends on the request threshold. Now it is necessary to determine how many channels are requested by one user if its message length, L is negative exponentially distributed. It is quite convenient to use an exponential model so that the request probability distribution of each user will be stationary throughout the duration of the message (memoryless property). Solving the request probability distribution we have

$$Pms_k(m) = \Pr\{m \text{ slots requested} \mid k \text{ users transmitting}\}$$

$$Pms_1(m) = \Pr\{m \mid k = 1\} = e^{-\frac{(m-1)v}{L}} - e^{-\frac{mv}{L}} \quad m=1,2,3,\dots \quad (6.4)$$

where v is the request threshold which means each user is eligible to reserve a maximum of $\left\lfloor \frac{L(t)}{v} \right\rfloor$ slots per frame where $L(t)$ is the message size in time t . For k , ($k \leq N$) busy users, the resulting channel request probability will be a k -fold self convolution of Pms_1 denoted as

$$Pms_k = \Pr\{m \mid k\} = \mathcal{D}(k, Pms_1). \quad (6.5)$$

Then the mean number of reserved channels for k users is given as

$$n_{ch}(k) = \sum_{j=k}^{N-1} j Pms_k(j) + N \sum_{j=N}^{\infty} Pms_k(j) \quad (6.6)$$

and the average departure rate for k users, $\mu(k)$ is

$$\mu(k) = \frac{n_{ch}(k)}{L}. \quad (6.7)$$

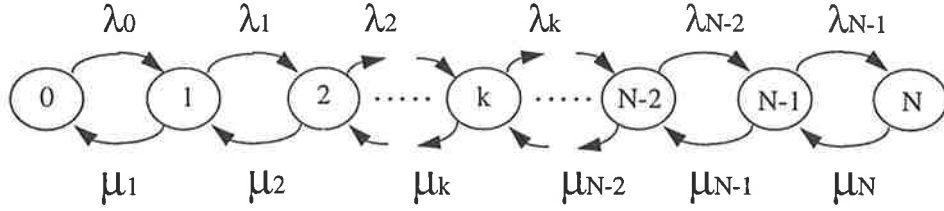


Figure 6.2 Markov Chain with N channels (λ arrival rate, μ departure rate)

Forming the Equilibrium state equation from the Markov Chain, we have

$$P_k = \Pr\{\text{state} = k\} = \frac{\lambda^k}{\prod_{j=1}^k \mu(j)} P(0). \quad (6.8)$$

Since $\sum_k P_k = 1$, $k = 0, 1, 2, \dots, N$ then we result to

$$P_0 = \frac{1}{\sum_i \frac{\lambda^i}{\prod_{c=1}^i \mu(c)}}. \quad (6.9)$$

After the state occupancy as well as the channel request probabilities are determined, then the throughput/delay performance are determined using Little's Result. The throughput equation becomes

$$S = \frac{\sum_{k=1}^{\infty} P_k n_{ch}(k)}{N} \quad (6.10)$$

and the message delay is calculated as

$$Dm = \frac{\sum_{k=1}^{\infty} k P_k}{\lambda}. \quad (6.11)$$

The delay unit is in frames and the previous equation does not account for the latency incurred in the channel allocation process of which is proportional to the throughput.

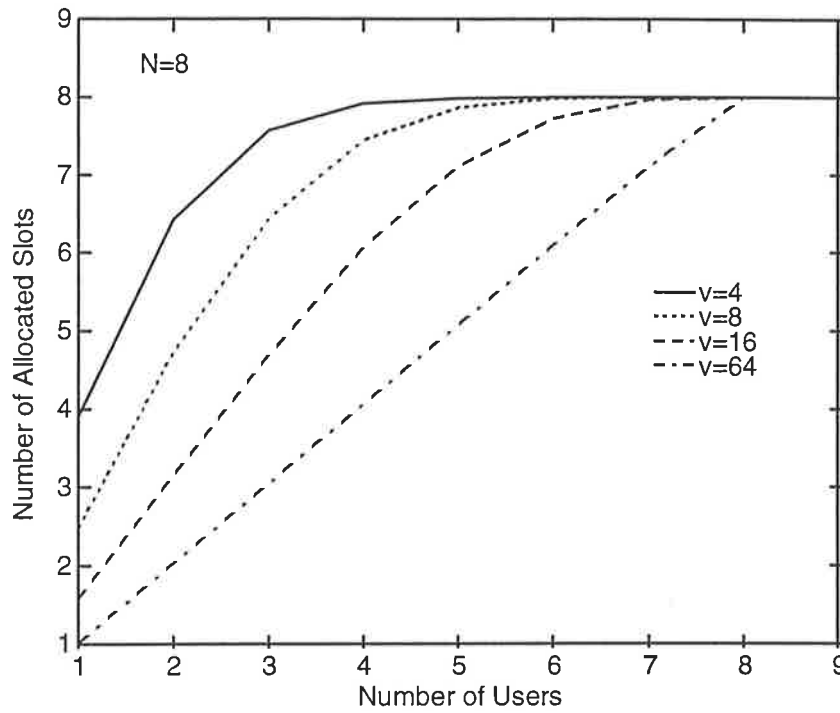


Figure 6.3 Plot of Number of Users vs the Number of Reserved Slots

The average message size, $L=16$ so that for a single user, approx. 4 slots are reserved. The increase in threshold, v reduces the average number of slots per user. $v=64$ resembles TDMA.

Sample calculations are plotted in Figure 6.3 - 6.5. The total channel request versus the number of users are shown. A typical delay curve is also produced with the different request threshold v . The results clearly favours for the lower values of threshold so that users can request more channels. A very large threshold would result in a performance comparable to TDMA with single slot. However, very small values of the threshold will reduce the fairness of the allocation policy as users with short messages will intend to require more slots than what is required.

§6.1.3 Approximate Analysis using Discrete Markov Analysis

The solution of TDMA in which users are assigned with fixed slots in every frame has been considered in [Lam77] [RomSi90]. In those papers, transform techniques with the use of generating functions simplify the solution. In this study, a solution based on Markov model is used for a fixed users dynamic TDMA. Which means, users are not assigned with fixed slots in every frame but rather assigned on demand basis. The use of Discrete Markov Analysis was used in [Mor84] in the Asynchronous-Reservation Demand-Assignment (ARDA) protocol.

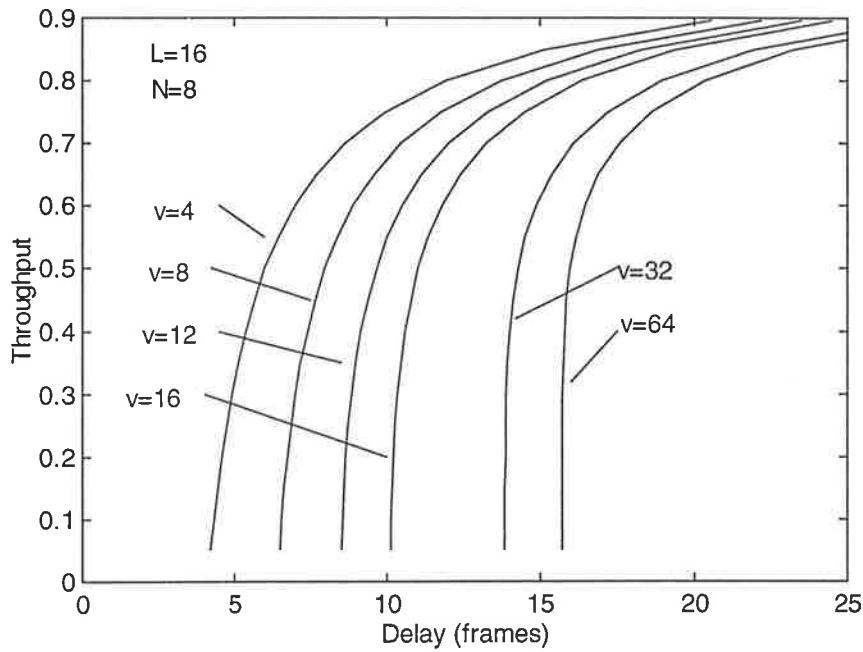


Figure 6.4 Plot of Throughput and Delay for Poisson Input

The small values of v is in favour for the mean throughput-delay performance. However, at higher loads, the effect of multislot reservation is marginal for most values of v . Much improvement is expected if the number of slots per frame is large.

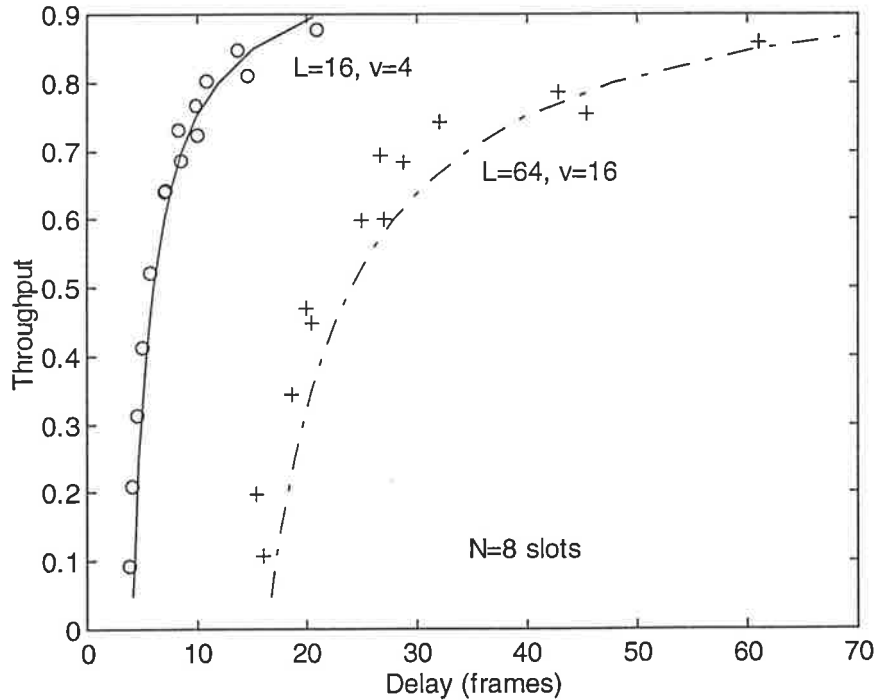


Figure 6.5 Simulation and Approximate Analysis of Multislot TDMA

A good fit of the solution to that of the simulation results. The discrepancy is due to the latencies involved in the simulation.

It was then used in [QiWyr94] in the analysis of PRMA. Some solutions for TDMA are also found in [Rub79] [RubT89]. Here we consider the Discrete Markov Analysis to obtain the performance of multislot reservation scheme. this solution is an extension of the method used in §4.4.1. This method is very useful in the multislot TDMA since the slot allocation algorithm will identify the transition probabilities in the system.

Here we consider a slotted channel in which a TDMA frame consist of N slots shared by M uniform users. Usually $M \gg N$ so that active users are served on a First Come First Served (FCFS) basis. We neglect the process how the users contend so that our problem will be concentrated on the analysis of TDMA itself. Every user alternates in two modes which could either be idle (no packets to send) or active (has ready to send packet(s)). User that are active are allocated with a slot(s) and remain on that slot(s) until the last packet of the current message is transmitted. Without the loss of generality, we assume that every time a user reserves a slot, it has only one message to transmit. If the number of active users are more than the available slots N some of the users are held in a queue. Therefore our users are either in one of the following states: idle, queued or transmitting. The number of users in each of these states are S, Q and R respectively. The transition of the users in these three states form an imbedded Markov chain which is our main interest.

Our main objective is to evaluate the system steady state from the state transition probabilities. Since our system is composed of three states, it is sufficient to describe the transition probability in terms of two variables. Let $P(m, n | Q, R)$ be the conditional probability that there are Q users in the queue and R number of users transmitting in frame k , and correspondingly m and n users in frame $k+1$. Then the probability that there are m and n users $\Pi(m, n)$ in the queue and transmission states is expressed as

$$\Pi(m, n) = \sum_{R=0}^N \sum_{Q=0}^{M-R} P(m, n | Q, R) \Pi(Q, R) \quad (6.12)$$

conditioned that

$$\sum_n \sum_m \Pi(m, n) = 1 \quad (6.13)$$

From the two expressions, we have to determine the joint one-step transition probabilities $P(m,n/Q,R)$ form the arrival and departure process in the system. Assuming that the user modes are both exponentially distributed, the probability that a user leaves the silent state is

$$\sigma = 1 - e^{-\left(1/T_s\right)} \quad (6.14)$$

where T_s is the mean silent duration. The probability of a user leaving the transmission state depends on the number of slots reserved by a user. The resulting transition probability becomes a combination of different distribution with corresponding weights (i.e. weighted exponential distribution). However, the determination of the weights is not trivial so that we recourse to the averaging of the transition probability. Thus every users leaves the transmission state with a probability

$$\gamma = 1 - e^{-\left(N/LR\right)} \quad (6.15)$$

where T_s is the mean idle period and L is the mean message length. Then the probability of x users becoming active in the incoming frame is

$$A(x|R,Q) = \beta \text{in}(M - R - Q, x, \sigma) \quad (6.16)$$

Similarly, the number of users departing from the transmission state is

$$D(y / R) = \beta \text{in}(R, y, \gamma) \quad (6.17)$$

Since every user has r slots where $r = \{1, 2, \dots, N-R+1\}$, then for y departing users, there are y_r slots becoming free where $y_r = \{y, y+1, \dots, N-R+y\}$. Since the conditional probability of y_r slots free given r departures is required to calculate the transition probabilities from the queue to the transmission state, the distribution of y_r must be known. Since the number of slots is relatively large, the use of multinomial distribution may be inappropriate as it will result to a multi-dimensional array. Here we can approximate the distribution using the binomial approach.

Since every user has at least one slot reserved, the number of slots available for multislot reservation is $N-R$. If we assume that all users have equal share to the number

of slots in every frame (fairness), if y users depart, the probability of any slot becoming free is y/R . Thus we can express y_r for i slots becoming free for every y departing users as follows

$$y_r(i/y) = \beta_{in}((N-R+y), (i-y), y/R), \quad i \geq y \quad (6.18)$$

The multislot reservation is based on a greedy allocation where arriving and existing users reserve all slots in the frame. This means, the number of reserved slots are either N or 0 . This assumption is valid if the average length of the message is much greater than the frame length or if the load is high. We are interested in the transition from state (Q, R) to state (m, n) based on the allocation procedure in the algorithm.

Consider Figure 6.6 where y_r slots become free on the departure of y users. If the queue is not empty and the number of users in the queue are greater than the number of free slots, the users in the queue including the new arrivals, A are allocated on a FCFS discipline. If there are more available slots, the excess slots are allocated to users with the relatively longer messages including the user that are already served. From the figure, the transition from Q to M depends on the arrival A and the number of slot becoming free. Whenever there is no arrival on the previous slot, the changes in Q and R occurs only when a departure happens in the current frame. Also, R is independent of the arrival. The allocation process can expressed in the following equations:

$$Q_{out} = \begin{cases} A + Q - m & \text{(single slot all users) if } A + Q > y_r \\ A + Q & \text{multislot for } \geq 1 \text{ users if } A + Q \leq y_r \end{cases} \quad (6.19)$$

This also mean that $n \geq R$ if $m > 0$ and $n = Q_0 + R - D$. If $\pi(a, d)$ is the probability of a arrivals and d departures for a pair of R and Q , and $\Delta(s/d)$ is the conditional probability of s slots free given d departures, then the transition probabilities are calculated as follows:

Case 1 $m=Q$ & $n=R$

$$P(m, n / Q, R) = \sum_{k=0}^N \pi(k, k) \Delta(k / k) \quad (6.20a)$$

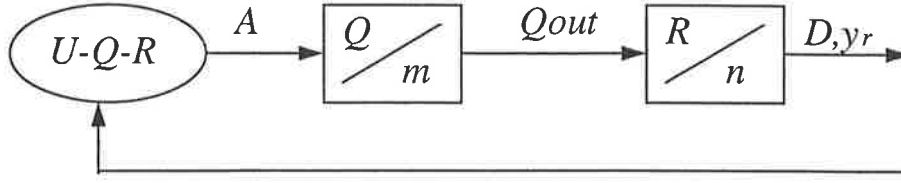


Figure 6.6 Transition Diagram of Multislot Reservation TDMA

Case 2 $m > Q$ & $n = R$

$$P(m, n / Q, R) = \sum_{k=(m-Q)}^N \pi(k, k - (m - Q)) \Delta(k - (m - Q), k - (m - Q)) \quad (6.20b)$$

Case 3 $m < Q$ & $n = R$ & $m \geq 0$

$$P(m, n / Q, R) = \begin{cases} \pi(0,1)\Delta(1/1) + \pi(0,2)\Delta(2/2) + \pi(2,3)\Delta(3/3) \dots, & (Q-m)=1, m > 0 \\ \pi(0,2)\Delta(2/2) + \pi(1,3)\Delta(3/3) + \pi(2,4)\Delta(4/4) \dots, & (Q-m)=2, m > 0 \\ \pi(0,1) + \pi(1,2) + \pi(2,3) \dots, & (Q-m)=1, m > 0 \\ \pi(0,2) + \pi(1,3) + \pi(2,4) \dots, & (Q-m)=2, m > 0 \\ \text{etc} \dots \end{cases} \quad (6.20c)$$

Case 4 $m = Q$ & $n > R$

$$P(m, n / Q, R) = \begin{cases} \pi(2,1)\Delta(2/1) + \pi(3,2)\Delta(3/2) + \pi(4,3)\Delta(4/3) + \dots, & (n-R)=1, Q > 0 \\ \pi(3,1)\Delta(3/1) + \pi(4,2)\Delta(4/2) + \pi(5,3)\Delta(5/3) + \dots, & (n-R)=2, Q > 0 \\ \pi(2,1)[\Delta(2/1) + \Delta(3/1) + \dots] + \pi(3,2)[\Delta(3/2) + \Delta(4/2) + \dots] + \dots, & (n-R)=1, Q = 0 \\ \pi(3,1)[\Delta(3/1) + \Delta(4/1) + \dots] + \pi(4,2)[\Delta(4/2) + \Delta(5/2) + \dots] + \dots, & (n-R)=2, Q = 0 \\ \text{etc} \dots \end{cases} \quad (6.20d)$$

Case 5 $m > Q$ & $n > R$

$$P(m, n / Q, R) = \begin{cases} \pi(3,1)\Delta(2/1) + \pi(4,2)\Delta(3/2) + \pi(5,3)\Delta(4/3) + \dots, & (n-R)=1, (m-Q)=1 \\ \pi(4,1)\Delta(3/1) + \pi(5,2)\Delta(4/2) + \pi(6,3)\Delta(5/3) + \dots, & (n-R)=2, (m-Q)=1 \\ \pi(4,1)\Delta(2/1) + \pi(5,2)\Delta(3/2) + \pi(6,3)\Delta(4/3) + \dots, & (n-R)=1, (m-Q)=2 \\ \pi(5,1)\Delta(3/1) + \pi(6,2)\Delta(4/2) + \pi(7,3)\Delta(5/3) + \dots, & (n-R)=2, (m-Q)=2 \\ \text{etc} \dots \end{cases} \quad (6.20e)$$

Case 6 $m < Q$ & $n > R$
when $m > 0$

$$P(m, n / Q, R) = \begin{cases} \pi(1,1)\Delta(2/1) + \pi(2,2)\Delta(3/2) + \pi(3,3)\Delta(4/3) + \dots, & (Q-m)=1, (n-R)=1 \\ \pi(2,1)\Delta(3/1) + \pi(3,2)\Delta(4/2) + \pi(4,3)\Delta(5/3) + \dots, & (Q-m)=1, (n-R)=2 \\ \pi(0,1)\Delta(2/1) + \pi(1,2)\Delta(3/2) + \pi(2,3)\Delta(4/3) + \dots, & (Q-m)=2, (n-R)=1 \\ \pi(1,1)\Delta(3/1) + \pi(2,2)\Delta(4/2) + \pi(3,3)\Delta(5/3) + \dots, & (Q-m)=2, (n-R)=2 \\ \text{etc} \dots \end{cases} \quad (6.20f)$$

when $m > 0$

$$P(m, n / Q, R) = \begin{cases} \pi(0,2) + \pi(1,3) + \pi(2,4) + \dots, & (Q-m)=1, (n-R)=1 \\ \pi(0,3) + \pi(1,4) + \pi(2,5) + \dots, & (Q-m)=2, (n-R)=1 \\ \pi(0,3) + \pi(1,4) + \pi(2,5) + \dots, & (Q-m)=1, (n-R)=2 \\ \pi(0,4) + \pi(1,5) + \pi(2,6) + \dots, & (Q-m)=2, (n-R)=2 \\ \text{etc} \dots \end{cases} \quad (6.20g)$$

Case 7 $m=Q$ & $n < R$ & $m=0$

$$P(m, n / Q, R) = \begin{cases} \pi(0,1) + \pi(1,2) + \pi(2,3) + \dots, & (R-n)=1 \\ \pi(0,2) + \pi(1,3) + \pi(2,4) + \dots, & (R-n)=2 \\ \pi(0,3) + \pi(1,4) + \pi(2,5) + \dots, & (R-n)=3 \\ \pi(0,4) + \pi(1,5) + \pi(2,6) + \dots, & (R-n)=4 \\ \text{etc} \dots \end{cases} \quad (6.20h)$$

Case 8 $m < Q$ & $n < R$, $m=0$

$$P(m, n / Q, R) = \sum_i \pi(i, (i+(m-Q)+(n-R))) \quad , \quad i=\{0,1,2,\dots\} \quad (6.20i)$$

After calculating the steady state probability of the number of users in the queue and transmission states, the mean values of throughput and delay can be solve. The throughput is the probability that at least one user is transmitting.

$$S = 1 - \sum_q \Pi(q,0) \quad q=0,1,2,\dots,M \quad (6.21)$$

The from Little's Result, the message delay is calculated as

$$Dm = \frac{\bar{B}}{[M - \bar{B}]\sigma} \quad (6.22)$$

where $\bar{B}$ is calculated the mean number of busy users taken from $P(m, n/Q, R)$ as

$$\bar{B} = \sum_i \sum_j (i+j) \Pi(i,j) \quad i=0 \text{ to } M-j, \quad j=1 \text{ to } N \quad (6.23)$$

Since the message delay have two components, we have to solve one its component first. A simpler approach is solving the transmission delay first as

$$D_t = \frac{L}{N/M_t} \quad (6.24)$$

The ratio N/M_t is actually the average number of slots reserved per user where

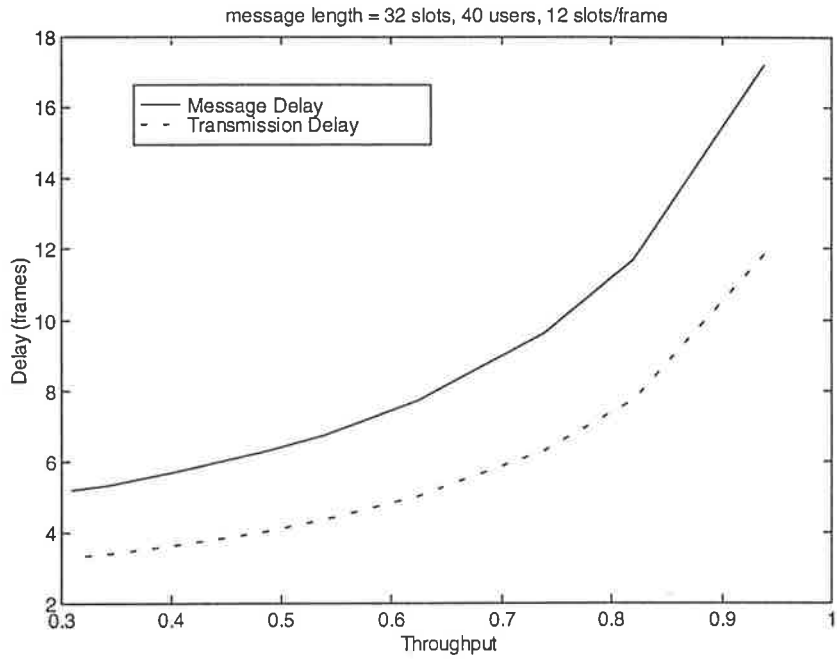
$$\overline{M_t} = \frac{\sum_{i=1}^N i \sum_{q=0}^{M-i} \Pi(q,i)}{1 - \sum_{q=0} \Pi(q,0)} \quad (6.25)$$

Then the queuing delay D_q is the difference of the message and transmission delays.

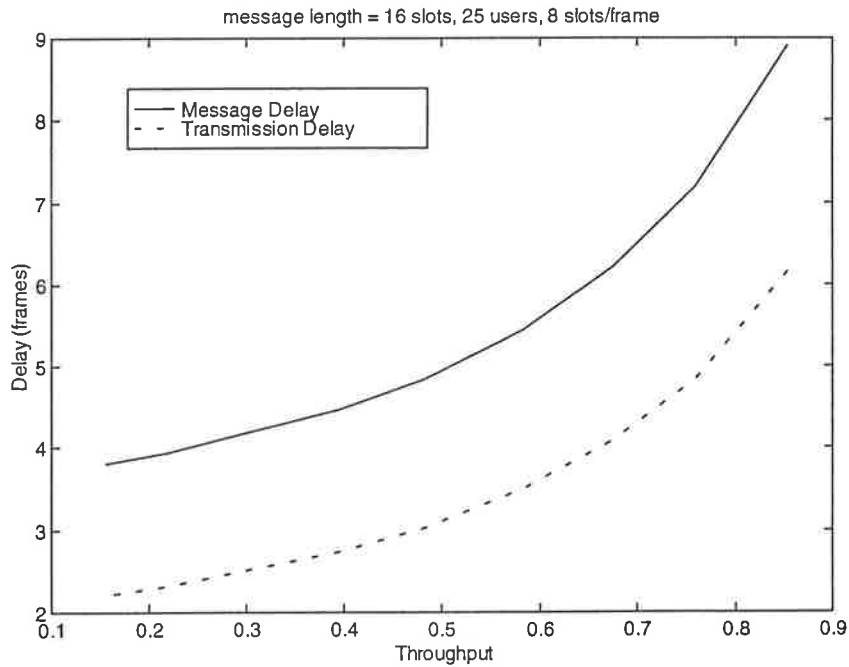
$$D_q = D_m - D_t \quad (6.26)$$

The model is tested using a 12 and 8 slot TDMA for a corresponding message length of 32 and 16 slots. The average sojourn time at the silence state is varied to change the load while maintaining the number of users. Sufficient number of iterations were performed such that the difference between succeeding iterations was less than or equal to 10^{-4} . The transition probabilities that are relatively small were also neglected. To satisfy the condition of total probability, the sum of transition probabilities in each iteration is normalised to 1. The results are plotted in Figure 6.7 and 6.8.

In this section, we described the performance of multislot reservation TDMA system using Markov analysis. As seen in the results, the multislot reservation TDMA achieved very short message delays at low load region. In addition, even in high load region the message delay is smaller than the average message length. This means that every user is always allocated by at more than one slots per frame. This suggests that smaller queuing delays can be achieved compared to a non-multislot scheme. This study also showed that the Markov analysis is a good tool to approximate the system performance of multislot TDMA.



(a)



(b)

Figure 6.7 Throughput/Delay Characteristics of Multislot Reservation

The plots were generated by varying the mean silent (idle) duration in order to obtain the delay at various throughputs. The steady state of the system was obtained using the iteration method until the difference between each iteration is less than 0.0001. The message delay is the sum of the queuing delay and the transmission delay.

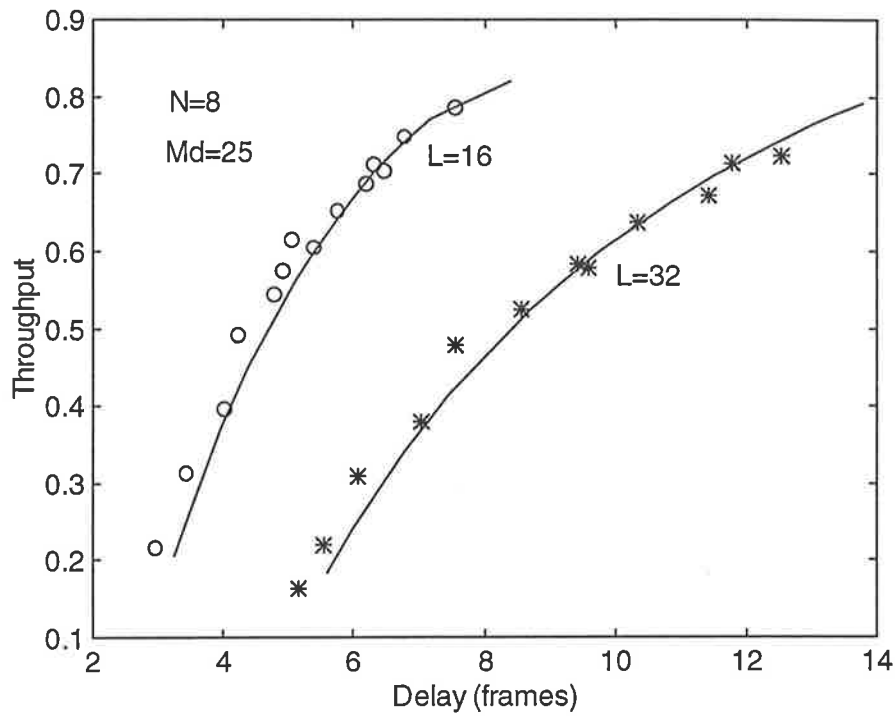


Figure 6.8 Simulation and Analysis Comparison of Multislot TDMA

The method is compared with simulations for two values of L . The results shown are in good agreement. The slight difference is due to the latency which is proportional to the throughput as the simulation increments coincide with the beginning of each frame.

§6.2 Multislot Reservation with Multiclass Users

Prioritisation at the random access level has already been addressed in §5.5 - §5.6 wherein an algorithm based on a stack implementation was proposed. This section introduced a prioritisation at the channel allocation level. the multislot reservation capability as well as the delay tolerance of data services were exploited in order to provide a prioritisation subject to the available resources. The existence of multiple classes of users in the system has an advantage in the design of wireless protocols in the sense that the number of users that a potential high priority user will compete for reservation is reduced into each own class and that of the higher class. However, this requires the low priority users to renegotiate their channel demand as long as their QoS is maintained. Because low priority users are less vulnerable to longer delays (in the case of data services), then a higher overall channel throughput can be attained while maintaining the QoS of each class. This principle is well known as the conventional method of implementing prioritisation. The three most common prioritisation strategies are the First Come First Serve (FCFS) priority queue, Preemptive Resume strategy and the Fixed or Moving Boundary Strategy. A restriction on the sharing of the channels are sometimes implemented in which the system could either use a Complete Sharing (CS), Complete Partitioning (CP), or Partial Sharing and Reservation Scheme [EpsS95].

The effectiveness of the channel allocation scheme highly depends on the amount of information available in each user during the access phase (e.g. channel request). However, the amount of information that can be transmitted by a user depends on the size of the request packet (e.g. random access burst is). Usually, a large portion of the random access burst will be allocated for synchronisation purposes (i.e. training sequence, guard bits, etc.) at the same time employing a more reliable FEC mechanism which significantly reduce the information field. In this paper, we consider only the following information such as: *terminal identity*, *priority*, *message size*, and optionally the an *age field* of the message that is ready to transmit. The class priority of a user is often determined from the terminal identity but the message type is required for the priority of the current message. Some of these information maybe transmitted on the first packet of the first reserved slot if cannot be accommodated into the channel request packet. The age field could be a byte of information depending on the timeout duration during the random access process.

Data services with no delay constraints are restricted only on the random access delay timeout. Thus the large component of the total transmission delay will be the waiting time on the channel allocation queue and the packet transmission time, making the slot allocation procedure simple. For data services with no delay constraints, a simple channel allocation procedure can be as follows: 1) allocate first the terminals in the channel allocation queue then 2) allocate as many slots to other users in the reservation mode giving priority to terminals with more packets to send, and 3) optionally, some slots could also be reserved for the upcoming terminals in order not to lock-out the queue. In this scheme, multislot reservation occurs only during low load condition in which most of the time the average number of terminal transmitting is less than the number of slots per frame. For delay constrained data, it is favourable to allocate multislot to users with large expected message delay (Longest Expected Processing Time First discipline). This reduces the delay variance caused by the random access delay and variable packet length.

§6.2.1 Simple Algorithms for Multislot Systems with Heterogeneous Users

Prioritisation in the resource allocation of TDMA systems may adopt simple algorithms. Some priority schemes for data systems were evaluated in [Buot95a]. In [KarS94], fairness for heterogeneous users was given with much importance. Popular techniques like FCFS prioritised queue, preemptive and non-preemptive strategies and the moving and no boundary resources sharing were evaluated in an ordinary TDMA scheme. The prioritisation in the channel allocation queue can control the access delay of the various classes of users. The concern in this section is the provision of prioritisation to control the transmission delay in each class whereby exploiting the multislot reservation capability. For simplistic algorithms for multipriority systems, the technique previously mentioned can be used. Examples of simple algorithms for the multislot case are as follows:

- FCFS Prioritised Queue / Multislot Reservation Scheme
- FCFS Prioritised Queue / Moving or Fixed Boundary / Multislot Reservation Scheme
- FCFS Prioritised Queue / Class Reservation Scheme / Multislot Reservation Scheme

The FCFS queuing discipline with priority is normally used in order to provide an initial prioritisation for controlling the access delay. In the first algorithm, the multislot reservation will not provide any distinction between classes. Therefore a uniform multislot algorithm applies to all classes. The second algorithm provides a boundary between each class so that the users will compete only to users of the same class. The users compete for multislot allocation. In the third algorithm, a portion of the channel will be reserved to a certain class(es) and some portion will be shared by all classes. Users are now allowed to compete with users from different classes. Thus a prioritised multislot algorithm will comprise of a *queuing discipline*, *resource sharing scheme*, and a *multislot reservation policy*.

§6.2.2 Best Effort Algorithms for Prioritised Multislot Systems

The aim of a prioritised access protocol is to provide an efficient rejection of the lower priority class users during high load condition. In this section, a multi-level prioritised access algorithm is introduced utilising the available parameters of each user. The algorithm is developed to minimise the variance of the overall transmission delay in each class [Buot96b]. Considering a system with N timeslots in a TDMA frame. Every user is capable of multislot reservation and divided into three classes. We then define the parameters used in the algorithm as for terminal i follows:

Symbol	Notation
Dt_i	total expected delay
da_i	total access delay (random access + queue)
dw_i	channel allocation delay
dx_i	transmission delay
Lb_i	remaining packet(s) in the buffer
L_i	message length
r_i	number of slots currently reserved
dt_w	multislot threshold for class k users
n_{bk}	multislot boundary for class k users

Let i the number of the user in the system, $U = \{i = 1, 2, 3, \dots, M\}$ where M is the total number of user. Then $U_c \subseteq U$ which is a set of users in the random access state, $U_Q \subseteq U$ the set of users in the channel allocation state, and $U_R \subseteq U$ the set of users in the reservation state. If the incoming slot is free then the algorithm are as follows:

Algorithm 1 (Base Station Controlled)

In this algorithm the users request only for the first slot to be reserve. Then the base station will broadcast the channel (slot) grant signal to the users for the subsequent slots to be reserved. Users in the queue are given with preference over the users already transmitting. In order for the lower class users to be able to transmit in multislot mode, a multislot threshold is set for each class. All the remaining free slots are utilised by the existing highest priority users. The allocation of multislot in each class is based on the fairness criteria (hungry users first) using the ratio

$$\eta = \frac{\text{number of queued packets}}{\text{number of reserved slots}} \geq K$$

where the parameter K is used to implement a "do not be a pig rule". The procedure of the multislot allocation algorithm is as follows:

1. if $[\exists i \ni i \in U_Q] \wedge [da_i = \max\{da\} \forall (i \in U_Q)] \wedge [\omega_i = 1]$ then
allocate the slot to i , else proceed to step 2.
2. if $[\exists i \ni i \in U_R] \wedge [Dt_i > dt_\omega] \wedge [Lb_i > K_1 r_i] \wedge [\omega_i = 1]$
 $\ni Dt_i = \max\{Dt\} \forall i; K_1 \geq 1$ then allocate the slot to i , else
go to step 3. Dt_i is calculated as $Dt_i = da_i + \frac{Lb_i N}{r_i}$.
3. Repeat Step 1 for $\omega = 2$, else go to step 4
4. if $[\sum r_i < n_{b2} \forall \{i \in U_R\}] \wedge [\exists i \ni i \in U_R] \wedge [Dt_i > dt_\omega]$
 $\wedge [Lb_i > K_2 r_i] \wedge [Dt_i = \max\{Dt\}]; K_2 \geq 1, \omega = 2$, then allocate
the slot to i , else go to step 5.
5. Repeat step 3 & 4 using $\omega = 3$ and n_{b3} for class 3 users, else go
to step 6.

6. if $[\exists i \ni i \in U_R] \wedge \left[\frac{Lb_i}{r_i} = \max \{Lb/r\} \forall i \right], \omega = 1$, then allocate the slot to i , else go to step 7.
7. Repeat step 6 for $\omega=2$ and $\omega=3$.

Algorithm 2 (Mobile Station Initiated)

This algorithm is a slight modification of Algorithm 1 wherein no message length indicator is required during the initial phase of the transmission. However, the user class is indicated in the random access burst. In this case fairness is not guaranteed so that the algorithm allocate slots to old users until they are allocated with the number of slots they initially require. An alternative is also to allocate the slots in round-robin fashion until a user finally receive all the slot it has requested (*Round Robin with Threshold*). Algorithm 2 is as follows:

1. if $[\exists i \ni i \in U_Q] \wedge [dw_i = \max \{dw\} \forall (i \in U_Q)] \wedge [\omega_i = 1]$ then allocate the slot to i , else proceed to step 2.
2. if $[\exists i \ni i \in U_R] \wedge [Dt_i > dt_\omega] \wedge [Lb_i > K_1 r_i] \wedge [\omega_i = 1]$
 $\ni [dw_i = \max \{dw\} \forall i]$ then allocate the slot to i , else proceed to step 3.
3. Repeat Step 1 for $\omega = 2$, else go to step 4.
4. if $[\sum r_i < n_{b2} \forall \{i \in U_R\}] \wedge [\exists i \ni i \in U_R] \wedge [Dt_i > dt_\omega]$
 $\wedge [Lb_i > K_2 r_i] \wedge [dw_i = \max \{dw\}]$; $K_2 \geq 1, \omega = 2$, then allocate the slot to i , else go to step 5. Alternatively, round robin reservation can be used instead of $[dw_i = \max \{dw\}]$.
5. Repeat step 3 & 4 using $\omega = 3$ and n_{b3} for class 3 users, else go to step 6
6. if $[\exists i \ni i \in U_R] \wedge [dx_i = \max \{dx\} \forall i], \omega = 1$, then allocate the slot to i , else go to step 7.
7. Repeat step 6 for $\omega=2$ and $\omega=3$.

Algorithm 3 (Predetermined Allocation)

This algorithm is suitable for multirate systems. However, a user upon random access will indicate the number of slots it require. Then the base station will allocate as much slots as available until the user is fully allocated. This form of best effort policy is also called *Available Bit Rate* scheme (ABR).

1. if $[\exists i \ni i \in U_Q] \wedge [dw_i = \max\{dw\} \forall (i \in U_Q)] \wedge [\omega_i = 1]$ then
allocate the slot to i , else proceed to step 2.
2. Let $\Psi = \{\psi_1, \psi_2, \psi_3, \dots, \psi_{Md}\}$ be the required number of slots
needed for every terminal in the reservation state. ψ_i is
calculated as

$$\psi_{\omega,i} = \begin{cases} \left(\frac{L_i}{v_{\omega}} \right) & \text{if } L_i > v_{\omega} \\ 1 & \text{otherwise} \end{cases}$$
 where v_{ω} is the request threshold parameter
3. if $[\exists i \ni i \in U_R] \wedge [\psi_{\omega,i} > r_i] \wedge [dw_i = \max\{dw\} \forall i] \wedge [\omega = 1]$ then
allocate the slot to i , else proceed to step 4.
4. Repeat Step 1 for $\omega = 2$, else go to step 5.
5. if $[\sum r_i < n_{b2} \forall \{i \in U_R\}] \wedge [\omega = 2]$
 $[\exists i \ni i \in U_R] \wedge [\psi_{\omega,i} > r_i] \wedge [dw_i = \max\{dw\} \forall i]$ then
allocate the slot to i , else go to step 6.
6. Repeat step 4 & 5 using $\omega = 3$ and n_{b3} .

§6.2.3 Simulation Parameters

The effect of the channel allocation algorithm can be demonstrated only if there are many slots in the TDMA frame and the average message length is long. Longer message lengths are also essential in all reservation type protocols to achieve higher throughput. The algorithms were tested in Variable Rate Reservation Access (VRRA) and ATDMA protocols using simulation method. A S-ALOHA without capture random access was employed with a binary exponential feedback contention resolution algorithm. The data message size was exponentially distributed and every message was

segmented into packets where one packet is equivalent to a slot in the frame. There was no limit on the maximum number of packets per reservation. It was also assumed that a terminal contends for reservation only when it has packets stored in its buffer. A single message arrival model was used [Lam80]. For the channel allocation queue, an unlimited buffering capability was assumed. The algorithm was tested using different values of load and the results were taken from equal proportion for all classes.

VRRA is a variant of R-ALOHA but the framing structure of each timeslot allocates a control slot at regular intervals both for the uplink and downlink used for acknowledgment and paging purposes. It also considers the TDMA frame structure for bidirectional traffic by having the control slots paired for the uplink and downlink directions. For a N -slot VRRA frame with control slots interval in every m slots in each timeslot, the VRRA frame repeats every Nxm slots or every m TDMA frames in every timeslot called a block. There are two block structures of VRRA, *rectangular* and *diagonal*. In the simulations, we considered only the diagonal case.

As a brief description of the VRRA protocol, every terminal that has packet(s) to send, listens to the status of each timeslot in the TDMA frame as indicated by the control slots. Once a timeslot is free, it will remain free until the central control allocates a terminal to it. The transmission of the reserved timeslot always start in the slot following the acknowledgment on the same timeslot number. The random access procedure is S-ALOHA with binary exponential back-off contention resolution algorithm, capture and timeout. Once an attempt is successful, it will be allocated with a slot(s). It is also possible that more than one terminal can successfully contend on the successive free slots prior to the acknowledgment. These terminals will be held in the channel allocation queue. Thus, a VRRA is a hybrid protocol as it employs a three-stage reservation with no fixed slots allotted for random access. The random access process occurs only in the free information slots. Details of this protocol is found in [Ham95].

In the simulations 16 slot TDMA frame was used ($N=16$) having $m=9$ leaving 8 information slots for every control slot (see Figure 6.10). In the contention process, the case where successive free slots occur prior to an acknowledgment slot, a terminal can only attempt to access once prior to the acknowledgment except for class 1 terminals where they can retransmit as if the previous attempt was a collision. In the event of collision, a binary exponential back-off was used. A terminal that experienced more than 8 collisions retransmit as a new user.

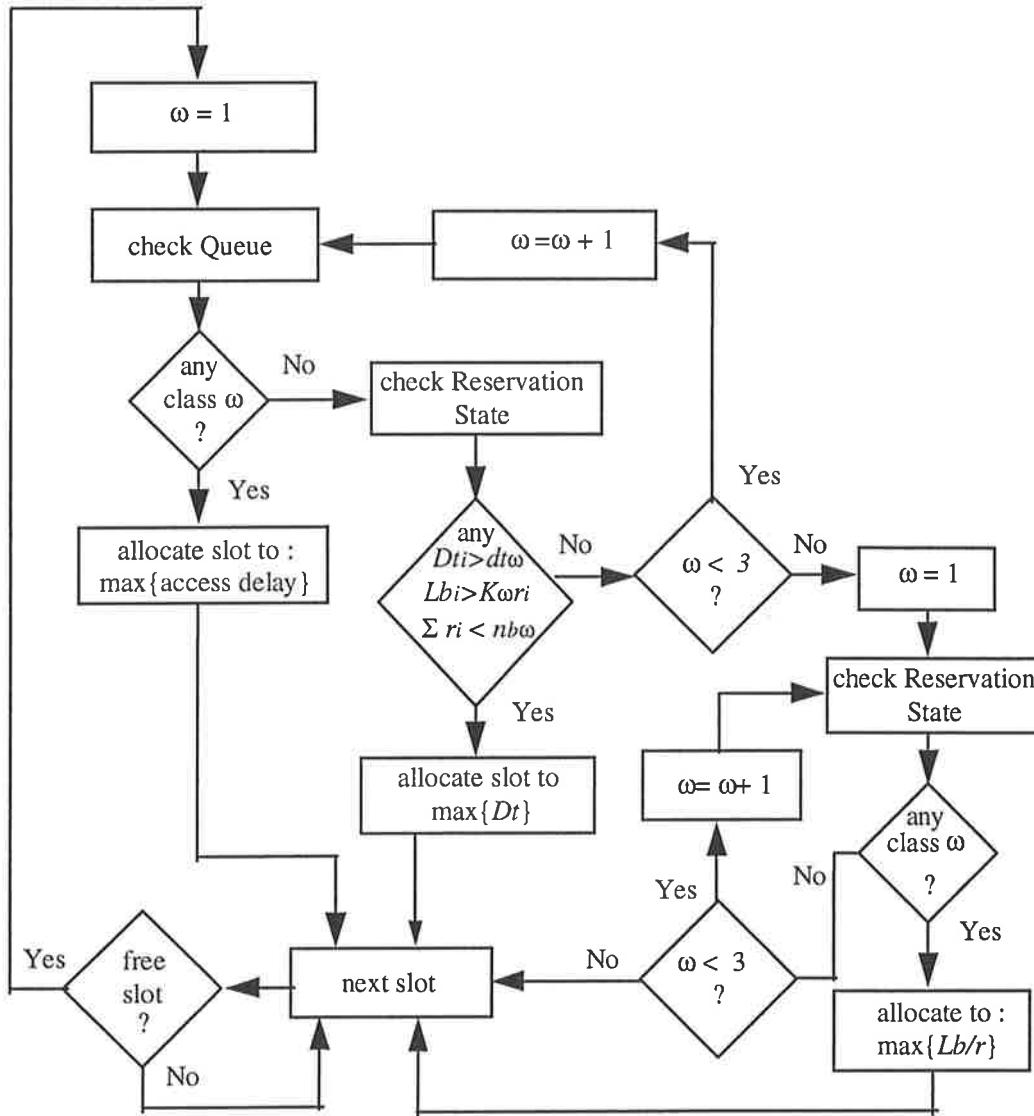


Figure 6.9 Flow Chart of Algorithm 1

		timeslot $\longrightarrow$														
frame number $\downarrow$	Ack0	1,8	2,7	3,6	4,5	5,4	6,3	7,2	Ack8	9,8	10,7	11,6	12,5	13,4	14,3	15,2
	0,1	Ack1	2,8	3,7	4,6	5,5	6,4	7,3	8,1	Ack9	10,8	11,7	12,6	13,5	14,4	15,3
	0,2	1,1	Ack2	3,8	4,7	5,6	6,5	7,4	8,2	9,1	Ack10	11,8	12,7	13,6	14,5	15,4
	0,3	1,2	2,1	Ack3	4,8	5,7	6,6	7,5	8,3	9,2	10,1	Ack11	12,8	13,7	14,6	15,5
	0,4	1,3	2,2	3,1	Ack4	5,8	6,7	7,6	8,4	9,3	10,2	11,1	Ack12	13,8	14,7	15,6
	0,5	1,4	2,3	3,2	4,1	Ack5	6,8	7,7	8,5	9,4	10,3	11,2	12,1	Ack13	14,8	15,7
	0,6	1,5	2,4	3,3	4,2	5,1	Ack6	7,8	8,6	9,5	10,4	11,3	12,2	13,1	Ack14	15,8
	0,7	1,6	2,5	3,4	4,3	5,2	6,1	Ack7	8,7	9,6	10,5	11,4	12,3	13,2	14,1	Ack15
	0,8	1,7	2,6	3,5	4,4	5,3	6,2	7,1	8,8	9,7	10,6	11,5	12,4	13,3	14,2	15,1

Figure 6.10 VRRF Frame Structure (uplink)

§6.2.4 Discussion

The two-class system simulation results in Figure 6.11 and 6.12 shows that the algorithm is very effective in providing priority control both during low load and high load conditions. The growth of the mean message delay is very slow for class 1 traffic as the load increases which is essential for delay sensitive data services. The increase of the delay during high load condition for the class 1 users was mainly attributed by the transmission delay as fewer slots can be allocated per terminal. Further increase in the load can result to single slot reservation depending upon the load proportion. In contrast, the large component of the class 2 delay is attributed by the queuing delay. With this algorithm, the lower class users can be totally eliminated during congestion since the order of channel allocation does not allow the class 2 users to reserve a slot unless the multislot threshold of the higher priority users is attained. The multislot threshold also affect the delay distribution depending upon the load. This characteristic is exhibited in Figure 6.13 showing lower threshold is recommended at lower load conditions. Thus the threshold function as a gate for the lower class users. This value can be varied dynamically by the central control in Algorithm 1. In this way the lower priority users can achieve very low message delays at low load conditions. The throughput of the system also depend on the threshold since the average number of slots reserved per user affects the utilisation. This is because the ratio of the number of information slots to the access slots decreases with more slots per user. It must also be considered for system stability.

In the comparison between the three algorithms (see Figure 6.14) it is shown that Algorithm 1 is slightly better than Algorithm 2. This comparison is taken for the same delay of class 1 traffic. It shows that the delay difference for class 2 and class 3 is noticeable at all throughput values. If looking at the performance of Algorithm 3, its throughput/delay characteristics does not achieve much improvement for class 1 traffic at low load condition. This is because the algorithm limits the number of slots per user to the users' request upon access.

The performance of Algorithm 1 was also tested in ATDMA. In the simulations, better throughput-delay performance was achieved (see Appendix H). This is because ATDMA does not suffer from the reduction between information slots to access slots during multislot reservation as its access slots are fixed. The very low delays at lower loads was achieved since slots can be allocated immediately unlike in VRRM where the

allocation is allowed only right after the acknowledgment slots of each timeslot. The behaviour of multislot ATDMA at lower load is very much the same to Slotted-Idle Signal Multiple Access (S-ISMA) [WuMF94] where only one terminal can transmit at a time. This happens in multislot ATDMA during periods where the interarrival times are much larger than the message transmission time. In fact S-ISMA is advantageous during lower loads. Since the allocation of ATDMA easily allows a single terminal to reserve all free slots in the frame, larger multislot thresholds are necessary.

§6.2.5 Summary

In this section, channel allocation algorithm for multislot reservation TDMA was proposed. The algorithm performed well in VRRM and Advanced TDMA giving good rejections of the lower priority users at high load conditions. The simplicity of the algorithm is attained by having a non-preemptive priority control and the use of few tuning parameters to achieve the required performance. The advantage of such algorithm is the low message delays at lower loads and the rejection of lower priorities at higher loads. In multislot TDMA systems the priority control can easily be implemented if a priority queue is maintained so that arriving terminals can bribe the queue according to its priority and multislot reservation parameters.

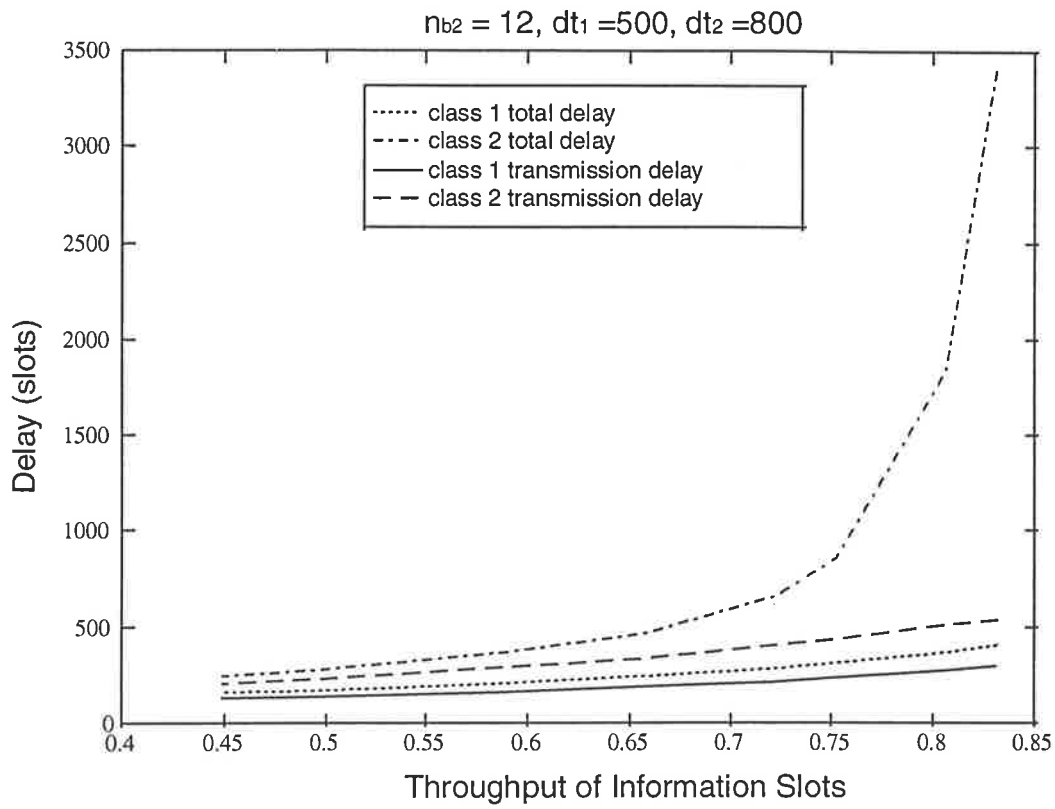


Figure 6.11 Message Delay (VRR)

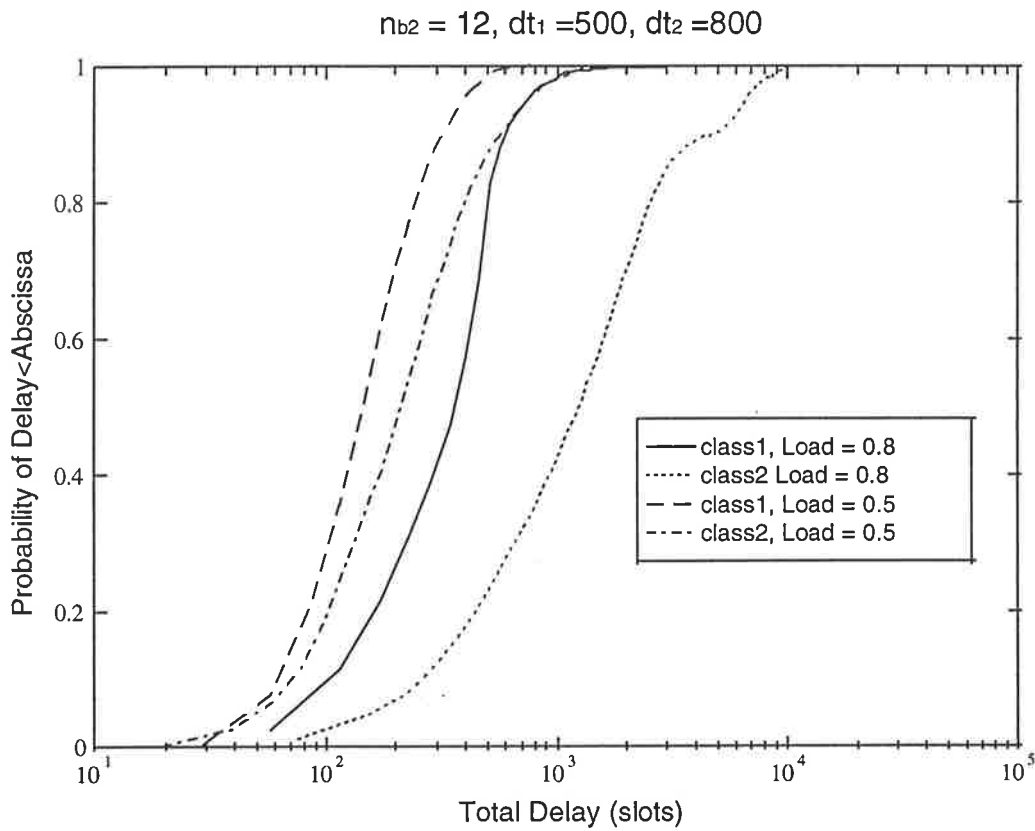


Figure 6.12 Cumulative Delay for Two Priority Systems (VRR)

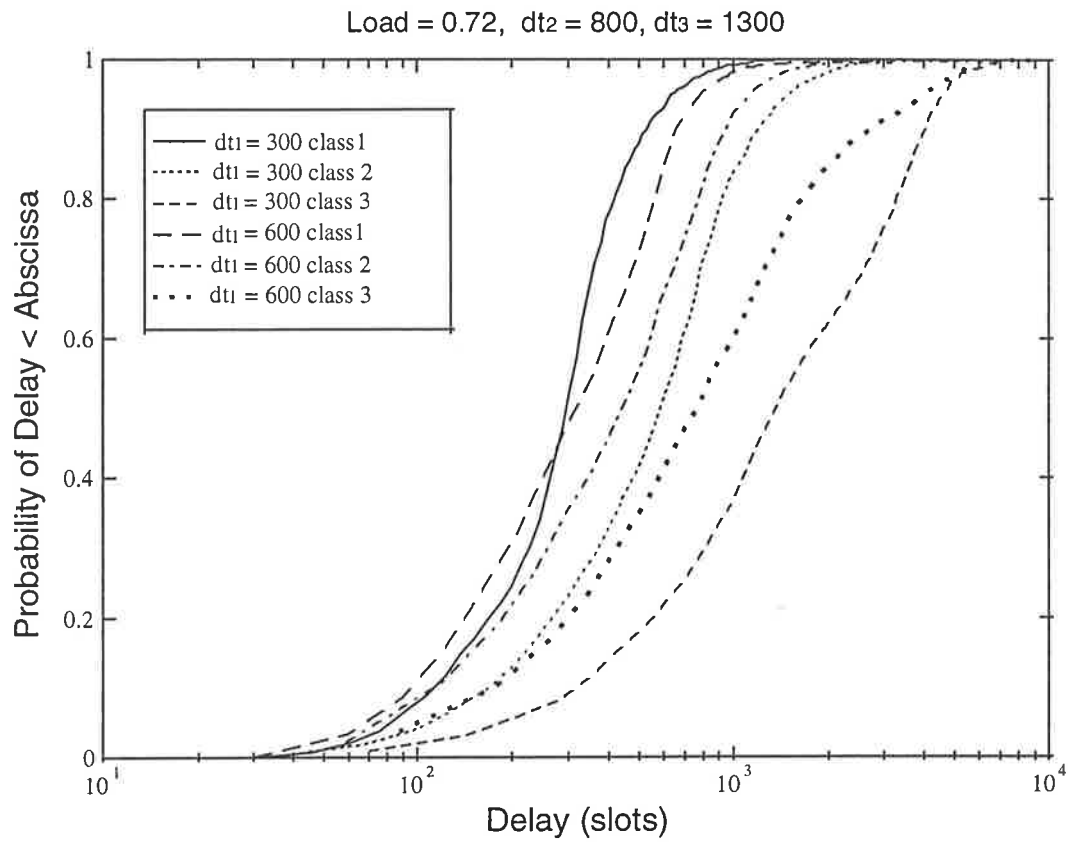


Figure 6.13 Cumulative Delay for Different Multislot Thresholds (VRRM)

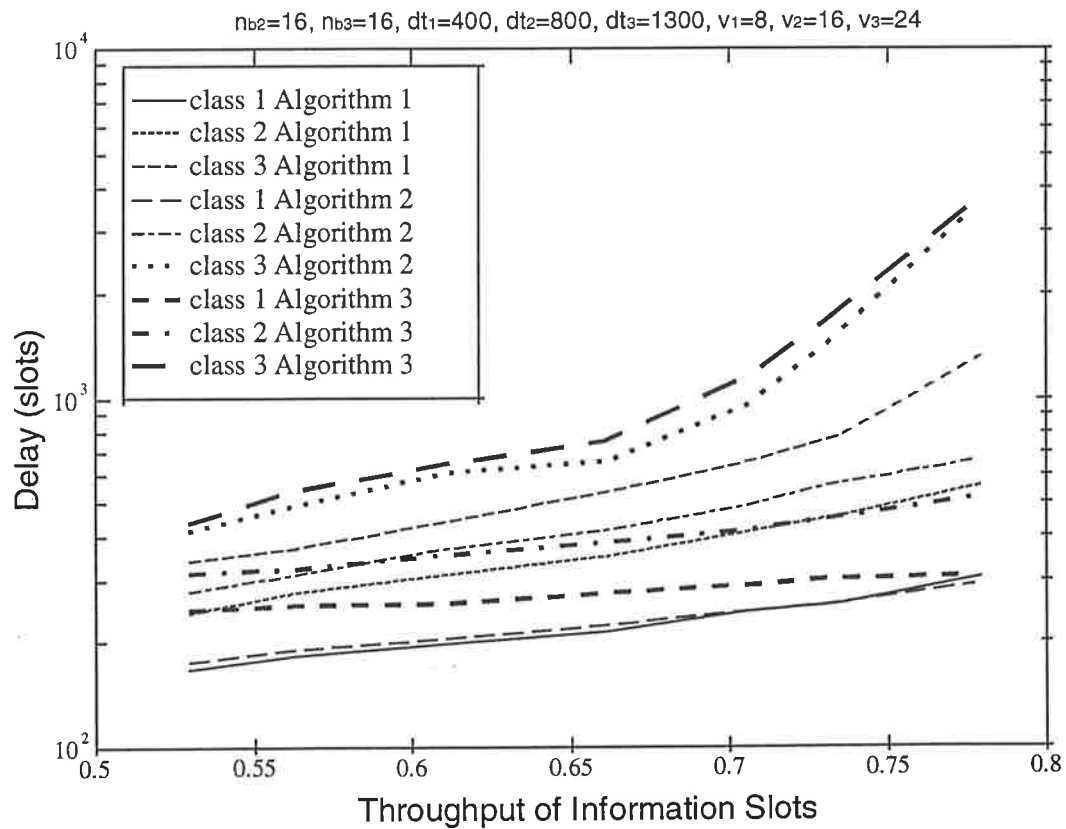


Figure 6.14 Comparison of the Three Algorithms in VRRM

§6.3 Multislot Reservation with Mixed Traffic

Prioritisation in multislot reservation has two types. One is the prioritisation for user class or terminal type discussed in the previous section. The other prioritisation is for the different applications that are supported by the system. The algorithms in §6.2 were designed and tested primarily for multiclass data systems where each class can tolerate certain allowable delay requirements. Thus it was appropriate that the algorithm was tested in a data system such as VRRM. In most cases as in WPC, a mixed traffic scenario often occurs and therefore different resource allocation algorithms will be suitable. In this section, a multislot reservation scheme in a mixed traffic environment is considered.

Here we will describe the reservation policies for the different cases of traffic mix in the radio channel pool in which multislot TDMA is used. We define three possible ways in which multislot data can exist or co-exist in the channel pool. They are the *Narrowband and Wideband Data* system (NB/WB), *Multipriority Data* system (MPD) and the *VBR-Data and Packetised Voice* system (VBRD/PV). The different mixed traffic reservation policies are illustrated in Figure 6.15. The policies considered are based on a reservation scheme [Eps95] and partial sharing of the available resources. The advantage of the following channel allocation schemes is their simplicity as they are all centrally controlled.

§6.3.1 Reservation Policy for Mixed Traffic

The next paragraphs will describe the slot allocation strategies in each of the aforementioned mixed traffic multislot systems.

Narrowband and Wideband Data

This traffic mix requires a reservation scheme since both users have different QoS criteria. Wideband (WB) users are given preferences to narrowband users. However, the NB users can reserve additional channels R_{nm} to meet the QoS criteria. WB users are blocked if there are no available channels in the R_{wb} . Channel Borrowing from R_{nm} to R_{wb} is also possible on a non-pre-emptive manner.

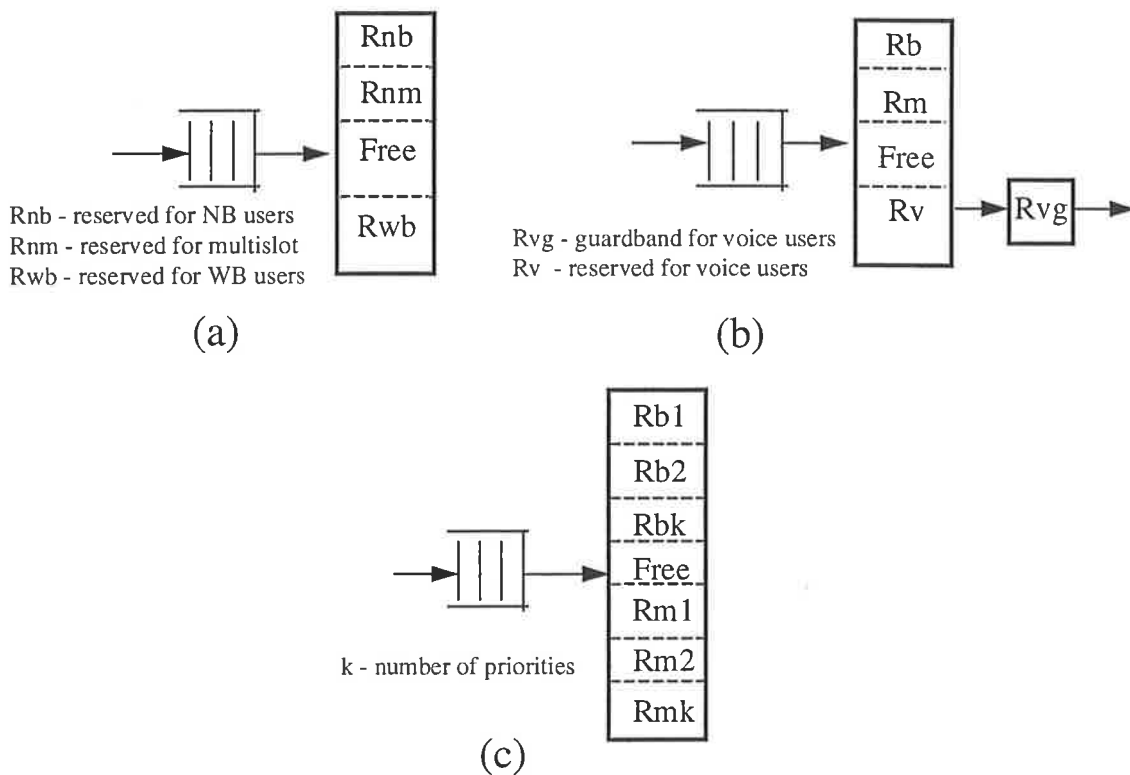


Figure 6.15 Different Traffic Mix in PCS
(a) NB/WB (b) VBRD/PV (c) MPD

VBR Data and Packetised Voice

The statistical multiplexing of speech packets is one of the main criteria of 3rd generation system. To achieve the QoS (percentage of speech packets dropped) criteria, a guardband, R_{vg} is necessary due to the random arrival process. After the voice QoS is met, then the remaining capacity can be used for data. In this case, a boundary for voice and data must be maintained depending on the number of conversations supported. Allowable load for data must also be determined to maintain the required QoS despite of using multi-channel reservation.

Multi-Priority Data

The multi-priority data can be implemented in different ways. The main principle is the same as in the data-only system where multichannel reservation is necessary to meet the QoS criteria. Therefore each class is allotted with its basic rate and multislot reservation. If a complete sharing of the spare channels in the multislot reservation is implemented, threshold type reservation policy is appropriate to maintain the QoS of each class. Higher multiplexing gain can also be achieved compared to a partitioned scheme.

§6.3.2 Simulation Model

The performance of the different traffic mixes were evaluated by simulations in a R-TDMA system. We neglect the MAC protocol in order to isolate the problem on the reservation policies. For the data traffic, a Poisson arrival process was assumed coupled with a negative exponential message length distribution. The WB users were assumed to be having bit rates of 5 times than the basic rate. For the voice traffic, the parameters were referred to [LeeUn86] with a slow activity speech detector and a hangover period of 125 ms resulting to a mean talkspurt duration of 1.1028 seconds and mean silent duration of 1.9556 seconds. The channel is assumed errorless.

For the NB and WB traffic mix, a 50-slot frame was used to clearly demonstrate the performance. A boundary for NB users is implemented with a value of $R_{nb} + R_{nm} \leq 20$. WB users have no boundary. The multislot allocation procedure has an allocation threshold (queued packets/reserved slots) equal to 4. Equal Proportion of input traffic was assumed with average message length of 32 and 160 packets for NB and WB respectively.

For the multi-priority data a single traffic generator was used which generates random classes of messages. The proportion of arrivals were the same for all users. For the stream 1 traffic, a *messaging* type traffic was used with a mean message length of 20 slots. The stream 2 traffic was an *FTP* type with a mean message length of 1000 slots while for the stream 3 message, an *Email* type message of mean message length of 60 slots. Different allocation thresholds and delay thresholds were imposed in each class. In this case the channels were completely shared. The channel has 32 slots per frame.

For voice and data system, a 32-slot frame was used in consonance with the RACE macrocell assumptions. A frame is assumed 10 ms duration for mapping the speech packets. If the speech packet is not transmitted within one frame, it is immediately dropped. A boundary of 15 slots for data was imposed leaving more post-reservation slots for voice. A non-greedy allocation threshold of 12 was used for data.

§6.3.3 *Simulation Results and Observations*

Before the performance of the mixed traffic was evaluated, a multislot data with single class and multiclass users was considered. This simulation differs from that of §6.2 since this time the concern was with the QoS criteria based on the maximum delay allowable. In the data system, the performance was taken at different average message size. Delay thresholds were imposed regarding of the message size. This is necessary in order to determined how effective the multislot reservation is. Longer delay timeout or thresholds were imposed on the lower class. Since the QoS is defined in terms of the delay percentile, the 95, 90 and 80 percentile mark were used. In Figure 5.16, the results shows a linear relationship between the maximum load with the average message size for a single user system. In the case of multipriority (3-class system), the average message size is fixed at 32 slots and the probability of delay exceeding the threshold is each class was determined. The results are plotted in Figure 6.17 showing the QoS for each class was maintained. This was achieved through properly tuning the system.

The performance of a mixed traffic data system was then evaluated. This time, we only used the throughput-delay criteria since different applications have different QoS parameters (but a function of the delay). Two schemes were used which are the threshold type multislot reservation and the threshold type multislot with greedy algorithm. The threshold type algorithm imposed a threshold (message size/number of reserved slots) of 15, 250 and 20 for type 1, type 2 and type 3 respectively while the greedy algorithm just allocate extra slots according to the largest expected transmission time first. The throughput-delay performance are plotted in Figure 6.18 and Figure 6.19. A difference in the performance can be noticed between the two schemes especially for the FTP traffic. This suggests that various control mechanisms can be implemented in providing prioritisation in a mixed data traffic environment. Thus the parameter are needed to be optimised by a centralised controller.

For the WN and NB system, comparable performance for both traffic types can be achieved (see Figure 6.20). This is attained by using a moving boundary between for NB users or by varying the allocation threshold. The use of higher priority for WB users is due to its blocked calls cleared assumption. Even if the WB users has no boundary, it has only little effect to the data users since the multislot reservation can compensate after WB users free the borrowed slots. Also, WB user has a very small probability of borrowing channels from the NB users if QoS is maintained. This scheme

is a form of pre-reservation for WB users. The same scheme is actually used for the voice/data system as a QoS maintenance for voice, (see Figure 6.21). The voice users eventually implement a post reservation [Rap91] and pre-reservation scheme as R_{vg} is allocated as well as no boundary is imposed after consuming all the R_{vg} slots.

Another important aspect for voice quality is the short term speech packet loss probability which was noticed in the results. If it is assumed that a severe packet loss in a duration of 3 seconds is noticeable by a user, then speech packet multiplexing can suffer from this behaviour. As seen in Figure 6.22, even if the average speech loss is very low, large spikes of packet dropping rates were observed. The plotted short term speech packet loss in Figure 6.22 is for a voice only system. It is shown that even for very low average loss of 0.74 percent (1 percent is tolerable), a 5 percent short term loss was observed. However, in voice/data system, the speech loss is evenly distributed (see Figure 6.23). Short term severe speech packet loss was then avoided as data users can back-off during voice congestion even without preemption as the average service time of data users is decreased by the multislot reservation. This suggests that a partitioning of the channels between voice and data is not attractive as considerable amount of guard band is required for voice to minimise the occurrence of short term severe packet losses. Although statistically, it cannot be avoided, the periodicity of the severe packet losses depends on the traffic load. However, it is not so clear if this the 3-seconds short term speech packet loss is very significant in identifying the MOS. Experimentation is recommended to evaluate such behaviour.

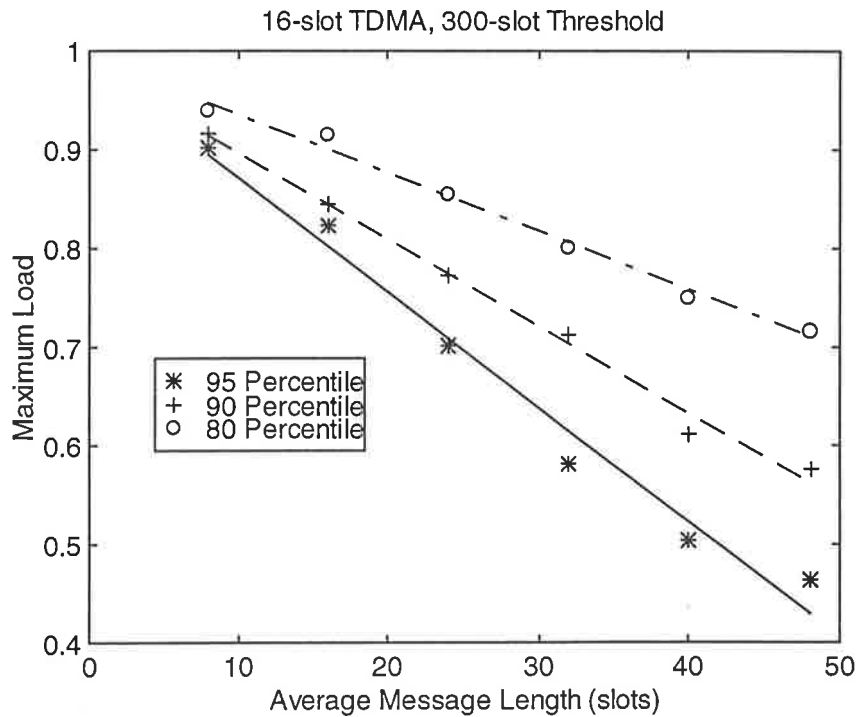


Figure 6.16 Linear Fit of the Maximum Load in ATDMA

The fitted curve is based on the set of data tested for the various average message length. The reduction in the maximum load is necessary so that the multislot reservation can take place. Since there is a limit on number of channels per frame, the delay threshold must be adjusted to the actual message length.

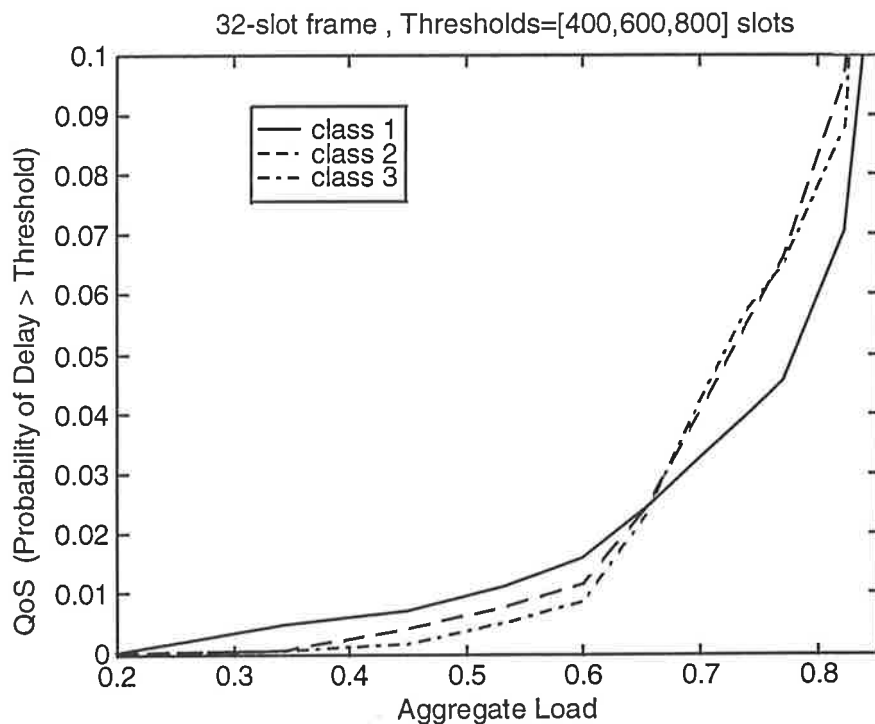


Figure 6.17 Performance of 3-Class Multislot Data Uniform Users

Average Message Length = 32 slots

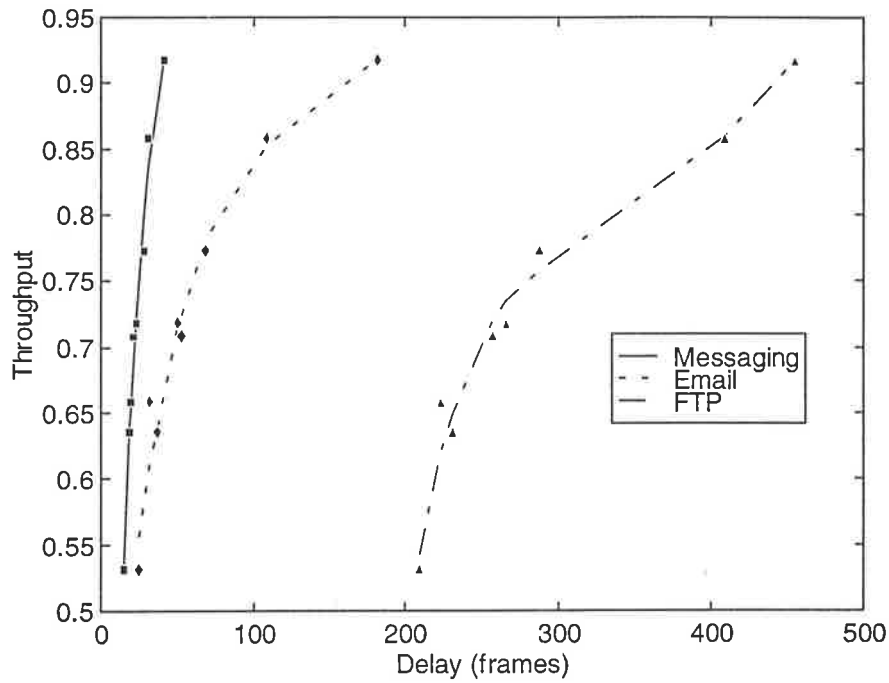


Figure 6.18 Throughput/Delay of Mixed Data Traffic

The number of slots per frame is 32 and the ave. message sizes are 20,60 and 1000 for messaging, email and FTP with multislot threshold of 151, 20 and 250 respectively. The large threshold for FTP resulted in a large average delay.

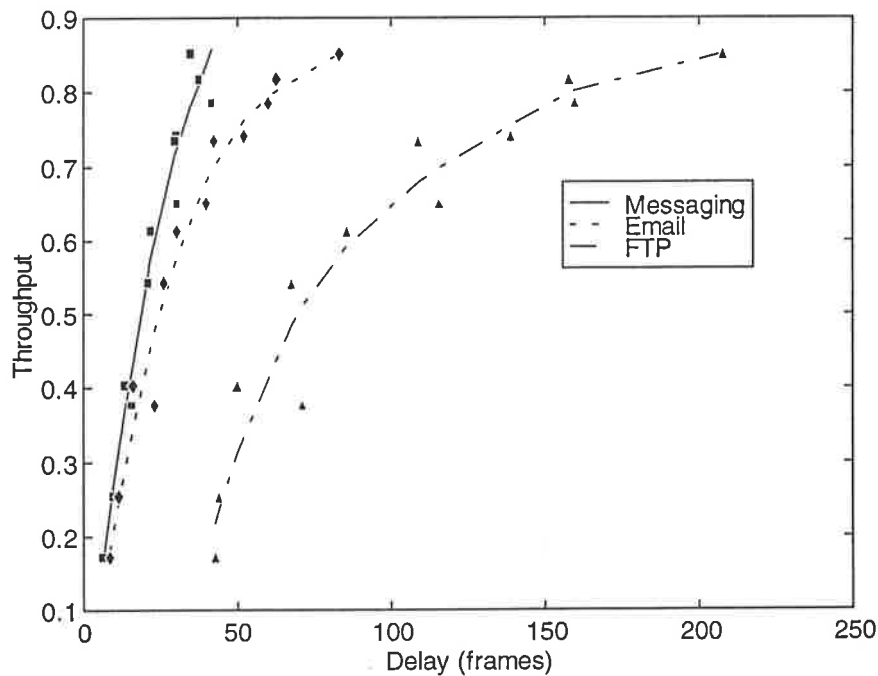


Figure 6.19 Throughput/Delay of Mixed Data Traffic

Results taken from a greedy multislot allocation (best effort). This time the FTP type can reserved of up to a total of 24 slots leaving fewer slots for email and messaging. As a result, the FTP average delay is reduced significantly.

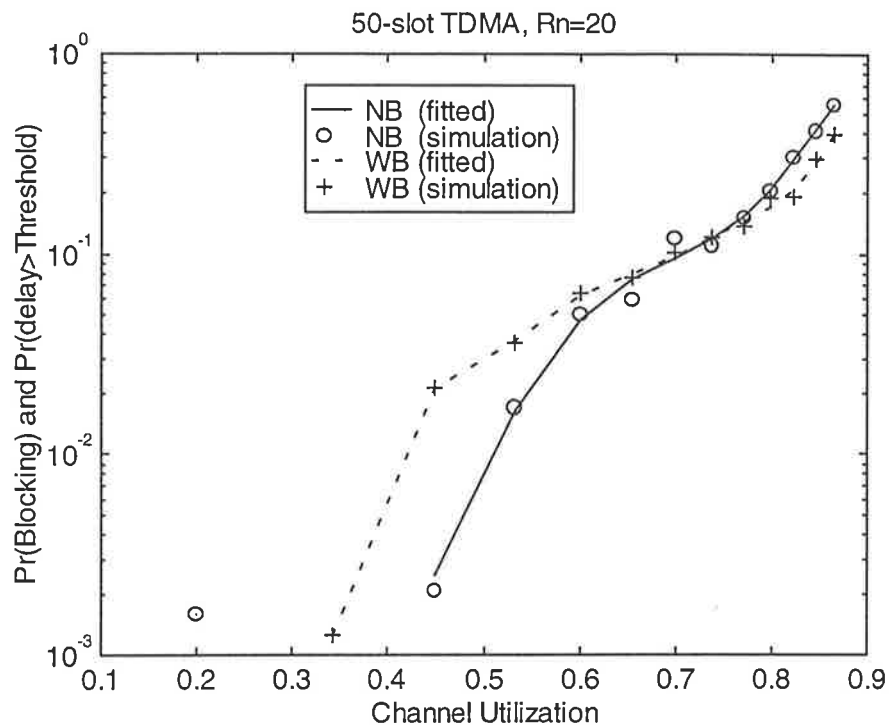


Figure 6.20 QoS Performance of NB and WB users

The QoS criteria for WB users is in terms of the blocking probability while the NB is in terms of delay percentile. NB employed multislot with a maximum number of reserved slots per user of 4 and the delay threshold is 1000 slots. The figure shows the difficulty of accommodating WB at higher loads.

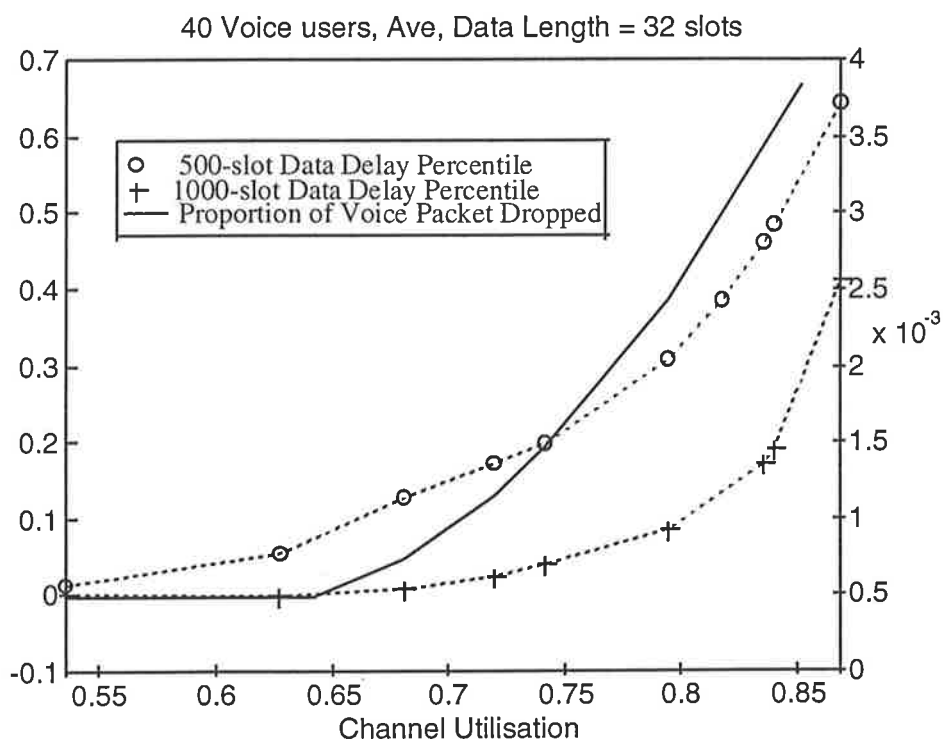


Figure 6.21 Plot of Voice/Data Integrated System

Probability of Data Delay > Threshold (left) and Proportion of Speech Packets Dropped (right)

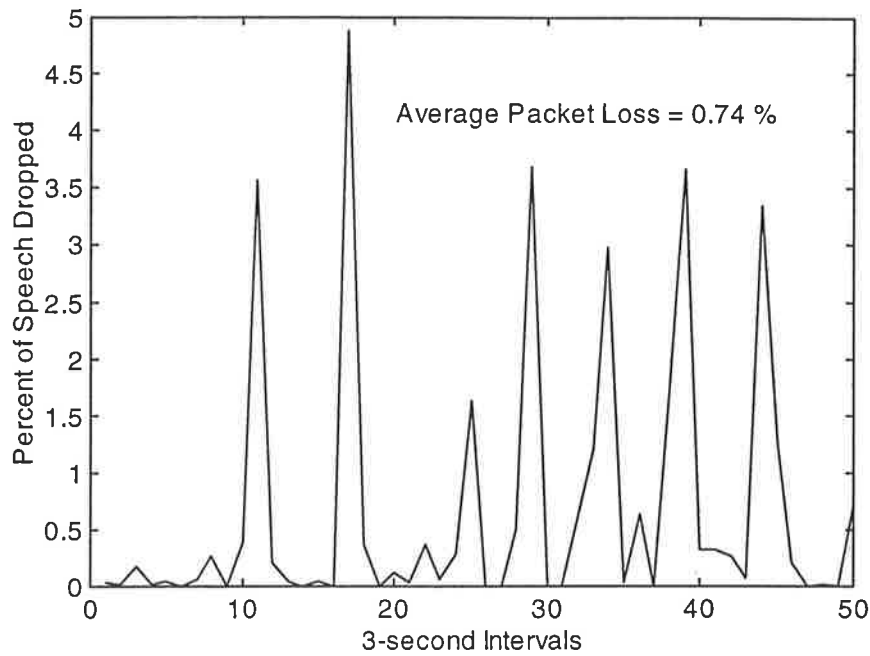


Figure 6.22 Short Term Speech Quality Measure of a Voice-Only System
Results were taken from a 32-slot TDMA frame with 70 users.

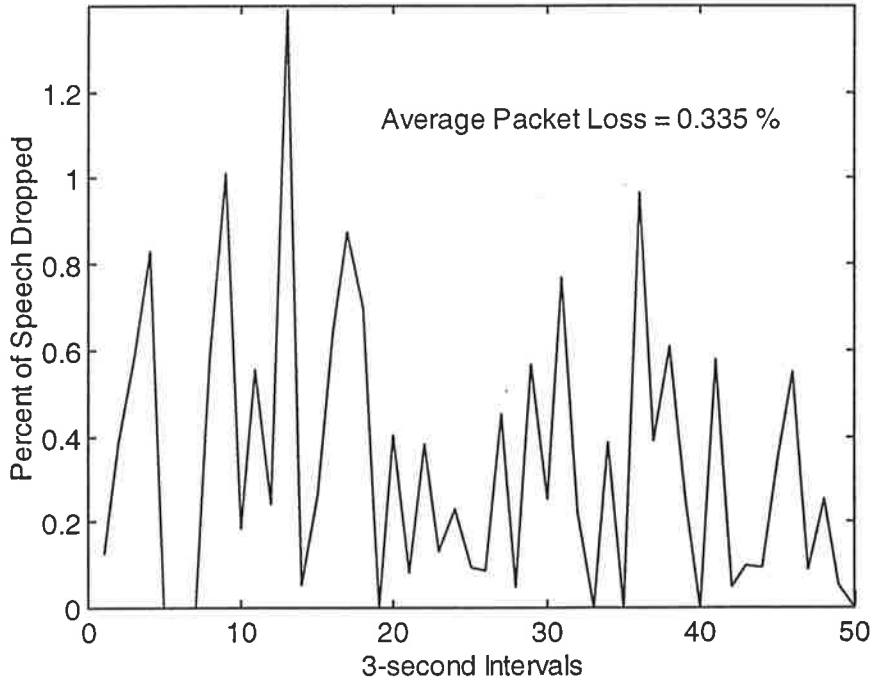


Figure 6.23 Short Term Speech Quality Measure of a Voice/Data System
The number of users was reduced to 40 user and a 40 percent data of the channel load was attributed by data. The data delay >1000 slots was measured at 21 percent (very high load). The QoS degradation of the system was transferred to the data users.

§6.4 Variable Coding Rate Multislot Multimedia System

As mentioned earlier, resource allocation is one of the problems in multimedia system in which an increase in the flexibility of the protocol is managed by a central resource allocator. For a Linear Time Invariant (LTI) channel where the capacity of the channel is optimised by a selected physical layer configuration, the multislot capability is the main component that can be exploited to manage the QoS of different services. It is then assumed that every channel has equal capacity which simplifies the unit of resources as channels (e.g. slots for TDMA). For a time varying channel, the capacity is optimised by selecting appropriate coding rates and/or transmission power to each terminal. With the limited power control^{*} in most wireless access technologies, the selection of appropriate coding rates becomes more important to maximise the channel capacity. Eventually, a variable coding rate system can dramatically affect the delay performance of every traffic type. For a single slot reservation scheme, a trade-off between error retransmissions and channel speed suggests an optimisation problem in minimising the transmission delay. However, the problem of delay performance degradation caused by the time varying channel can be overcome by an appropriate resource allocation. That is, to use a multislot reservation scheme.

Again, one of the objectives in a multimedia access protocol is to improve the message delivery performance of asynchronous services. In a LTI channel this is a relatively easier task compared to a time varying channel where a wrong choice of coding rate in the latter will eventually result in massive packet errors. The fragility of a wireless link is also affected by the channel overloading but conversely a waste of resources will result if a channel is lightly loaded. The waste of resources is detrimental to multislot reservation scheme because the QoS is achieved by allocating extra capacity to hungry users.

In the massive packet error scenario, the delay of the erroneous packet also depends on the backward error correction scheme as well as the back-off time interval. When the channel quality is highly autocorrelated, it is might as well necessary to implement a longer back-off time until the channel recovers from bad error performance.

^{*} Power control is used in wireless systems to increase/decrease the transmission power in order to maintain the required signal to noise ratio. However, there is a limitation on the power control since the frequencies in a cellular environment are reused at a certain distance hence, causing interference.

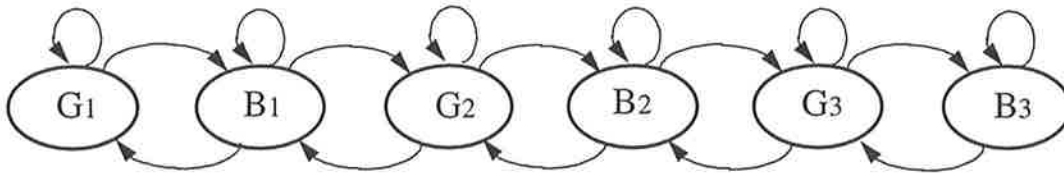


Figure 6.24 Fritchman's Model for Time Varying Channel
Three Good and Three Bad States Configuration

Since the random back-of time will introduce longer delays, then the use of variable coding rate becomes essential. By lowering the coding rate but allocating more channels, the problem of transmission delay can be addressed. This idea is applicable in the wireless environment since the users are dispersed in a geographical area where the individual channel qualities are independent.

§6.4.1 Channel Model

The modelling of the instantaneous bit errors is necessary in order to evaluate the performance of the system. In this case a slow fading channel is required in order to demonstrate a long term channel variation that affects the packet error probabilities. For simplicity, a Fritchman's model [Fritch67] was adopted in which the channel states at the bit level are alternates of good and bad states. In this case the average run lengths of erroneous and correct bits are dependent on the state of the channel. An illustration of the Fritchman's Model is shown in Figure 6.24. The transition probabilities are chosen such that the effect of long term fading are pronounced as a result of users mobility and power control limitation.

The other consideration for the channel quality is the power control which is usually employed in a mobile channels bringing the effective *BER* at a predefined level. However, WPC based on R-TDMA cellular systems have limitations on power control because a dramatic increase in transmission power of one mobile terminal in one cell means an increase in co-channel interference to the neighbouring cells using the same frequency group. This is the main reason why power control alone is not enough to combat the problems of a noisy channel. To incorporate its behaviour in the Fritchman's Model a careful selection of the transition probabilities is required. In the channel modelling, a $[g,b]$ channel which represents g good states and b bad states. For example, in a $[3,3]$ channel with states G_1 and B_1 represents a normal channel condition while going states G_3 and B_3 means an increase in the average

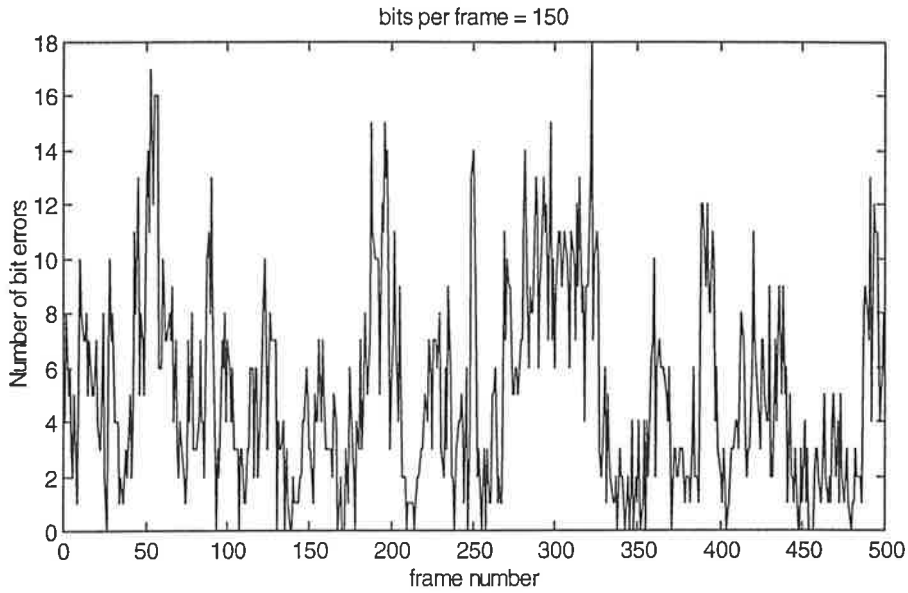


Figure 6.25 Sample Channel Histogram of a [4,4] Fritchman's Model

bit error rate. The power control tends to push the channel state back towards G_l and B_l . Therefore a practical channel suggest some transition probabilities where the channel state is always pushed towards G_l and B_l as the occurrence at the worse states is less frequent. In this case we can model a more practical channel with limited power control characteristics. A sample channel histogram for some chosen transition probabilities is shown in the figure above.

§6.4.2 Channel Coding

The coding scheme can be properly chosen once the channel characteristics is known. For simplicity of calculation, we adapt a simple assumption by choosing a fixed frame length of a certain number of coded information bits. Every coding rate has a corresponding number of correctible errors per frame. The selection of the coding rate depends on the estimated channel quality. Therefore, the scheme rely on the reported signal measurements. The channel estimation takes a window of E_w frames where the average bit error rate is measured. From the measured channel quality, a coding rate is selected by introducing an error margin, E_m . The measurements of the reverse channel are taken by the base station which is responsible for the decision of which coding scheme is to be used. The calculated coding scheme is then reported to the mobile station via a feedback channel (i.e. downlink pair or associated channel). Similarly, the implementation on the forward channel requires the mobile station to calculate the best

coding scheme and is transmitted to the base station via the reverse control channel periodically (not as frequent as for the reverse case).

From the error cluster distribution, the maximum channel capacity can be obtained based on the selected coding rates supported. The maximum capacity is simply calculated as

$$\text{Capacity for a given coding rate, } C_{ch} = \sum_i Cr(i) Pe(i) \quad (6.27)$$

where: $Pe(i)$ = probability of i errors per TDMA burst

$Cr(i)$ = maximum coding rate to combat i errors

However, the actual capacity of the channel can be obtained based on the margin and measurement window parameters. To obtain the capacity, a channel is firstly simulated taking the number of errors per burst. Then based on the chosen window and margin, the number of wrong error predictions are counted. This parameter is known as the packet or burst error probability. After it is obtained, the capacity is calculated as

$$C_{ch}(w, m) = \sum_i Cr_m(i) Pe_{w,m}(i) [1 - E_p(w, m)] \quad (6.28)$$

where E_p is the packet error probability and Cr_m is the coding rate for a chosen window and margin. The optimal parameters can be selected by maximising the capacity.

§6.4.3 Simulation Model

The Variable Coding Rate Variable Bit Rate (VCR/VBR) scheme was tested in a 16-slot R-TDMA system. Messages were generated with Poisson arrivals where every message is assumed to be owned by different users. Slots were allocated on FCFS bases. Every user must own a slot first before it can request for additional slots. The request packet for additional slots was transmitted on the first reserved slot (e.g. in a form of a request flag). The requests for additional packets are held in a request queue and is treated with lower priority than the first slot allocation queue. A user cannot make another request until the current request is granted. The coding rate is independent to each user and uniform coding rate is used to all slots owned by one user. Every message is composed of one or several packets and is exponentially distributed in

length. Once a user has reserved one or more slots, it owns the slot until all packets are transmitted. In the event that there are packet errors, it was assumed that an error is retransmitted immediately (Immediate Error Recovery).

Four coding rates were supported by the system. They are 3/4, 2/3, 1/2 and 1/3 rates. The FEC capacity of each rate can correct up to 6, 8, 12 and 16 errors respectively. The arrival process is Poisson and the message length distribution is exponential. Multislot reservation was based on the threshold policy where the threshold parameter is the ratio of the remaining packet size to the total transmission rate (coding rate multiplied by the number of slots reserved). To emphasise the effect of multislot and variable coding rate, the channel load is only 0.5 packet/slot unless specified. Every TDMA frame (in one slot) carries 150 bits. The channel transition parameters are as follows:

Model 1:

$$P_x = \begin{bmatrix} .98 & .02 & & & & & & & \\ & .94 & .02 & .04 & & & & & 0 \\ & & .002 & .94 & .058 & & & & \\ & & & .94 & .04 & .02 & & & \\ & & & & .003 & .90 & .097 & & \\ & & & 0 & & .92 & .05 & .03 & \\ & & & & & & .005 & .88 & .115 \\ & & & & & & & .90 & .10 \end{bmatrix}$$

Model 2:

$$P_x = \begin{bmatrix} .96 & .04 & & & & & & & \\ & .92 & .04 & .04 & & & & & 0 \\ & & .004 & .93 & .066 & & & & \\ & & & .92 & .06 & .02 & & & \\ & & & & .003 & .90 & .097 & & \\ & & & 0 & & .90 & .07 & .03 & \\ & & & & & & .005 & .86 & .135 \\ & & & & & & & .80 & .20 \end{bmatrix}$$

where the row represents the current state and the column for the next state.

The channel characteristics is plotted in Figure 6.26 showing the PDF of the number of bit errors in a TDMA frame. It shows the time varying characteristics of the channel which confirmed that a Fritchman's Model is capable of modelling such scenario. The capacity of the channel in Model 1 is firstly determined in order to show

the effect of the measurement window and error margin. It shows that there is an optimal combination of both parameters to maximise capacity (see Figure 6.27 - 6.28). Since the maximum capacity does not correspond to the minimum packet error probability, this means that for a single slot system, the delay budget must consider the trade-off between queuing delay and retransmission delay. High data rates would mean a shorter transmission delay but having a higher probability of retransmission. The variable coding rate with multislot can easily compensate the lower data rate by the use of more channels. The results are plotted in Figure 6.29 - 6.31.

§6.4.4 Summary

This study had combined capacity optimisation with the resource allocation. The use of variable coding rate in a multiaccess system can compensate the capacity variation of the channel. This scenario is typical in wireless multimedia where QoS is only met by an efficient resource allocation scheme. Optimal tuning involves E_m , E_w and the resulting delay. The effect of using a variable coding rate minimise the use of power control and has an advantage in the power control management of the channel (i.e. looser control). Power control can attain a constant coding rate but can affect the entire system. In contrary, variable coding rate can maintain the system interference but can affect the QoS of the service. This problem is overcome by multislot reservation in order to average the capacity loss as a result of a time varying channel.

The parameters E_m and E_w were found to be sufficient in minimising the packet error probability. The results showed that larger margin can decrease the FER . But larger margin would mean lower capacity due to lower coding rate. However, when delay is considered, a slightly different optimal operating points that are close to the static optimisation were obtained. This is due to the channel loading that is only at 0.5 packets per slot. Which means, at this load, there are always excess capacity making a static optimisation slightly different from that of the actual performance.

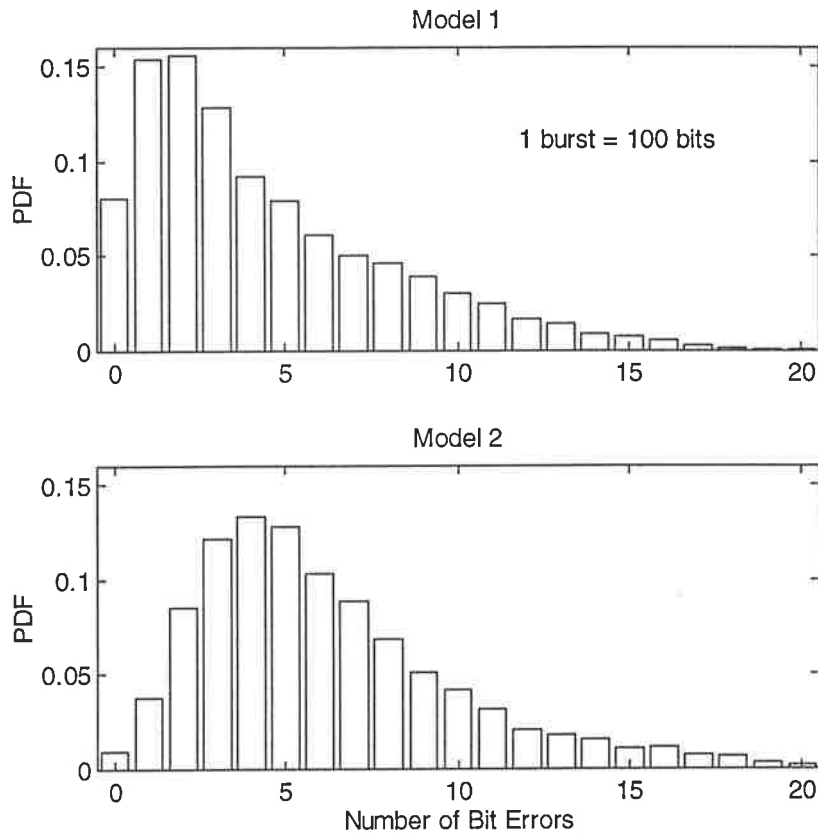


Figure 6.26 Error Distribution

The channel model has a typical error run distribution which is Rayleigh-like. The variation of the parameters can practically obtain most of the channel characteristics.

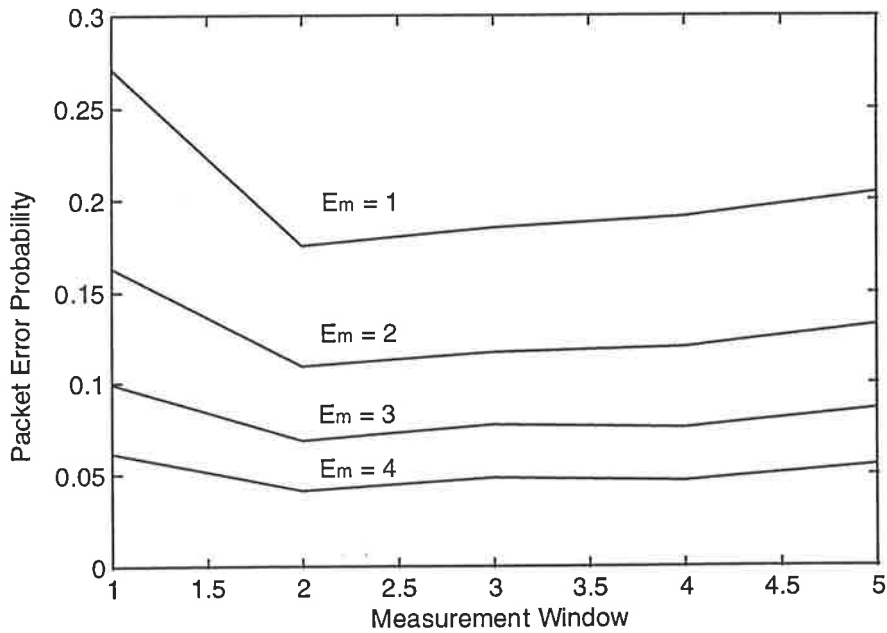


Figure 6.27 Packet Error Probability due to Wrong Prediction

A simulation result of the channel in Model 1 by varying the measurement window, E_w .

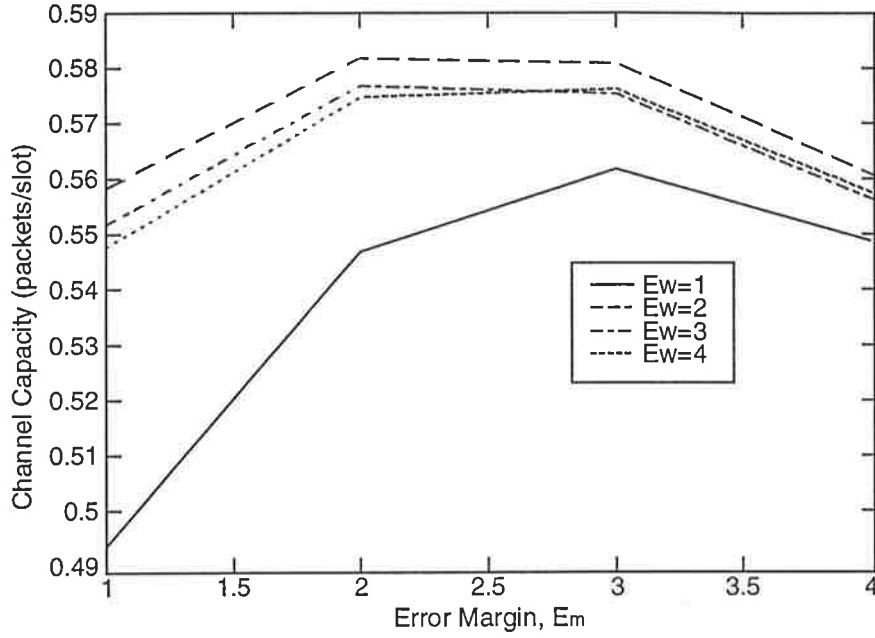


Figure 6.28 Capacity for Various Control Parameters

Model 1 simulation result for the capacity after using coding rate switching based on Eq. 6.28. It shows that optimal channel capacity requires proper tuning of the system.

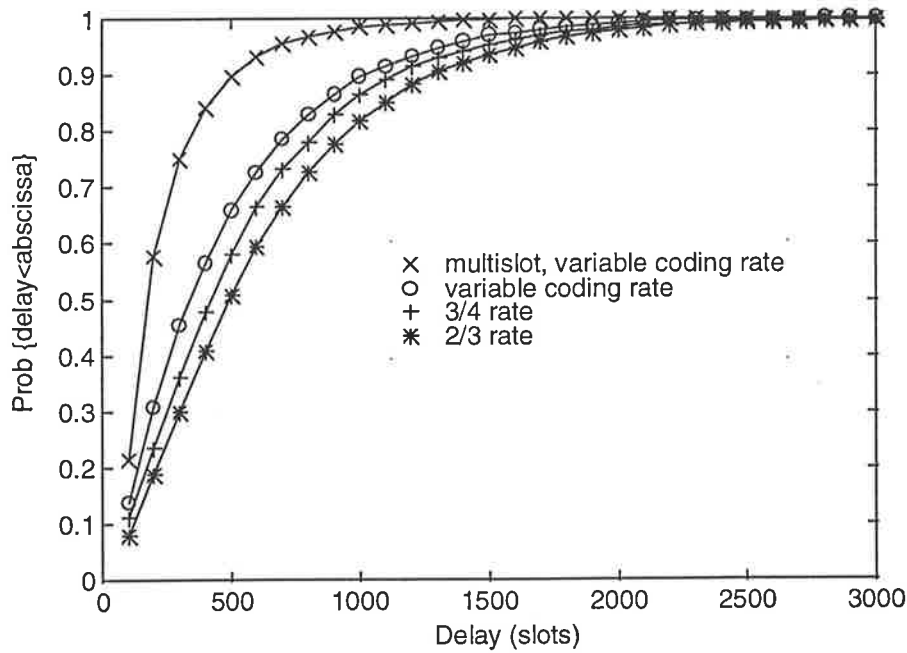


Figure 6.29 Comparative Performance of Fixed Rate, Variable Rate and Variable Rate with Multislot

The simulation was conducted using $L=16$, and channel Model 1. It shows that even for the single slot reservation, the variable coding rate can achieve better performance. Multislot threshold for is 16. ($E_w=3$, $E_m=2$, $N=16$).

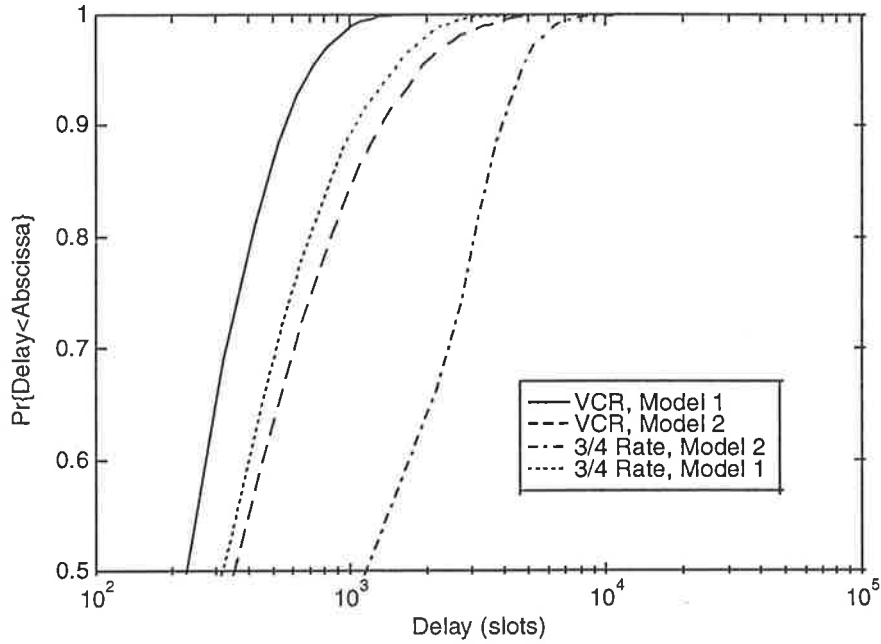


Figure 6.30 Comparison of VCR and Fixed Rate

Much difference can be noticed when channel Model 2 was used due to severe bit error rate. The figure shows that even for a higher multislot threshold value of 16, VCR performance is much better than the fixed coding rate of 3/4. ($L=32$, $E_w=3$, $E_m=2$, $N=16$).

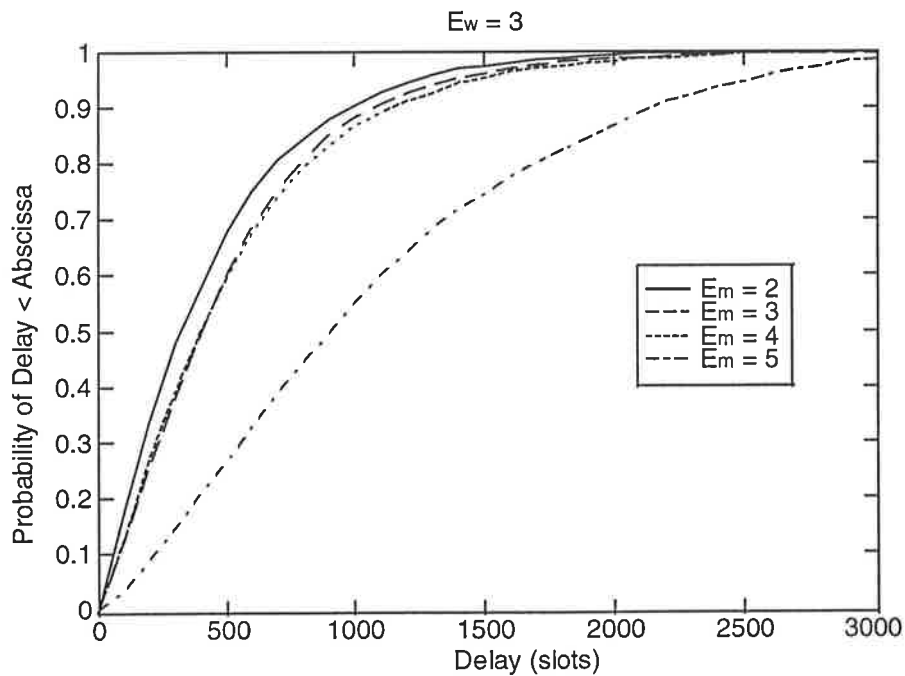


Figure 6.31 Multislot Performance at Various Error Margin

Channel : Model 1, ($L=32$, $E_w=3$, $N=16$).

§6.5 QoS Maintenance for Wireless Video Transmission

Perhaps the most difficult type of application that R-TDMA will support is a viable rate video. In the case where video is multiplexed with voice and/or data, the problem of QoS maintenance is inevitable. In situations like this, voice and video is always treated with higher priority because of the assumption that data services can tolerate a relatively longer delays to that of voice and video. Since video traffic varies significantly (i.e. VBR video), the QoS for voice and/or data will eventually be affected. However, if we compare the number of voice and data users to the video users, it is logical that we will treat video as a lower priority service. Consequently, its quality cannot be guaranteed so that the best achievable video quality will be the subject of investigation in this section. The technique of using a channel load feedback to the video encoder will be carefully examined. We model the system with a single video source multiplexed with data users in a R-TDMA protocol. Again, the multislot reservation capability will be used.

§6.5.1 *System Model*

The most challenging traffic type to be supported is the *variable bit rate* (VBR) video due to its strict Quality of Service (QoS) requirements. Real-time video requires an adaptive access mechanism to support the strict delay, fast bit rate variation and relatively larger bandwidth requirements. With the significant reduction in video bit rates due to the advancement in coding and signal processing techniques, real-time high quality video can be transmitted with bit rates less than 1 Mbps. Thus the support of high quality video in the wireless environment is inevitable in the near future. One serious problem for video transmission is that of the hand-off of multislot reservation. This is due to the unpredictability of the load of the adjacent cells as well as the desired QoS to be maintained by the video user. This problem is aggravated in cellular systems with only small cell overlaps, and thus reduce the hand-off region and hand-off time. In this case some video transmission interruptions (i.e. visible frame losses) may result if the number of available slots is less than the minimum requirement. Accordingly, minimum QoS is expected during hand-off if the load of the new cell is considerably high (insufficient to carry full video requirement). The main objective of this section is the provision of a mechanism in which the video encoder and the multiaccess layer will negotiate for the best video quality dynamically.

The system consists of a single video user and data users with Poisson arrivals. The data users generate a single message while being active that are exponentially distributed. In the event that data users generate large messages, a multislot reservation scheme is used on a dynamic basis. The video user employs an Auto-Regressive VBR video model with feedback (*see Chapter 3*). However, a feedback from the MAC layer is used as a control to adjust the video quantisation level in the event that the MAC layer cannot handle the required rate. For video frames that are not immediately transmitted due to the channel availability, a buffer is provided. A queue is also provided for data users.

The AR Model is used in this study which accounts for the fixed, random and correlation components of a video traffic. The output $I(n)$ is autocorrelated and is assumed to represent a constant quality video i.e. using an adaptive quantisation scale. In this case, the video quality parameter can only be expressed in terms of *Frame Loss Rate* (FLR) as a result of channel availability provided, the minimum *FLR* is satisfied (e.g. $FLR_{th}=0.05$). In video encoding where QoS is maintained, a target quantisation size for a given video frame will serve as the reference [Dalg95]. Which means, a video frame cannot be represented with higher quality beyond what the target quantisation size can achieve so that using a smaller quantisation size from the target is just a waste of capacity (smaller quantisation size will result to bigger frame sizes). In this case, we can incorporate the relative quantisation, Rq which is the ratio of the actual quantisation size to the target quantisation size to the video quality measure. Heuristically, a simple expression to describe the video quality is given as

$$V_{qos} = \left[\frac{Rq - Rq_{th}}{1 - Rq_{th}} \right]^{kv} [1 - FLR] \quad (6.29)$$

where: $FLR < FLR_{th}$ and $Rq > Rq_{th}$. The Rq_{th} and FLR_{th} stands for the threshold values and kv is the relative video quality factor. In this case the QoS can be expressed in terms of the *relative quantisation* and the corresponding *frame loss rate*. The first component is related to the video coding characteristics while the other is a result of the channel quality and channel load and as a consequence, there is always an optimal quantisation size as well as frame loss rate at certain load conditions to achieve the best quality. A sample plot of the relative video quality is in Figure 6.32.

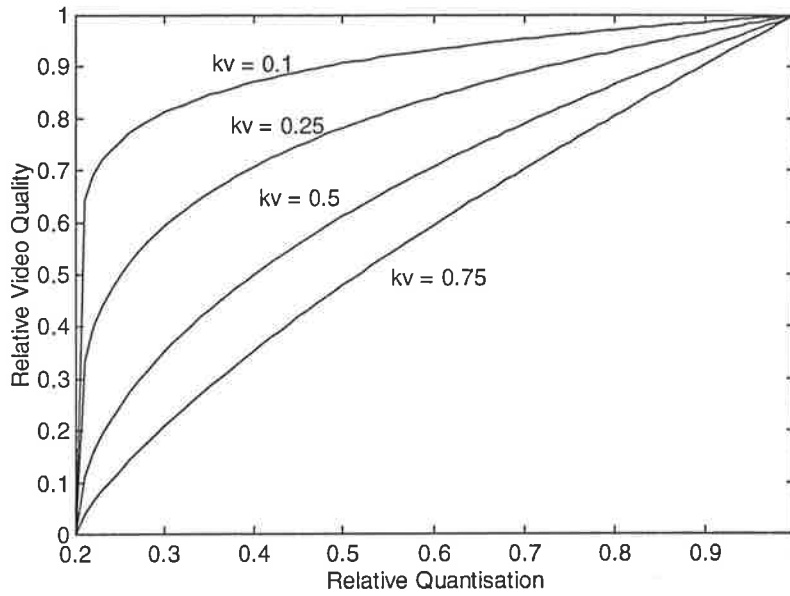


Figure 6.32 Plot of the Relative Video Quality with no Frame Loss
($Rqth=0.2$)

§6.5.2 Static Optimisation for Video/Data System

As mentioned earlier, an optimal source coding and frame loss rate is necessary to achieve a better video quality. This time we can test a video/data system based on the video quality measure described in Eq. 6.29. Consider a video/data sample system using a single R-TDMA radio channel where data users are identically independent Poisson sources. Messages are generated with a negative exponential distribution with a data activity factor α . Data traffic is given preference over video such that the slot occupancy from the data traffic is a truncated binomial defined as $\beta_{in}(M, i, \alpha)$ for $i=1, 2, \dots, N$. The video traffic exploits the multislot reservation capability. The number of slots required for each video frame is simplified which is given in the formula

$$fv(k) = \operatorname{erfc} \left(\frac{(k - VoRq) / \sigma_v}{\sqrt{2}} \right) \quad (6.30)$$

This simple model uses a one sided Gaussian distributed frame size variation with minimum size of $VoRq$. Therefore, the average frame size varies according to the chosen quantisation ratio.

The resulting *FLR* is the probability that the number of unused slots is less than the frame size. Combining with the chosen quantisation ratio, the video QoS for a static system based on Eq 29. is shown in Figure 33. The results demonstrated an optimisation curve for balancing the QoS based on the *FLR* and *Rq* parameters. This is due to the increase in *FLR* as *Rq* increases while a increasing in the source coding quality. Secondly, there is a change in the optimal value of *Rq* at different loads, suggesting an adaptive method of controlling the QoS is required as the load varies. This problem is discussed in the next section.

§6.5.3 Video with Dynamic Load Feedback

A better way of maintaining the QoS of video is to adjust the quantisation level dynamically as a function of the channel load so that the expected delay of the video frames can be controlled. The load information can be determined through the size of the buffer or the measured throughput via a dedicated feedback channel that incorporates all the necessary signalling information (i.e. channel quality, bit error rates, channel load, etc.). To evaluate the performance of this system, a simulation model is required. The system consists of a single video user competing for slots with data users with Poisson arrivals multiplexed in a 64-slot TDMA. The data users employs multislot reservation with threshold type policy. The video frames are generated at every 10 TDMA frames equivalent to 100 ms periodicity. The video source consists of a buffer which stores the subsequent frames. The video user requests for slots only at each generation of the video frame (every 640 slots) according to

$$request_slots = buffer_size/10 + guard_slots - number_reserved_slots \quad (6.31)$$

When the expected transmission delay of the arriving frame exceeds 100 ms due to the lack of the number of reserved slots, the frame size is reduced in order to compromise the delay against the quality. In order to quantify the frame size reduction of the current frame, we need to model the quality reduction as a function of frame size reduction. Firstly, we model the number of slots required for a given quantisation level given as

$$Nrs(k) = 1000 \left(k_1 (1 - w_1) w_1^{k-1} + k_2 (1 - w_2) w_2^{k-1} \right) \quad (6.32)$$

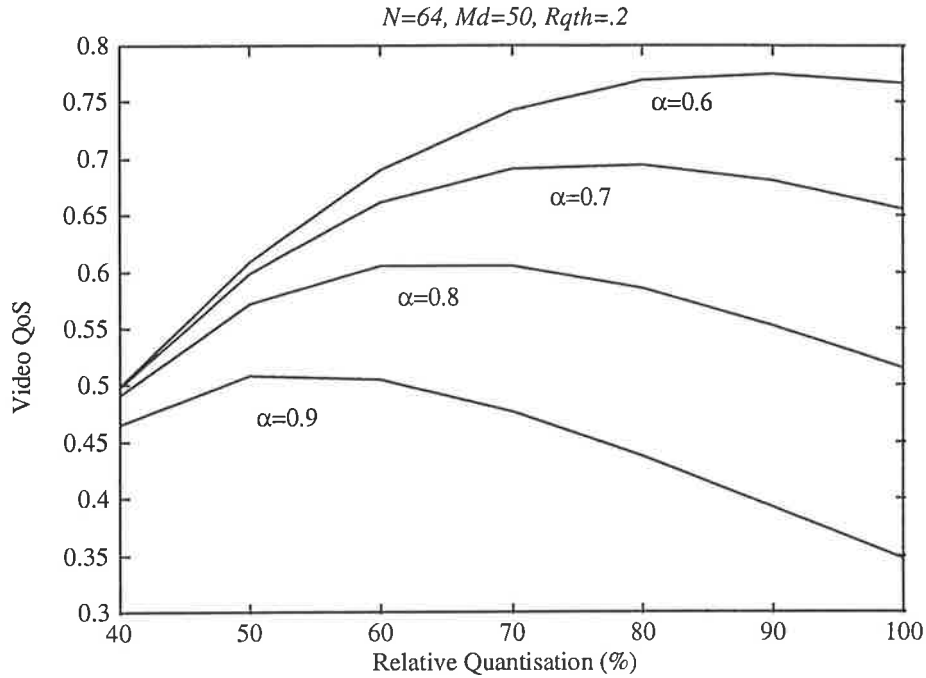


Figure 6.33 Plot of a Static Performance of a Video and Data System
 $\sigma_v = 20 Rq$ and the minimum frame size is $VoRq$

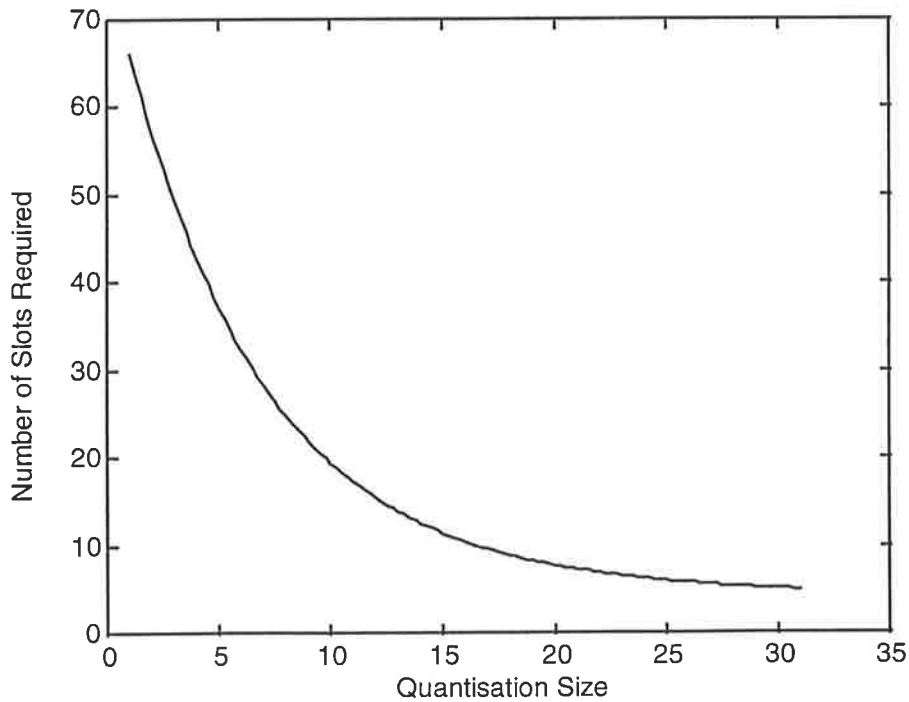


Figure 6.34 Reference Model for Video Frame Size

This curve is a typical curve for constant quality video encoding fitted from the results in [Dalg95]. However, the accuracy of the Quantisation size vs frame size varies in many video frame sequences like fast motion, still image, movie, etc. It also depends on the type of video encoder.

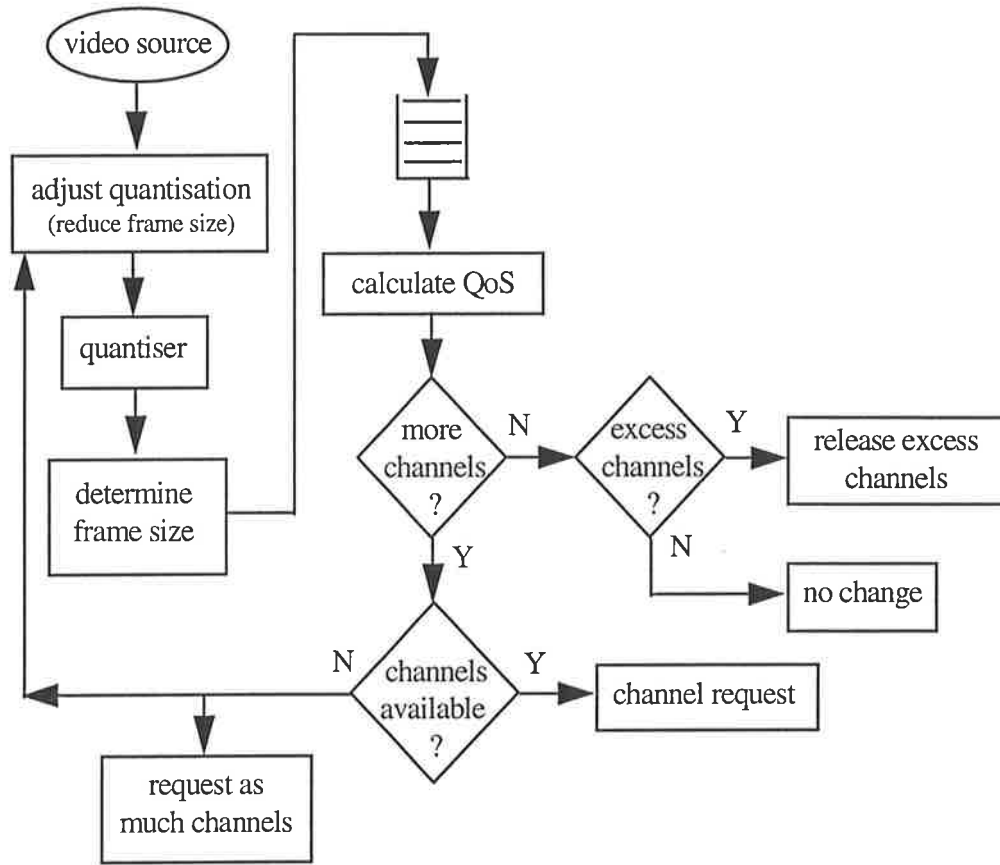


Figure 6.35 Video Coding and Channel Allocation Diagram

where $k_1=.4$, $k_2=.6$, $w_1=.85$, $w_2=.99$. The algorithm to reduce the frame size at higher loads is as follows. First, the number of required slots, N_{rs} and the number of free slots, N_{fs} are determined. If $N_{rs} > N_{fs}$, then the video frame size reduction takes place. By employing a best effort policy, the reduction of the quantisation size, D_q is simply an offset of the x -axis of Figure 6.34 due to the difference of the number of slots required to the number of slots allocated. The corresponding reduced frame size, $f_v(t)'$ is scaled from Figure 6.34. This method of maintaining the QoS is shown in Figure 6.35. Upon the reduction of the frame size, the video user maintains some guard slots reserved to accommodate the rapid frame size variation while being at a lower priority the data.

§6.5.4 Simulation and Observations

In the simulation, the data arrival rate was varied to increase/decrease the traffic load of the system. Since the video frames arrive every 100 ms, this correspond to a cycle of 640 slots assuming that the TDMA frame has a 10 ms duration. The channel bit rate was 819.2 Kb/s as 128 bits of user information per timeslot was used. During the simulation, the video quantisation reduction was measured in order to check the performance. The other parameter is the buffer size which can be used to determine the effective difference between the bit rate variation and the availability of video slots, The relative quantisation as well as the number of frame losses were also counted.

The simulation result is tabulated in Table 6.1 showing the average Rq and FLR for a given data load, α_d . As expected, large reduction in Rq is necessary in order to maintain the transmission of video frames. The relatively smaller values of FLR suggests that both the multislot reservation and the dynamic load feedback are effective in handling video transmission. The results also shows that no quality reduction will result if the channel load is low. The histogram also showed a steady QoS is achievable at moderate load. Other results are shown in Figure 6.36 and 6.37

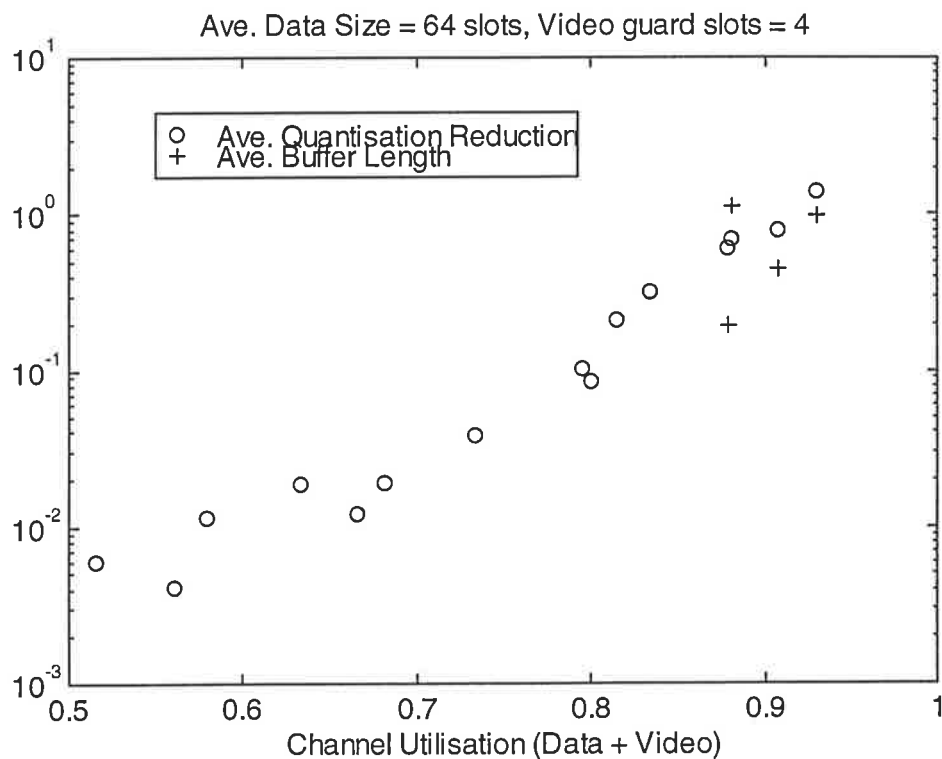
§6.6 Summary of Chapter 6

This chapter was concerned with the slot allocation strategies for maintaining the QoS of multimedia services. The multislot reservation capability of R-TDMA was being stressed as a means to provide variable bit rate and handle bursty traffic. The multislot reservation in TDMA was analysed using Birth and Death Markov chains (infinite users) as well as Discrete Markov Analysis (finite users). Both solutions were in good agreement with the simulations. Then the multiclass user and mixed traffic system performance were evaluated exploiting the multislot reservation capability. The simple algorithms were used in order for the ease of implementation. Later, a variable coding rate with multislot reservation was considered. This is essential in the slot allocation policy where delay is a prime consideration. Significant improvement was achieved when compared to a fixed coding scheme. Then lastly, the VBR video transmission was considered. The main improvement was the use of channel load feedback in optimising the video QoS. The scheme was tested in a video/data integrated system where the video user was taken with lower priority. The simulation results showed a steady QoS for video for a fixed number of data users.

Table 6.1 Video Quality Simulation Results

α_d	Rq	FLR	$Vqos$
0.71	0.97	0	0.981
0.75	0.97	0	0.981
0.80	0.91	0.004	0.938
0.85	0.77	0.038	0.812
0.91	0.61	0.072	0.783

Number of video guard slots = 2, $L = 64$, $N = 64$, $Vqos$ is calculated at $Rq_{th} = 0.2$ and $kv = 0.5$

**Figure 6.36 Minimal Overall Reduction of Video Quality**

The expected video quality degrades with channel load. At very low channel utilisation or load, the buffer is almost empty all the time. Therefore, the average quantisation reduction is small. At higher loads, the quantisation reduction mechanism cannot compensate all the variation in the video frame sizes, hence the buffer starts to build up.

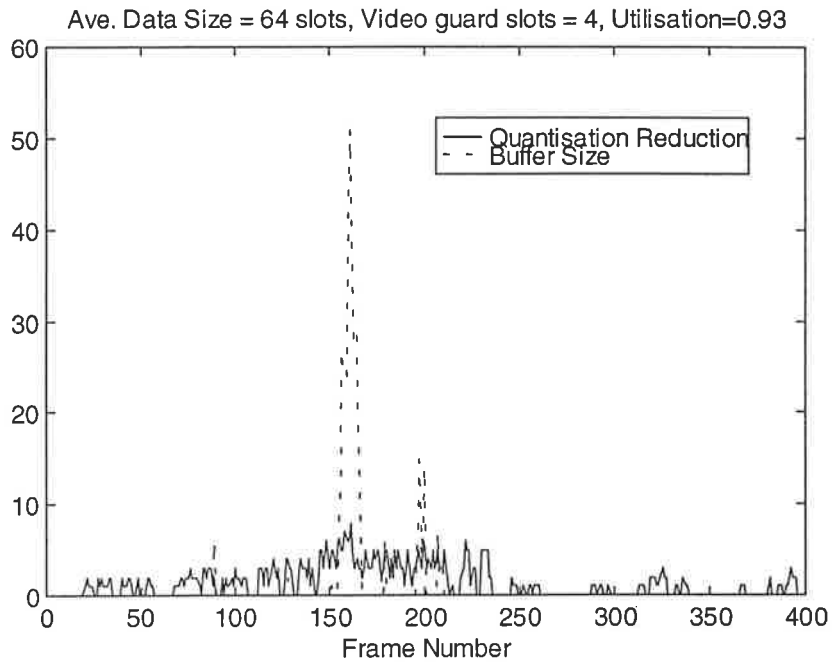


Figure 6.37 Instantaneous Reduction of Video Quality in Clusters

The simulation was conducted which assumed that video frames with very long delays were not dropped. As a result, the buffer size sometimes store a lot of video frames. This illustration suggests that frame losses is common in video transmission over a limited capacity channel. The quantisation size has to catch-up with the buffer size.

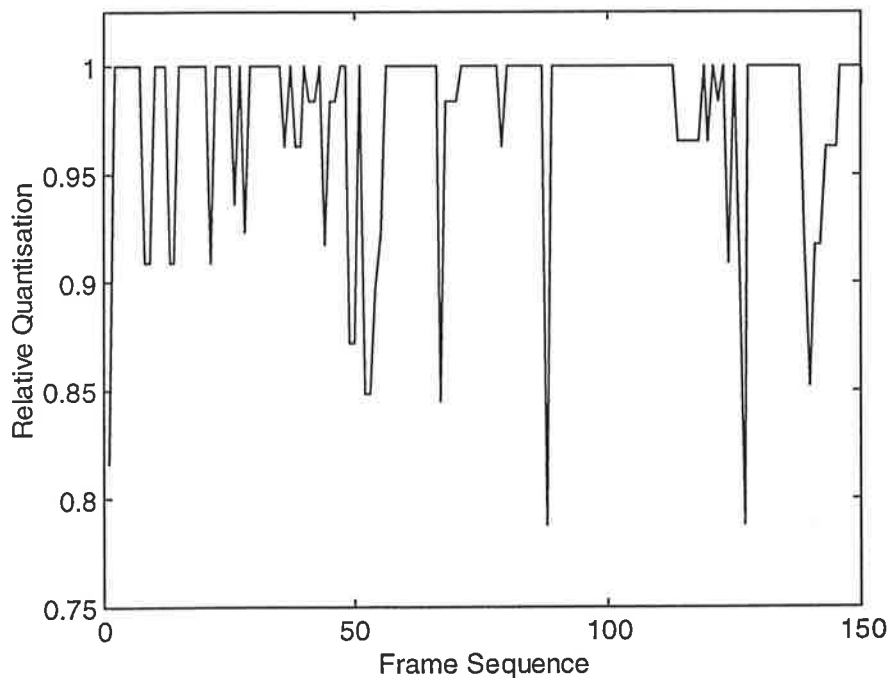


Figure 6.38 Relative Quantisation Histogram

Results taken at 0.9 throughput and video frames exceeding 100 ms delay were discarded. The number of video guard slots was 2. The observed *FLR* was almost zero.

Chapter 7

Conclusions

The aim of the future wireless personal communications is to provide a truly ubiquitous form of communication. However, a significant advancement in this area is required. One aspect of which is in the multiple access protocols. The contribution of this thesis is in the design improvements and theoretical performance analysis of Reservation based Time Division Multiple Access in order to achieve the requirements of the various wireless services. In this study, it was found that R-TDMA protocols have the potential of being a candidate technology for future wireless personal communications.

§7.1 Thesis Summary

In the study of R-TDMA protocols, two phases namely design improvements and performance analysis, were considered. The design improvements were centred on the three design criteria of WPC, fast channel access, variable rate transmission and fast error recovery. This thesis had considered mainly the first two criteria. The channel access speed was improved by 1) improving the channel structure of the TDMA frame to be more adaptive, 2) the use of prioritisation to give favour to users which requires faster access, 3) the use of random but fast polling and 4) the improvement of the network topology through the exploitation of the receiver capture capabilities. On the other hand, variable rate transmission was considered in order to support the various data services. The multislot capability was carefully studied under different scenario. The multislot reservation with multiclass users, with mixed traffic as well as with variable coding rate were also considered.

The performance analysis was divided into three parts. They were the source modelling, analytical modelling, and simulations. In the analyses, the Markov Models were used in evaluation of the queuing delays while the Transient Fluid Approximations were used for the system analysis. The S-G formulae were also derived in order to calculate the maximum capacity as a function of the channel load. For the channel access, the success probability was carefully considered as a design benchmark.

The behaviour of a reservation based protocol is highly influenced by the nature of its access mechanism. In this study, two schemes were identified namely, the random access or RMA and the polling mechanism. Most protocols were in favour for the RMA so that an approximate but rather accurate analysis is provided in §4.3. The commonly used RMA is the Slotted ALOHA due to the nature of the R-TDMA channel. For simplicity of the design, two collision resolution algorithms or CRA were widely accepted, firstly, the fixed retransmission probability which is analysed in §4.3.1, and secondly, the binary exponential back-off which is analysed in §4.3.2. Based on these two CRAs, R-TDMA can achieve a fast channel access if the load of the access slots is very low i.e. in the region of less than 0.1. However, in the event that a request is not successful (e.g. request packet collision) the nature of these CRAs cannot provide an immediate retransmission. In the case of the fixed retransmission probability, a higher retransmission probability, p can be used but this can decrease the stability of the system (see §4.5.2 and §4.5.3). For example, p is in the region of 0.1 to 0.3 for voice which means, in the event of a collision, a terminal has to wait for an average of 3.33 to 10 slots before it can retransmit.

This problem is partially solved by employing a more effective RMA. The stack algorithm addressed this problem in two ways. Firstly, the stack algorithm can achieve better performance than the Slotted ALOHA due to the processing of the feedback information. Secondly, the stack algorithm can provide an efficient prioritisation (see §5.5.3) so that users which requires faster access can be treated with higher priority. The prioritisation in the random access is also essential for protocols like PRMA since no centralised resource allocator is employed. Another way to solve this problem in ATDMA is to employ a dynamic frame configuration so that the number of access slots can be increase/decreased depending on the required performance (see §5.4.3). Although this scheme requires a proper balancing of the access delay and the queuing delay, a pseudo-bayesian control is sufficient to provide a good performance.

The support of multislot reservation has been mentioned in the various proposed protocols for the 3rd generation wireless networks. However, only a few studies had been made regarding its effect to the system performance. The flexibility of a multiple access protocol can be increased if a variable rate transmission is supported. This is the main advantage of a packet mode access over the circuit-switched mode. This is due to the nature of multimedia traffic which is characterised by a mixture of bursty and

continuous traffic. The delay improvement achieved by multislot reservation were evaluated in §6.1.2. and §6.1.3. It was shown that the improvements achieved by multislot reservation is noticeable when the channel utilisation is in the region below 0.8. This figure is often acceptable since in most applications and network design, the utilisation of 0.8 for the traffic slots is already high. Furthermore, when prioritisation was introduced in the multislot environment, better performance were achieved even beyond the 0.8 utilisation (see §6.2). This increased in flexibility is due to the different delay criteria of the different classes of users.

After identifying the design issues in R-TDMA, its performance with the various traffic was tested. Voice traffic capacity is always the basis for the comparison of the various multiple access technologies and was considered in this study. Prior to the performance evaluation, the nature of the voice packet generation was firstly studied. It was found in §3.4 that the temporal speech parameters are dependent to the hangover value used at the speech coder with fast speech activity detector. Each hangover value corresponds to transmission efficiency which is the ratio of the speech packet to the total number of packets. From a given hangover value, the maximum capacity of ATDMA for voice traffic was identified in §5.2 based on the packet dropping probability criteria. For ATDMA, the need for an optimal frame structure is required.

The second traffic that was considered was the data traffic and a simple model with Poisson arrival and negative exponential message size distribution was used. Since the throughput-load and throughput-delay characteristics are required for data systems, a mean delay analysis was developed in §4.5.3 based on TFA. The analysis of multislot reservation in §6.1 were also intended to data services. A simulation for the multislot reservation with multiclass users in §6.2 was also intended for data.

The last traffic was the VBR video. Since the bit rate variation varies significantly, an efficient transmission of VBR video requires a multiplexing of video with voice and data. In this study, it was assumed that only one video user can be accommodated in one radio channel which competes with a number of data users. In §6.5, a method for integrating video and data in which data was treated with higher priority was considered. To maintain the video quality, a video source coding with dynamic load feedback was proposed. A performance criteria using the frame loss rate and the relative quantisation was introduced in order to assess the proposed scheme. The results in §6.5.4 showed an acceptable performance under a typical load scenario.

One of the problems in the transmission of delay sensitive and error sensitive data is the presence of packet errors. One way to solve this problem is to provide a fast error recovery in the protocol. However, the delay performance can be jeopardised if massive packet errors will occur. The best way to handle this problem is to support multiple coding rates which can be varied dynamically. In this scheme, a number of coding rates must be supported by the system and the terminal can select the best coding scheme at any time. In this way, speed and accuracy are taken into account. In §6.4, the multislot reservation together with the variable coding rates scheme was proposed as a means to combat the massive packet errors that may occur in a time varying channel. It achieved a better performance as confirmed by the simulation (see §6.4.3).

§7.2 Future Work

In this thesis the design improvements as well as the performance analysis of R-TDMA were addressed. However, the study was only limited to the conceptualisation and theoretical studies of the R-TDMA protocols and the physical characteristics of the channel was not considered in detail. The next step of this work is recommended to include the channel characteristics of the 3rd generation spectrum band and the underlying physical layer characteristics in order to determine the realistic performance of R-TDMA protocols.

It is found that a strong relationship between the physical layer and the multiaccess layer exists in R-TDMA. This is demonstrated by the increased in the throughput as a result of high random access capture probability at the receiver. Knowing that the capture capability can be incorporated in the physical layer design, it must be initially considered so that a great deal of improvement can be achieved at the multiaccess layer. The strong receiver capture is the only way to improve the access speed of S-ALOHA protocol. It was also found in this work that the increased in the number of receivers for some topological arrangement can increase the access speed. It is therefore recommended to continue this work in a realistic environment.

Since this work assumed an errorless channel in most part (except in §6.4), the effect of transmission errors as well as signalling feedback errors can cause a performance degradation to R-TDMA. The transmission errors requires error recovery for data while causing a quality degradation to voice and video. The feedback error in random access can also cause large access delay especially if the error detection and

recovery procedure is slow. The future work regarding the effect of errors requires to two things, to investigate the effect of errors to voice and video traffic, and the need to incorporate the error recovery procedure in the R-TDMA protocol design.

The other area to be considered is the design of high speed multiaccess protocols for radio channels at the range of 10 Mbps. Protocol at this speed are expected to multiplex video sources as well as other broadband services. Therefore faster error retransmission and acknowledgment as well as faster channel allocation are necessary at higher speeds.

The other direction for further research is to design a common multiaccess protocol for WPC and WLAN. Although most of the problems are expected at the physical layer since the channel characteristics for indoor and outdoor environments are quite different, the need for the integration of these two environments is essential. WLAN is characterised by high speed but low mobility while WPC has relatively lower speed with higher mobility. This results to a challenging task in the design of the multiaccess protocol. However, it should be noted that both WLAN and WPC will eventually support similar applications.

Lastly, the performance of R-TDMA requires some detailed analytical work. This work covered the performance approximations based on the commonly accepted performance parameters and generic traffic source models. A single radio channel is also assumed throughout the analyses. It is therefore recommended to further study the teletraffic performance of R-TDMA together with more realistic traffic sources.

Appendices

Appendix A Sample Data Source with Buffering

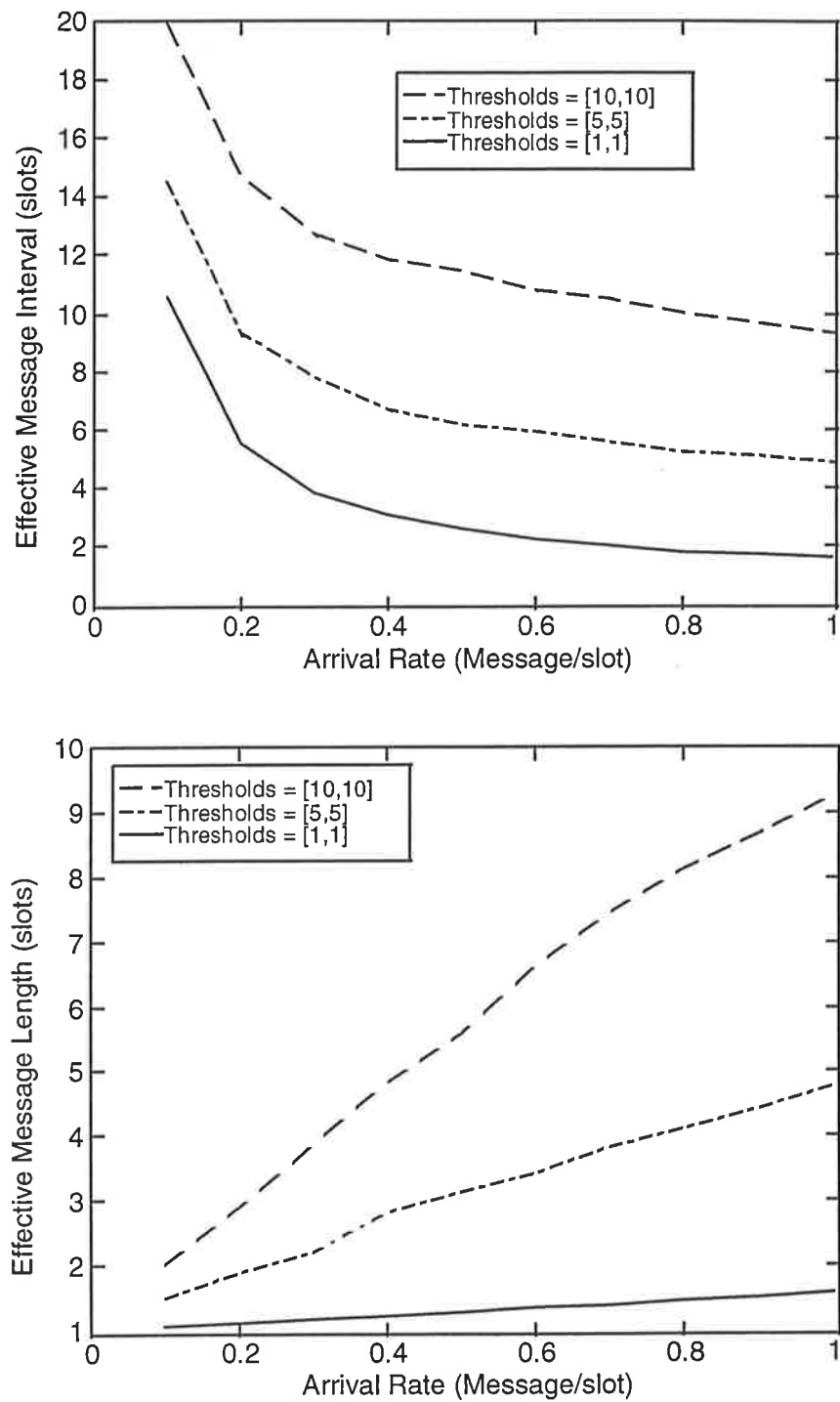


Figure A.1 Simulation Results of the Model of Queued Users with Thresholds

The top figure shows the increase of the effective message interval with large threshold values. Bottom figure shows the increase in the effective message length. $L = 1$

Appendix B ATDMA Performance

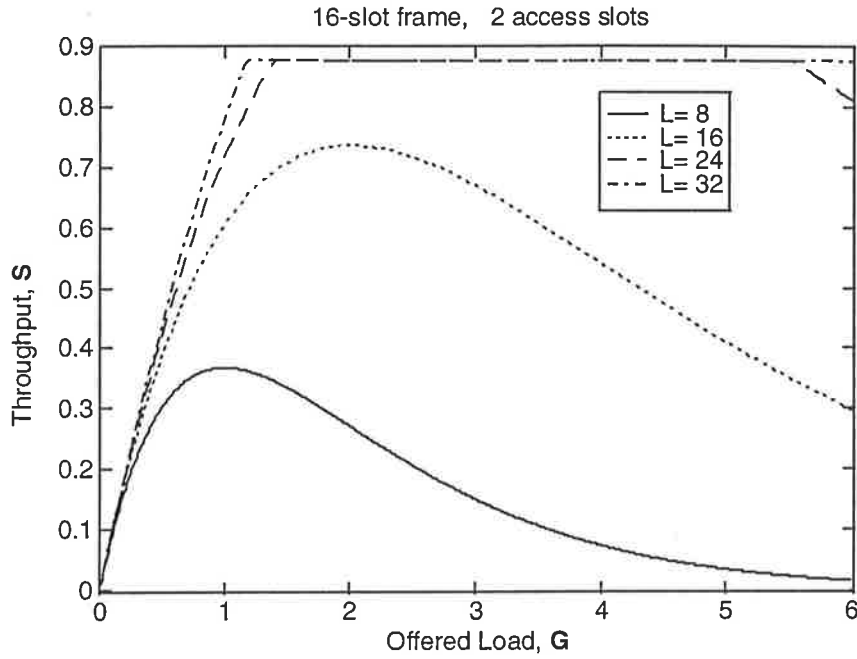


Figure B.1 ATDMA S-G Performance for [16,2] Channel

The figure shows the need to optimise the frame structure of ATDMA. Smaller values of L corresponds to a faster contention rate and in effect, the throughput is jeopardised. However, the acceptable operating region in cases where the delay is considered is only at regions where G/S is close to unity.

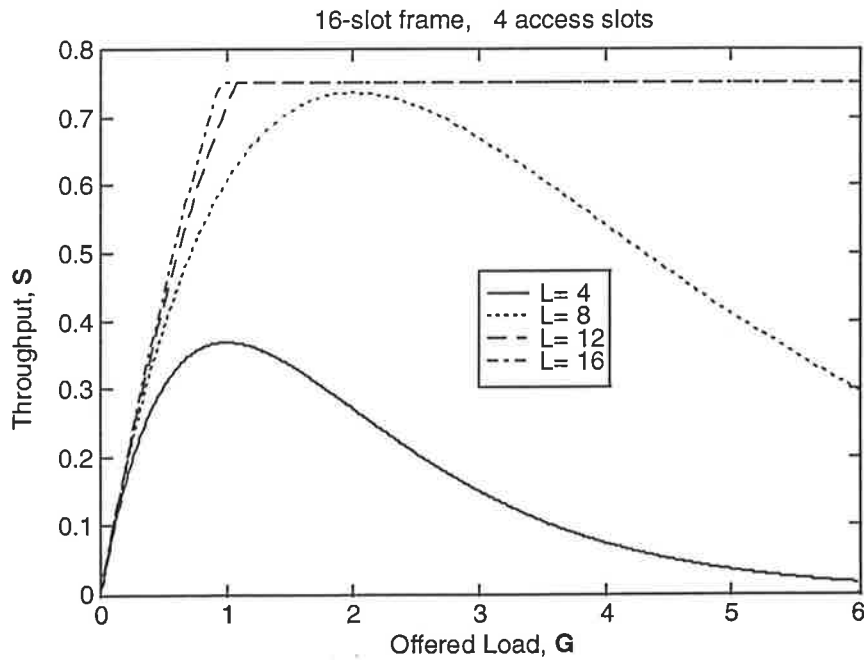


Figure B.2 ATDMA S-G Performance for [16,4] Channel

The figure shows an improved performance for $L=8$ almost doubling the maximum throughput. However, the maximum achievable throughput in case of large L (i.e. $L=16$) is reduced to only 0.75. This suggests that a dynamic frame configuration or a variable number of access slots is necessary to increase the flexibility of ATDMA protocol.

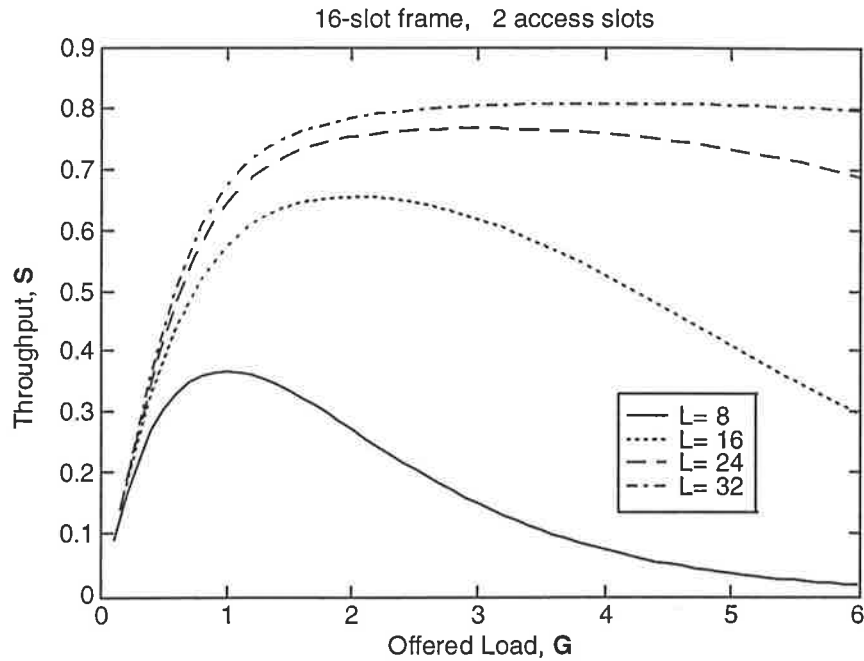


Figure B.3 ATDMA S-G Performance with Blocked Arrivals Cleared

S-G Analysis is always applied to random access protocols. ATDMA can also be a random access protocol if the queue is removed (Blocked Calls Cleared). In this case the G/S parameter is still essential. The blocking probability is simply $1-S/G$.

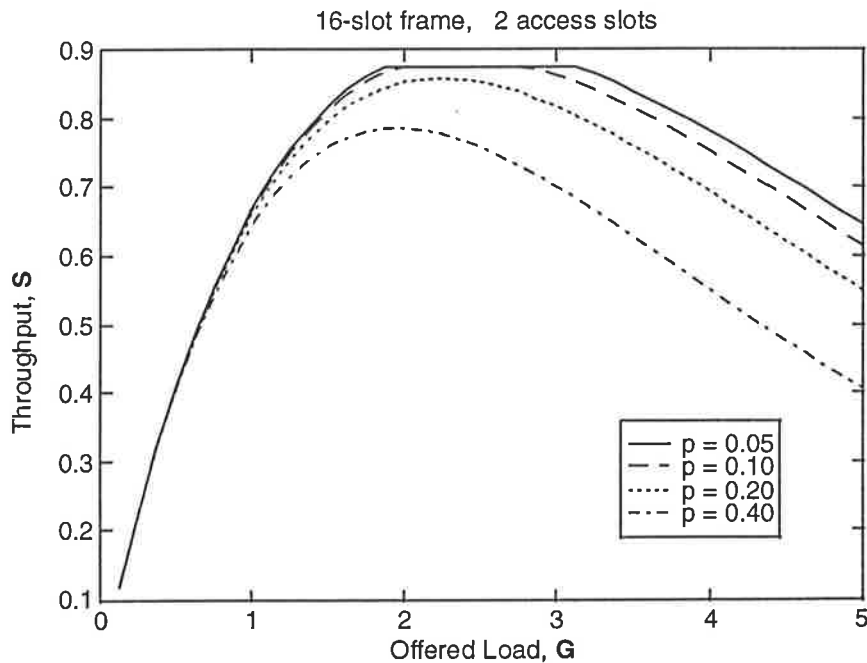


Figure B.4 S-G Performance for Different Retransmission Probabilities

The figure shows the effect of the retransmission probability at higher load for $L=20$. The S-G performance shows only the advantage of lower values of p . However, the delay performance is also required to determine the optimal value. Stability is also one parameter to consider.

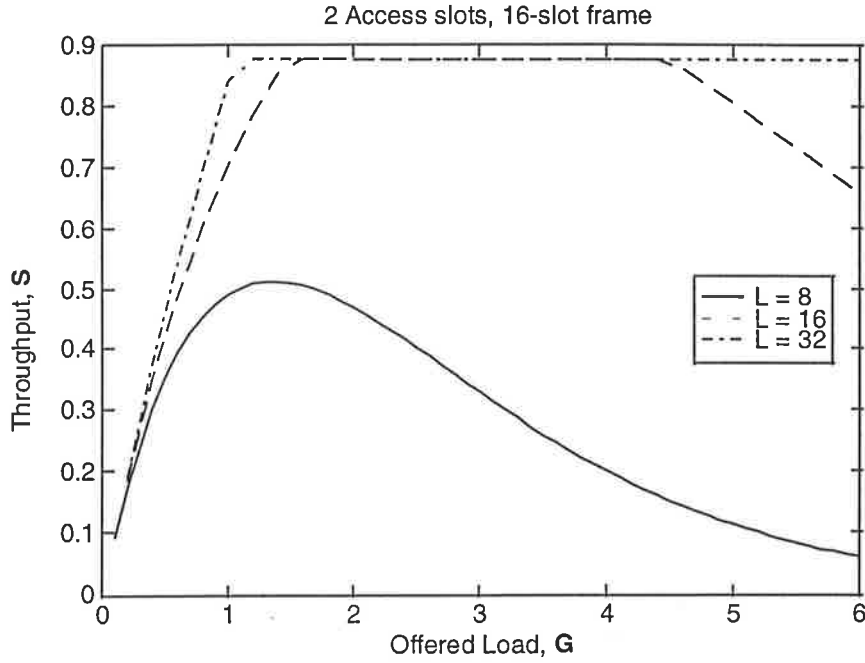


Figure B.5 ATDMA S-G Performance with Packet Capture

Capture capability is one parameter that can enhance the performance of R-TDMA protocols. The figure shows a dramatic improvement in the S-G performance. For $L=32$, the S/G ratio is almost close to unity at $G<1$. This shows that capture can improve both the throughput/load and throughput/delay performance.

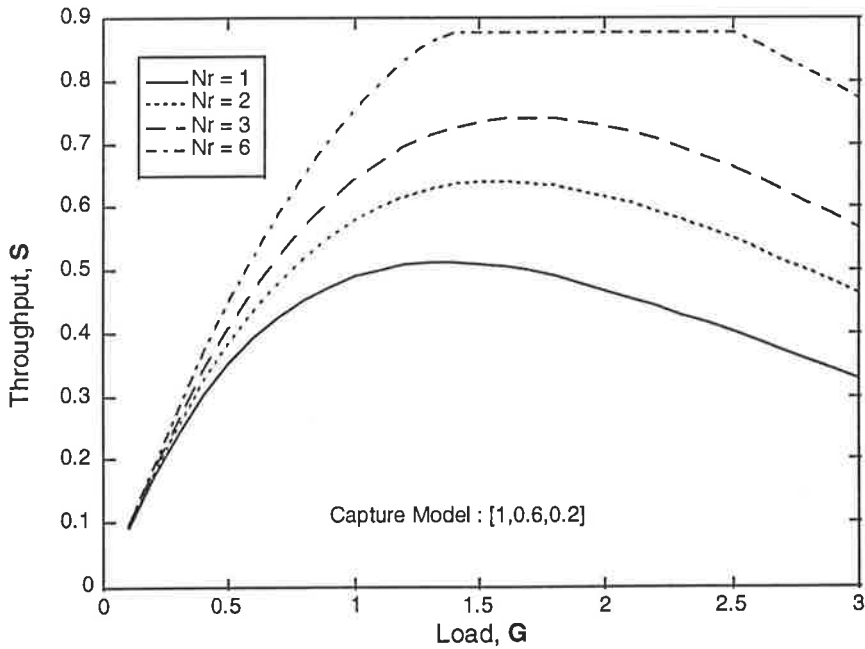


Figure B.6 S-G Performance with Different Number of Receivers

The figure is taken for $L=8$ and $[2,16]$ channel with $p=0.1$. The number of receivers is varied. The value of $L=8$ which achieved only around 0.35 maximum throughput in conventional ATDMA has achieved the maximum achievable throughput for 6 receivers.

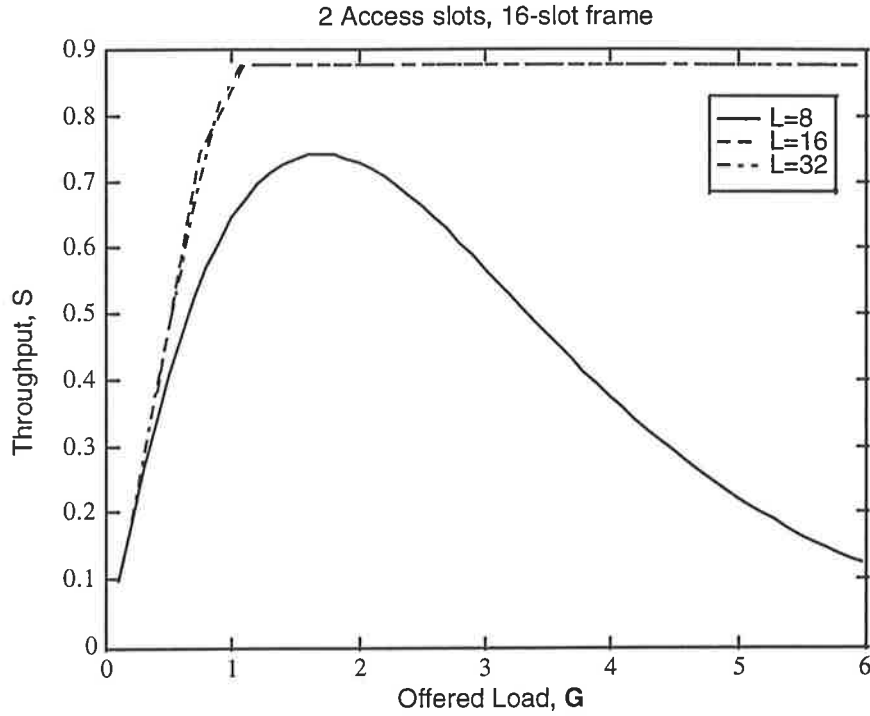


Figure B.7 ATDMA S-G Performance with Packet Capture and Multiple Receivers
($N_r=3$)

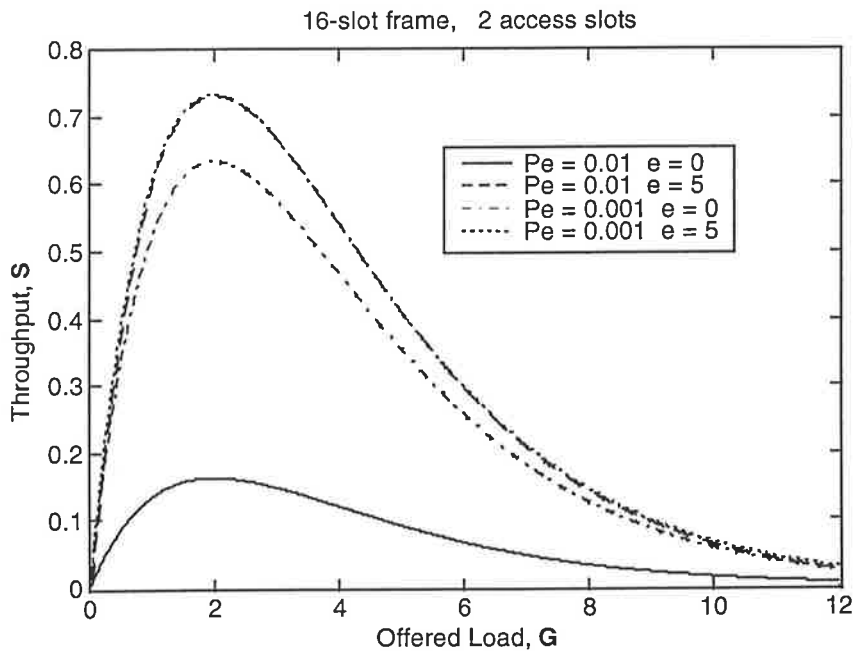


Figure B.8 ATDMA Performance with Bit Errors and FEC

Bit Errors are known to degrade the performance of access protocols. Therefore FEC is necessary to combat such problem. An error correcting capability for less than or equal to e errors is used and compared to a no FEC case. $L = 16$

Appendix C Simulation on the Stability of S-ALOHA

The stability of S-ALOHA is measured in terms of the First Exit Time [Klie75]. Stability can be achieved if proper selection of the retransmission probability is observed. However, if the traffic is not stationary, stability is not guaranteed (i.e. changes in the arrival rate). Most of the stability measure are taken from a Poisson input traffic. In the simulation, both Poisson and bursty traffic using Two-state MMPP were used. For the MMPP source, a sojourn time of 200 slots for both states were selected and the arrival rates were chosen such that the effective arrival rate for the first state is twice that of the second state. In this way, a bursty traffic was generated. The IDC was measured at 100 slots intervals. An IFT assumption with retransmission probability of $p=0.1$ was used. Backlogged packets refresh after the 8th collision. The total arrival rate was varied and results are shown in Figure E.1 and Figure E.2.

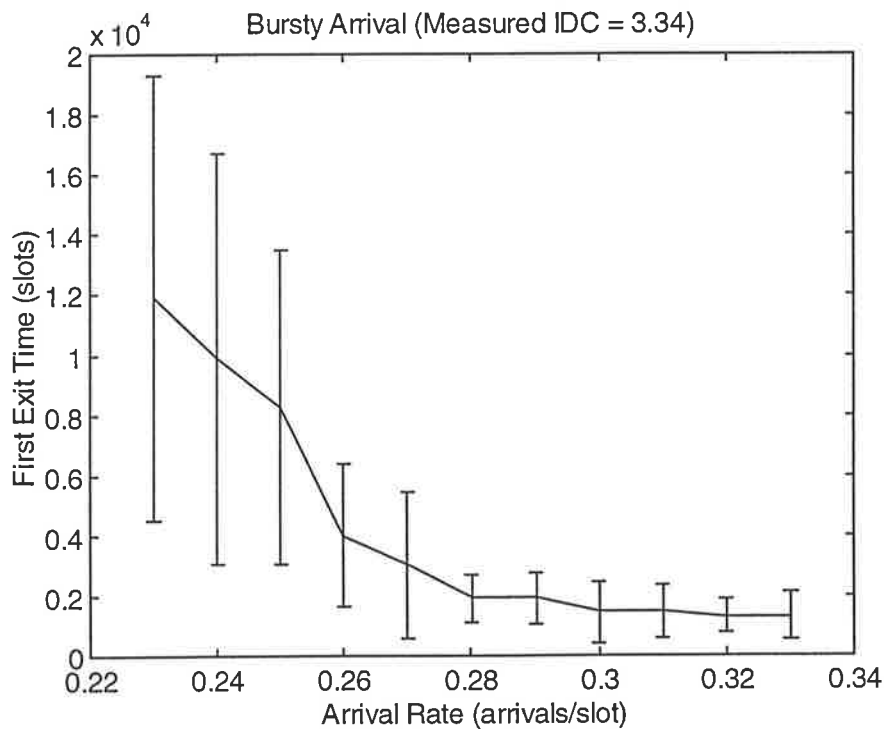


Figure C.1 Stability of Slotted ALOHA with Bursty Arrivals

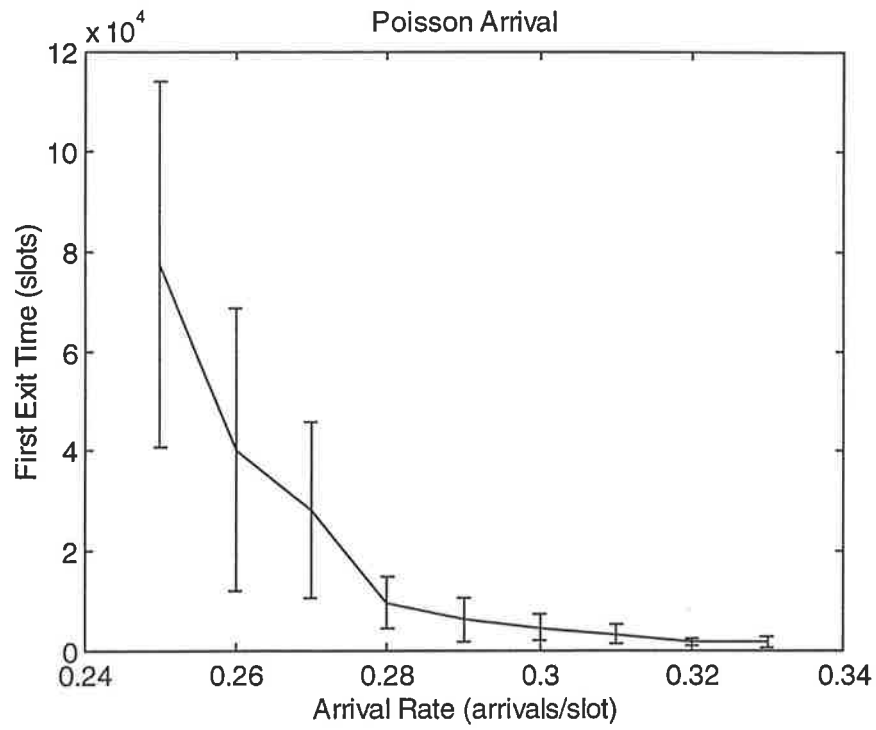


Figure C.2 Stability of Slotted ALOHA under Poisson Arrivals

Appendix D Rough Approximation of R-TDMA with Voice Traffic

The capacity calculation of R-TDMA protocols for voice traffic can be approximated as follows. From a speech statistics based on a chosen hangover period, the corresponding voice activity factor is α . To maximise the capacity we have

$$Na + Nv + Ng = N$$

where Na , Nv and Ng are the number of access slots, reserved slots, and guard slots for the incoming voice packets. For a 0.1 load of the access slots and a k standard deviation for Ng (Christensen's Method), our expression becomes:

$$\frac{10M\alpha}{Tt} + M\alpha + k\sqrt{M\alpha(1-\alpha)} = N$$

where Tt = mean talkspurt duration. By rearranging and squaring the square root term, we arrive to a quadratic expression

$$a^2 M^2 - [2aN + k\alpha(1-\alpha)]M + N^2 = 0$$

where $a = 10\alpha/Tt + \alpha$. From the standard solution of quadratic equations, we arrive to

$$M = \frac{-b \pm \sqrt{b^2 - 4a^2 N^2}}{2a^2}$$

where $b = -[2aN + k\alpha(1-\alpha)]$. Tabulating the results for $k=2$ chosen to provide an acceptable packet dropping probability, the results are comparable to the results of §5.2.3.

Hangover (ms)	Number of Voice Users
0	150
25	154
50	156
75	156
100	154
150	151
200	147
250	143
375	134
500	127

Appendix E Derivation of the Stack Splitting Parameter

Proof of $V = \frac{(1-\lambda) \sum_{z \geq 2} z \hat{P}(z, \lambda)}{\sum_{z \geq 2} z(z-1) \hat{P}(z, \lambda)}$ in §5.3.2

From $P_{succ} = \sum_{z \geq 2} [B_{in}(z, 0, \phi) P_s(1, \lambda) + B_{in}(z, 1, \phi) P_s(0, \lambda)] \hat{P}(z, \lambda)$

differentiating P_{succ} with respect to ϕ ,

$$\begin{aligned} \frac{d}{d\phi}(P_{succ}) &= \frac{d}{d\phi} \left(\sum_{z \geq 2} (1-\phi)^z \lambda e^{-\lambda} \hat{P}(z, \lambda) + \frac{z!}{(z-1)!} \phi (1-\phi)^{z-1} e^{-\lambda} \hat{P}(z, \lambda) \right) \\ &= \frac{d}{d\phi} \left(e^{-\lambda} \sum_{z \geq 2} [\lambda (1-\phi)^z + z \phi (1-\phi)^{z-1}] \hat{P}(z, \lambda) \right) \\ &= \sum_{z \geq 2} \lambda z (1-\phi)^{z-1} (-1) + z [\phi (z-1) (1-\phi)^{z-2} (-1) + (1-\phi)^{z-1}] \hat{P}(z, \lambda) \\ &= \sum_{z \geq 2} \left[-\lambda (1-\phi)^{z-1} - \phi (z-1) (1-\phi)^{z-2} + (1-\phi)^{z-1} \right] z \hat{P}(z, \lambda) \\ &= \sum_{z \geq 2} \left[(1-\phi)^{z-1} (1-\lambda) - \phi (z-1) (1-\phi)^{z-2} \right] z \hat{P}(z, \lambda) \end{aligned}$$

setting to zero we have:

$$0 = \sum_{z \geq 2} \left[(1-\phi)^{z-1} (1-\lambda) - \phi (z-1) (1-\phi)^{z-2} \right] z \hat{P}(z, \lambda)$$

dividing by $(1-\phi)^{z-1}$

$$0 = \sum_{z \geq 2} \left[(1-\lambda) - \phi (z-1) (1-\phi)^{-1} \right] z \hat{P}(z, \lambda)$$

$$0 = \sum_{z \geq 2} \left[(1-\lambda) z \hat{P}(z, \lambda) - \frac{\phi}{(1-\phi)} (z-1) z \hat{P}(z, \lambda) \right]$$

$$(1-\lambda) \sum_{z \geq 2} z \hat{P}(z, \lambda) = \frac{\phi}{(1-\phi)} \sum_{z \geq 2} (z-1) z \hat{P}(z, \lambda)$$

$$\frac{\phi}{(1-\phi)} = \frac{(1-\lambda) \sum_{z \geq 2} z \hat{P}(z, \lambda)}{\sum_{z \geq 2} z(z-1) \hat{P}(z, \lambda)}$$

Thus
$$V = \frac{(1-\lambda) \sum z \hat{P}(z, \lambda)}{\sum z(z-1) \hat{P}(z, \lambda)}$$

and rearranging
$$\phi = \frac{V}{1+V}$$

$$\phi_1 = \text{Load of stack 1} = \frac{V}{1+V}$$

In solving ϕ_2 we have to maximising the success in stack 2

Load of stack 2 = $(1-\phi_1)\phi_2 = \phi_1$

therefore

$$\phi_2 = \frac{\phi_1}{1-\phi_1}$$

Appendix F Results of the Multiclass Stack Simulations

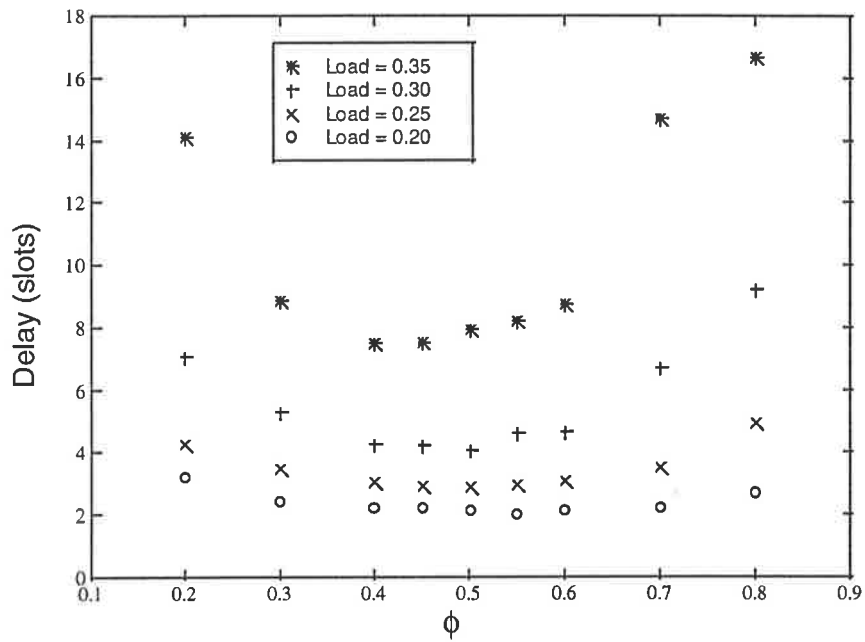


Figure F.1 Class 1 Delay against ϕ

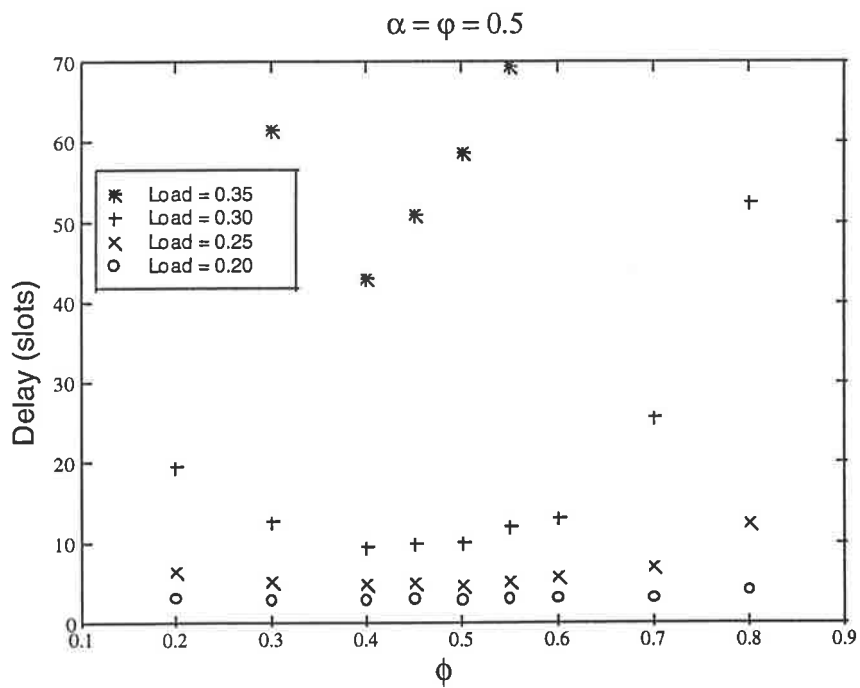


Figure F.2 Class 2 Delay against ϕ

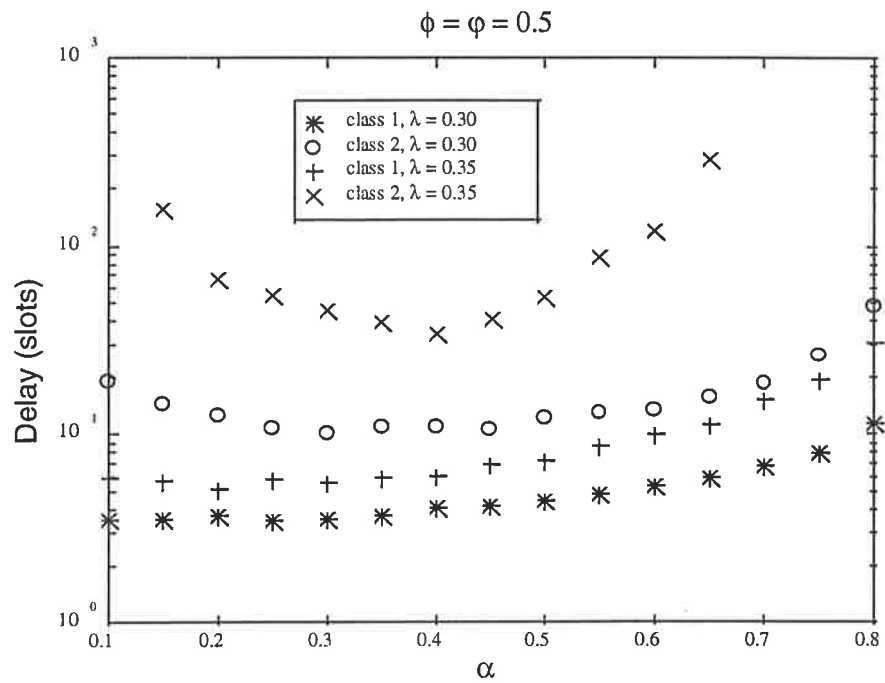


Figure F.3 Delay of a Two-class Algorithm for Various α

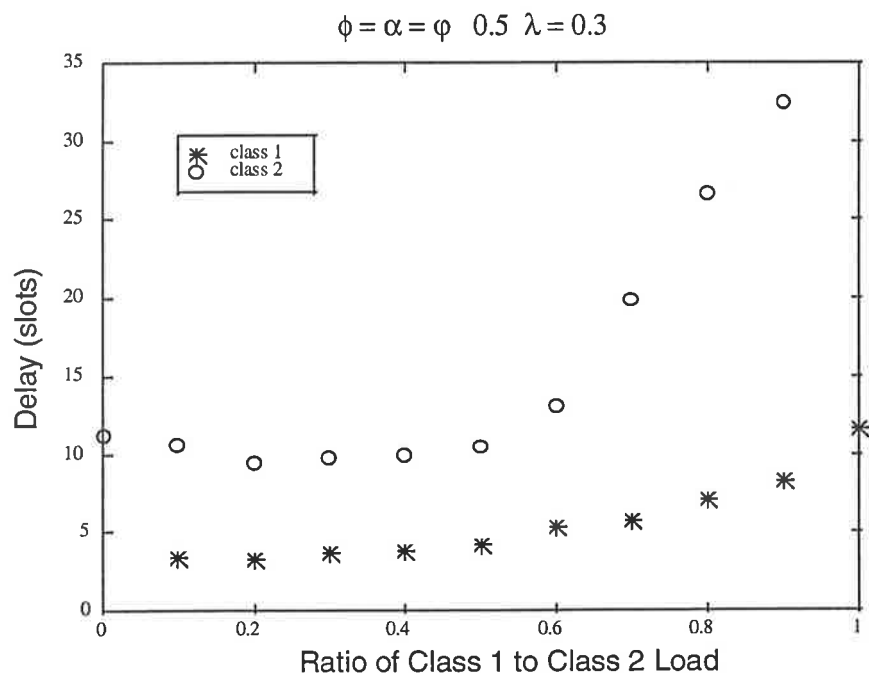


Figure F.4 Effect of the Load Proportion to Delay

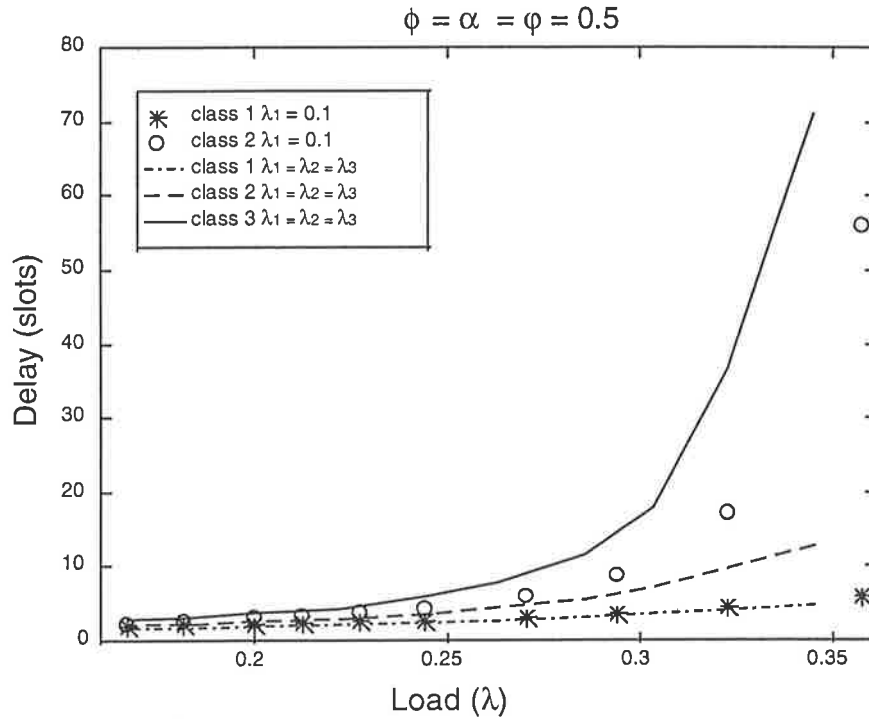


Figure F.5 Throughput/Delay Curve

Test of prioritisation for a 2-class system and a 3-class system. For the 2-class test, the class 1 load is fixed at 0.1 and the class 2 load is increased. In the second test, all loads are of equal proportion. The plot shows that the mean delay for class 1 is not affected by the class 2 load. Similarly, a good rejection of the class 3 users is indicated by the sudden increase of the delay when the load reaches near congestion.

Appendix G Results of the SCARP Simulations

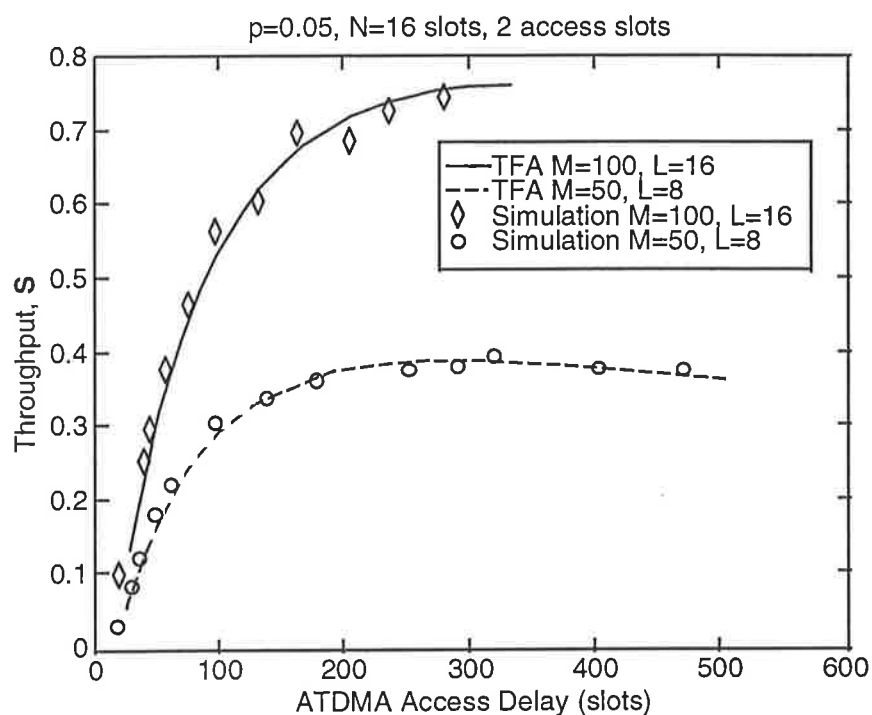


Figure G.1 Access Delay in Advanced TDMA Protocol

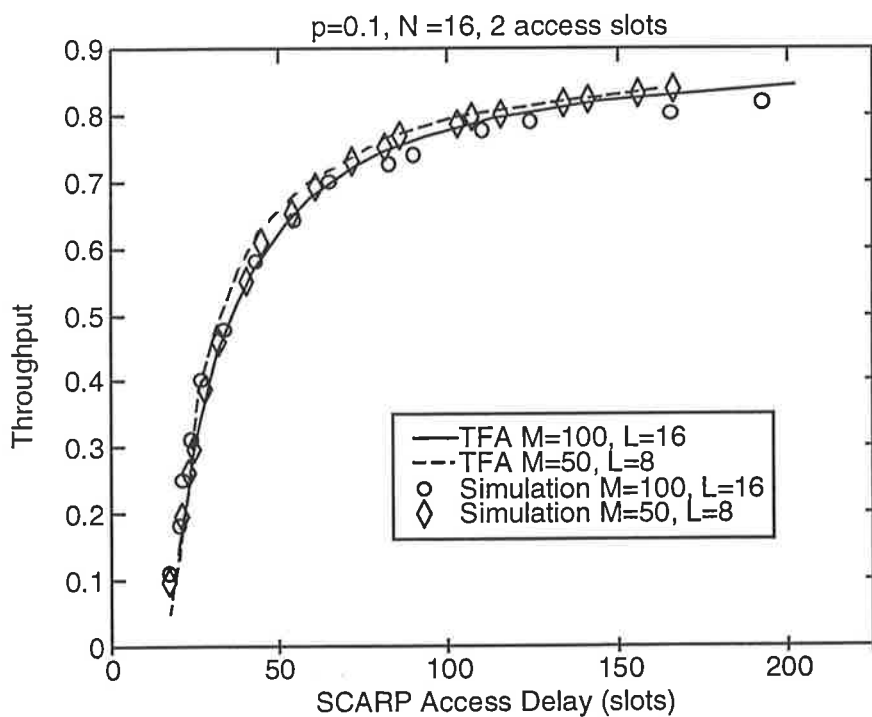


Figure G.2 Access Delay in SCARP

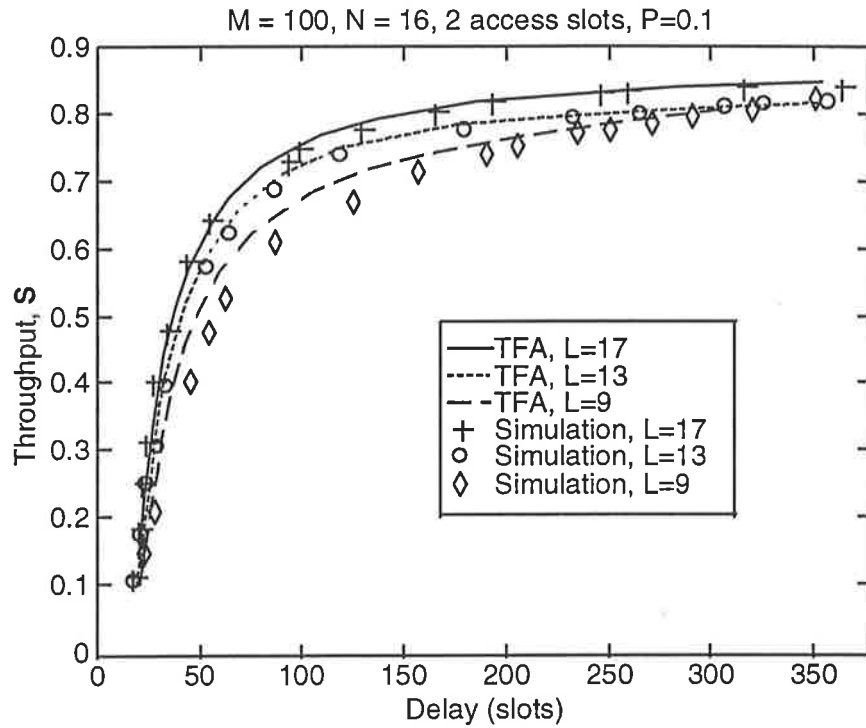


Figure G.3 Access delay for Different Message Length

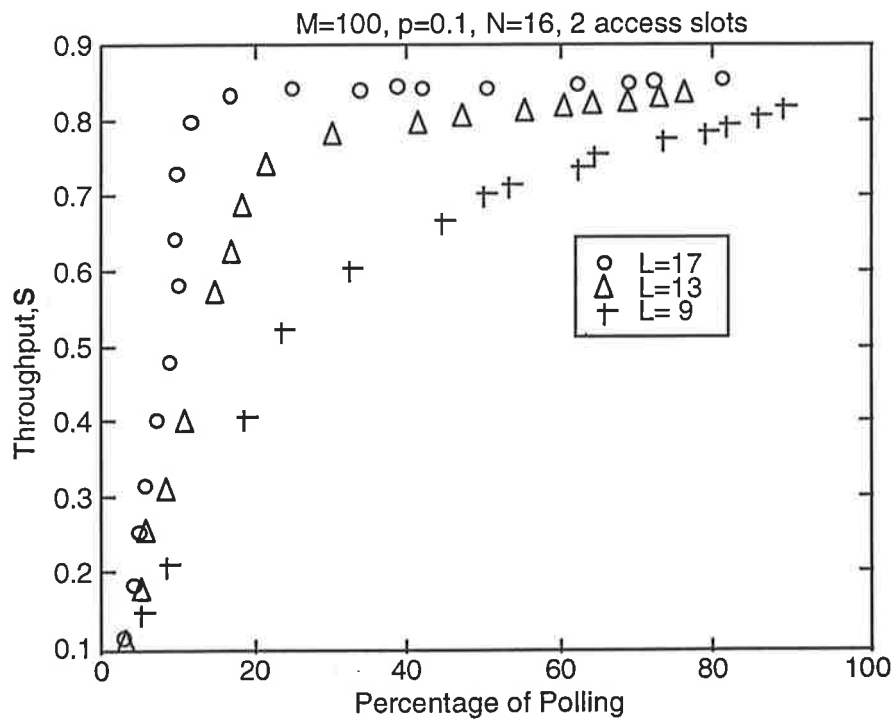


Figure G.4 Plot of the Percentage of Polling in During Access

Appendix H Results of the Multiclass Multislot Simulation

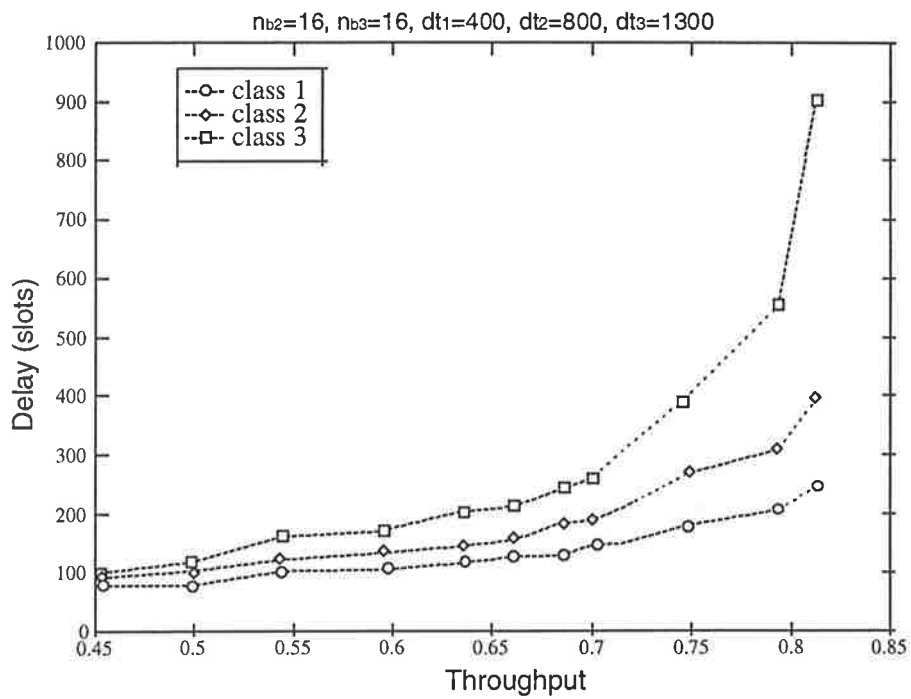


Figure H.1 Performance of Algorithm 1 in Advanced TDMA

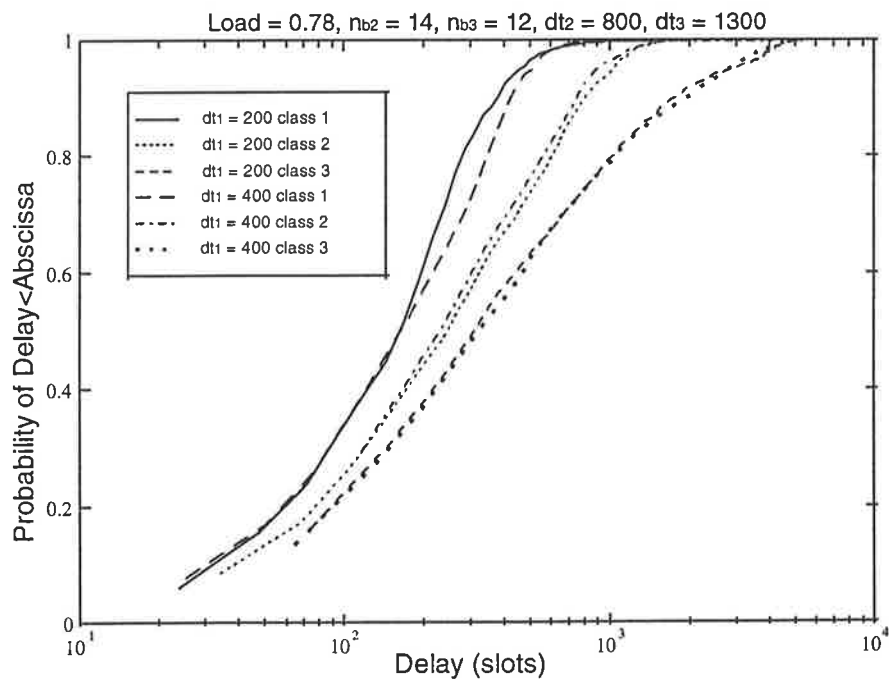


Figure H.2 Cumulative Delay in Advanced TDMA (Algorithm 1)

References

- [Abra70] N. Abramson, "The ALOHA System-Another Alternative for Computer Communications," in *Proc. 1970 Fall Joint Comput. Conf., AFIPS Press*, Vol.37, pp.281-285, 1970.
- [Abra94] N. Abramson, "Multiple Access in Wireless Digital Networks," *Proc. of the IEEE*, Vol.82, No. 9, Sep 1994, pp.1360-1370.
- [Abra96] N. Abramson, "Wideband Random Access for the Last Mile," *IEEE Personal Communications Magazine*, Dec 1996, pp. 29-33.
- [Akaiw92] Y. Akaiwa, T. Nomura, and S. Minami, "An Integrated Voice and Data Radio Access System," in *Proc. 1992 Vehicular Technology Conference (VTC)*, Denver Colorado, pp. 255-258.
- [Ami92] N. Amitay, "Resource Auction Multiple Access (RAMA): Efficient Method for Fast Resource Assignment in Decentralised Wireless PCS," *Electronics Letters*, Vol.28, No.8, 9th Apr 1992, pp. 799-901.
- [Ami94] N. Amitay and S. Nanda, "Resource Auction Multiple Access (RAMA) for Statistical Multiplexing of Speech in Wireless PCS," *IEEE Transactions on Vehicular Tech.*, Vol 43, No. 3, Aug 1994, pp 585-595.
- [Apos93] K. Apostolidis, et al "A Reservation Protocol for packet Voice and Data Integration in a Unidirectional Bus Network," *IEEE Transactions on Communications*, Vol.41, No.3, Mar 1993, pp.478-485.
- [Ben94] G. Benelli, G. Cau and A. Radaelli, "A Performance Evaluation of Slotted Aloha Multiple Access Algorithms with Fixed and Variable Frames for Radiomobile Networks," *IEEE Transactions on Vehicular Technology*, Vol. 43, No. 2, May 1994, pp. 181-193.
- [BerGal92] D. Bertsekas and R. Gallager, *Data Networks*, Second Ed., Prentice-Hall, Inc., 1992.
- [Bianch94] G. Bianchi, et al, "C-PRMA: the Centralized Packet Reservation Multiple Access for Local Wireless Communications," in *Proc. on Global Conference Record (GLOBECOM'94)*, San Francisco, Dec 1994, pp.1340-1344.
- [Bolla95] F. Bolla, et al, "Performance of Simple Priority Schemes for the RRA-ISA Multiple Access in Packet Voice and Data Cellular Systems," in *Proc. 1995 Vehicular Technology Conference*, July 25-28 Chicago IL, pp.939-944.

- [Brad69] P.T. Brady, "A Model for Generating On-Off Speech Patterns in Two-Way Conversation," *The Bell Systems Technical Journal*, Vol.48, No.7 Sept. 1969, pp.2445-2472.
- [Bra69] P.T. Brady, "A Statistical Analysis of On-Off Patterns in 16 Conversations," *The Bell Systems Technical Journal*, Jan 1967, pp.73-91.
- [Brad67] P.T. Brady, "A Technique for Investigating On-Off Patterns of Speech", *The Bell Systems Technology Technical Journal*, Vol XLIV Number 1 Jan 1965, pp1-22.
- [Buot96a] T. Buot and F. Watanabe, "Random Access Algorithm for Users with Multiple Priorities," *IEICE of Japan Transactions on Communications*, Vol.E79-B No.3, March 1996.
- [Buot96b] T. Buot and F. Watanabe, "Channel Allocation Algorithm for Multislot TDMA with Multiclass Users," *IEICE of Japan Transactions on Communications*, Vol.E79-B No.3, March 1996.
- [Buot96c] T. Buot, "Random Access, Reservation and Polling Multiaccess Protocol for Wireless Data Systems," *International Federation for Information Processing (IFIP'96) 14th World Computer Congress*, Canberra, Australia, September 2-6, 1996.
- [Buot95a] T. Buot, "Priority Schemes for Mobile Data Access Employing Reservation," *IEEE World Wireless Communications Symposium*, Long Island, NY, November 1995.
- [Buot95b] T. Buot, "Channel Allocation Strategy for Voice/Data TDMA Systems," *4th IEEE International Conference on Universal Personal Communications (ICUPC'95)*, Tokyo, Japan. November 6-10, 1995.
- [Cap79] J.I. Capetanakis, "Tree Algorithms for Packet Broadcast Channels," *IEEE Transactions on Information Theory*, Vol. IT-25, No.5, pp.505-515, Sep. 1979.
- [Carl75] A. Carleial, "Bistable Behavior of ALOHA-Type Systems," *IEEE Transactions on Communications*, Vol. COM-23, No.4, pp.401-409, April 1975.
- [Chang94] C-J Chang and C-H Wu, "Slot Allocation for an Integrated Voice/Data TDMA Mobile Radio system with a Finite Population of Buffered Users," *IEEE Transactions on Vehicular Technology*, Vol.43, No.1, pp.21-26, Feb.1994
- [Cho95] H.F. Chou, C.H. Lee and K.C. Chen, "Group Randomly Addressed Polling with Reservation for Wireless Integrated Networks", *Proc. Personal, Indoor and Mobile Radio Communications (PIMRC'95)*, Sep. 27-29, Toronto, 1995.

- [ChU95] W.S. Chung and C.K. Un, "Collision Resolution Algorithm for M-priority users," *IEE Proc.-Commun.*, Vol. 142, No. 3, June 1995.
- [Cleary94] A.C. Cleary and M. Paterakis, "An Investigation of Reservation Random Access Algorithm for Voice-Data Integration in Microcellular Wireless Environments" in *Proc. IEEE GLOBECOM'94*, 1994, pp 1333-1339.
- [Cox92] D. Cox,, "Wireless Network Access for Personal Communications," *IEEE Communications Magazine*, pp. 97-115, Dec. 1992.
- [Coop72] R.B. Cooper, *Introduction to Queueing Theory*, The Macmillan Company, 1972.
- [Dalg95] I. Dalgic and F. Tobagi, "Constant Quality Video Encoding," in *Proc. 1995 International Conf. on Communications*, pp.1255-1261.
- [Dav80] D. Davis and A. Gronemeyer, "Performance of Slotted ALOHA Random Access with Delay Capture and Randomized Time of Arrival," *IEEE Transactions on Communications*, Vol.COM-28, No.5, pp.703-710, May 1980.
- [Dev93] J.M. DeVile, "A Reservation based Multiple Access Scheme for a Future Universal Mobile Telecommunications System," in *Proc. of 7th IEE Conf. on Mobile and Personal Communications*, Brighton, U.K., pp.210-215, Dec. 1993.
- [Dunl94] J. Dunlop, P. Cosimini and D. Robertson, "Optimisation of Packet Access Mechanism for Advanced Time Division Multiple Access (ATDMA)," in *Proc. 44th IEEE Vehicular Technology Conference*, Stockholm, pp.1040-1044, June 1994.
- [Dunl93] J. Dunlop et al "Optimisation of Packet Access Mechanisms for Advantced Time Division Multiple Access (ATDMA)," in *Proc. of 7th IEE Conf. on Mobile and Personal Communications*, Brighton, U.K., pp.237-242, Dec. 1993.
- [Dunl95] J. Dunlop, D. Irvine, D. Robertson, and P. Cosimini, "Performance of a Statistically Multiplexed Access Mechanism for a TDMA Radio Interface," in *IEEE Personal Communications*, pp. 56-64, June 1995.
- [Eklun86] B. Eklundh, "Channel Utilization and Blocking Probability in a Cellular Mobile Telephone system with Directed Retry," *IEEE Transactions on Communications*, Vol.34, No.4, Apr 1986, pp.329-337.
- [EpsS95] B. Epstein and M. Schwartz "Reservation Strategies for Multi-Media Traffic in a Wireless Environment," *1995 IEEE Vehicular Technology Conf.*, Chicago IL, July 25-28.

- [Eve94] D. Everitt, "Traffic Engineering of the Radio Interface for Cellular Mobile Networks," *Proceedings of the IEEE*, Vol. 82, No. 9, pp. 1371-1381, Sep. 1994.
- [Falk83] G. Falk, et al "Integration of Voice and Data in the Wideband Packet Satellite Network," *IEEE Journal in Selected Areas in Communications*, Vol. SAC-1, No.6, pp.1076-1083, Dec 1983
- [Fisch76] M.J. Fischer, T.C. and Harris, "A Model for Evaluating the Performance of an Integrated Circuit and Packet-Switched Multiplex Structure," *IEEE Transactions on Communications*, Vol COM-24, No.2, pp.195-202, Feb 1976.
- [Fritch67] B. Fritchman, "A Binary Channel Characterization Using Partitioned Markov Chains," *IEEE Transactions on Information Theory*, Vol.IT-13, No.2, Apr 1967, pp 221-227.
- [Frull93] M. Frullone, G. Falciasecca, P. Grazioso, G. Riva, and A.M. Serra, "On the Performance of Packet Reservation Multiple Access with Fixed and Dynamic Channel Allocation," *IEEE Transactions on Vehicular Technology*, Vol.42, No.1, pp.78-86, Feb. 1993.
- [Frull94] M. Frullone, et al. "Comparison of Multiple Access Schemes for Personal Communications Systems in a Mixed Cellular Environments," *IEEE Transactions on Vehicular Technology*, Vol.43, No.1, pp. 99-109, Feb 1994.
- [Fuk83] A. Fukuda, and S. Tasaka, "The equilibrium point analysis - a unified analytic tool for packet broadcast networks," in *Proc. GLOBECOM'83*, pp.33.4.1 - 33.4.8, 1983.
- [Gil89] K. Gilhousen, I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver Jr. and C.E. Wheatley, "On the Capacity of a Cellular CDMA Systems," *IEEE Transactions on Vehicular Technology*, Vol.40, No.2, pp.303-312, May 1991.
- [Git75] I. Gitman, "On the Capacity of Slotted ALOHA Networks and Some Design Problems," *IEEE Transactions on Communications*, Vol. COM-23, No.3, pp. 305-317, March 1975.
- [Goodm89] D.J. Goodman, R.A. Valenzuela, K.T. Gayliard and B. Ramamurthi, "Packet Reservation Multiple Access for Local Wireless Communications," *IEEE Transactions on Communications*, Vol. 37, No.8, pp.885-890, Aug. 1989.
- [Goodm90] D.J. Goodman, "Cellular Packet Communications," *IEEE Transactions on Communications*, vol.38, no.8, Aug. 1990, pp. 1272-1280.

- [Gril93] D. Grillo, M. Frullone, P. Graziaso, and P.Valigi, "A Performance Analysis of PRMA Considering Speech/Data Traffic, Co-Channel Interference and ARQ Error Recovery," *Proc. of 7th IEE Conf. on Mobile and Personal Communications*, Brighton,U.K., pp.167-172, Dec. 1993.
- [Grub82] J.G. Gruber, "A Comparison of Measured and Calculated Speech Temporal Parameters Relevant to Speech Activity Detection," *IEEE Transactions on Communications*, Vol. COM-30, No.4, pp.728-738, April 1982.
- [Grub83] J.G. Gruber and N.H. Le, "Performance Requirements for Integrated Voice/Data Networks," *IEEE Selected Areas in Communications*, Vol.SAC-1, No.6, pp.981-1005, Dec. 1983
- [GSM6.31] ETSI-GSM Technical Specifications, European Telecommunications Standard Institute, 1992.
- [GSM5.01] ETSI-GSM Technical Specifications, European Telecommunications Standard Institute, 1992.
- [Ham95] J. Hämäläinen, H. Jokinen, Z. Honkasalo and R. Fehlmann, "Multi-slot packet radio air interface to TDMA systems - Variable Rate Reservation Access (VRRRA)," *Personal, Indoor and Mobile Radio Communications (PIMRC'95)*, Sep. 27-29, Toronto, 1995.
- [Heff86] H. Heffes and D.M. Lucantoni, "A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance," *IEEE Journal in Selected Areas in Communications*, Vol. SAC-4, No.6, Sep 1986.
- [Hob83] W. Hoberecht, "A layered Network Protocol for Packet Voice and Data Integration," *IEEE Journal in Selected Areas in Communications*, Vol. SAC-1, No.6, Dec 1983, pp.1006-1013.
- [IbaK88] T. Ibaraki and N. Katoh, *Resource Allocation Problems Algorithmic Approaches*, MIT Press 1988.
- [Karim86] M.R. Karim, "Packet Communications on a Mobile Radio Channel," *AT&T Technical Journal*, Vol. 65 Issue 3, May/June 1986.
- [KarS94] M.J. Karol and S.C. Schwartz, "Multiple-Access Protocols: Fairness in Heterogeneous Systems," *IEEE Transactions on Communications*, Vol.42, No.6, pp. 2276-2281, June 1994,.
- [Kim83] B.G. Kim, "Characterization of Arrival Statistics of Multiplexed Voice Packets," *IEEE Selected Areas in Communications*, Vol.SAC-1, No.6, pp.1133-1139, Dec. 1983.

- [Klie75] L. Klienrock and S.S. Lam, "Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation," *IEEE Transactions on Communications*, Vol. COM-23, No.4, pp.410-422, Apr. 1975.
- [Kli75] L. Klienrock, *Queueing Systems*, Vol I and II, John Wiley and Sons Publication, 1975.
- [KliT75] L. Klienrock and F. Tobagi, "Packet Switching in Radio Channels: Part I-Carrier Sense Multiple-Access Modes and Their Throughput-Delay Characteristics," *IEEE Transactions on Communications*, Vol. COM-23, No. 12, Dec 1975, pp.1400-1416.
- [Kob77] H. Kobayashi, Y. Onozato and D. Huynh, "An Approximate Method for the Design and Analysis of an ALOHA System," *IEEE Transactions on Communications*, Vol. COM-25, No.1, pp 148-157, Jan 1977.
- [Kohn95] R. Kohn, R. Meidan and L.B. Milstein, "Spread Spectrum Access Methods for Wireless Communications," *IEEE Communications Magazine*, pp 58-67, Jan. 1995.
- [LamK175] S.S. Lam and L. Klienrock, "Packet Switching in a Multiaccess Broadcast Channel: Dynamic Control Procedures," *IEEE Transactions on Communications*, Vol. COM-23, No.9, pp 891-904, Sep 1975.
- [Lam77] S.S. Lam, "Delay Analysis of Time Division Multiple Access (TDMA) Channel," *IEEE Transactions on Communications*, Vol. COM-25, No.12, pp.1489-1494, Dec. 1977.
- [Lam80] S.S. Lam, "Packet Broadcast Networks - A Performance Analysis of the R-Aloha Protocol," *IEEE Transactions on Computers*, Vol.C-29, No.7, pp.596-602, Jul. 1980.
- [LeeUn86] H.H. Lee and C.K. Un, "A Study of On-Off Characteristics of Conversational Speech," *IEEE Transactions on Communications*, Vol.COM-34, No.6, pp. 630-637, June 1986.
- [LeeUn96] H.H. Lee and C.K. Un, "Performance Analysis of Nonpersistent Idle-signal Casting Multiple Access with collision-Detection (ICMA/CD) Protocol," *IEE Proc.-Commun.*, Vol. 143, No.3, pp. 162-166, June 1996.
- [Li87] V.O.K. Li, "Performance Evaluation of Multiple-Access Networks: Introduction and Issue Overview," *IEEE Selected Areas in Communications*, Vol.SAC-5, No.6, July 1987.
- [LiG95] I Chi-Lin and R.D. Gitlin, "Multi-Code CDMA Wireless Personal Communications Networks," in *Proc 1995 International Conference on Communications*, Seattle, Jun. 18-22,1995, pp.1060-1065.

- [LiMer94] F. Li and L. Merakos, "Voice/Data Channel Access Integration in TDMA Digital Cellular Networks," *IEEE Transactions on Vehicular Technology*, Vol.43, No.4, pp.986-996, Nov. 1994.
- [Linn94] J-P Linnartz, et al., "Throughput of Inhibit Sense Multiple Access with Propagation Delays," *IEEE Transactions on Communications*, Vol.42, No.1, Jan 1994, pp 119-126.
- [LiP93] M. Liu and P. Papantoni-Kazakos, "A Protocol for Cellular Radio Signalling Channels Carrying Data and High Priority Accessing Requests," *IEEE Transactions on Communications*, Vol. 41 No.4, pp.570-582, Apr. 1993.
- [Lu94] C.C. Lu and K.C. Chen, "A combined polling and random access protocol for integrated voice and data networks," *Proc. Personal, Indoor and Mobile Radio Communications (PIMRC'94)*, Sep. 18-23, The Hague, Netherlands, 1994.
- [Mad92] G. Madungwe and A.Bateman, "Transmission Strategies for Integrated Multichannel Packet Radio Networks," in *Proc. 1992 Vehicular Technology Conference*, pp. 883-887.
- [Mad91] G. Madungwe, et al. "A Protocol for Combining Voice and Data Traffic on a Trunked Multichannel PMR System," *IEE Sixth Int'l Conf on Mobile Radio and Personal Comms.*, Warkick, UK, Dec 9-11 1991, pp. 49-53.
- [Mah83] S.A. Mahmoud, et al, "An Integrated Voice/Data system for VHF/UHF Mobile Radio," *IEEE Journal in Selected Areas in Communications*, Vol. SAC-1, No.6, Dec 1983, pp 1098-1111.
- [Merak92] L. Merakos and S. Jangi, " Voice Packet Losses and Data Integration in Reservation random Access Protocols for Wireless Access Networks" *IEEE Globecom Conf. Record*, 1992, pp.26-31.
- [Mitro90] N.M. Mitrou, TH.D. Orinos and E.N. Protonotarios, "A Reservation Multiple Access Protocol for Microcellular Mobile-Communications Systems," *IEEE Transactions on Vehicular Technology*, Vol.39, No.4, pp.340-351, Nov. 1990.
- [Mitro93] N.M. Mitrou, G.L. Lyberopoulos and A.D. Panagopoulou, "Voice and Data Integration in the Air Interface of a Microcellular Mobile Communications Systems" *IEEE Transactions on Vehicular Technology*, Vol.42,No.1, Feb.1993, pp 1-13.
- [Mor84] L. Moraes and I. Rubin, "Message Delays for a TDMA Scheme Under a Non-preemptive Priority Discipline," *IEEE Transactions on Communications*, Vol. COM-32, No.5, May 1984, pp. 538-588.

- [MukF81] K. Mukumoto and A. Fukuda, "Idle Signal Multiple Access (ISMA) Scheme for Terrestrial Packet Radio Networks," *Transactions of IEICE* Vol.J64-B No. 10, pp.66-74, Oct. 1981 (in Japanese).
- [Muku90] K. Mukumoto, "Performance Evaluation of Mobile Packet Communications Networks by Using Transient Fluid Approximation Method," *PhD Dissertation*, Shizuoka University, Japan, Apr 1993 (in Japanese).
- [Nand92] S. Nanda, D.J. Goodman and U. Timor, "Performance of PRMA: A Packet Voice Protocol for Cellular Systems," *IEEE Transactions on Vehicular Technology*, vol.40, no.3, Aug.1992, pp. 584-598.
- [NiVP95] J.A.M. Nijhof, R.D. Vossenaar and R. Prasad, "Stack Algorithm in Mobile Radio Channels", in *Proc 1995 Vehicular Technology Conference*, pp.1193-1197.
- [Nom89] M. Nomura, T. Fujii and N. Ohta, "Basic Characteristics of Variable Rate Video Coding in ATM Environment," *IEEE Journal of Selected Areas in Communications*, Vol.7, No.5, Jun 1989, pp. 752-760.
- [Nom87] M. Nomura, T. Fujii and N. Ohta, "A Study on Bursty Signal Modelling in a Low Bit Rate Video Coding," *Technical Report of IEICE*, IN 87-39, pp 7-11, 1987.
- [PaKaz89] M. Paterakis and P. Papantoni-Kazakos, "A Simple Window Random Access Algorithm with Advantageous Properties," *IEEE Transactions on Information Theory*, Vol.35, No.5, pp.1124-1130, Sep. 1989.
- [Pkaz89] P. Papantoni-Kazakos, "Multiple Access Algorithms for a System with Mixed Traffic: High and Low Priority," in *Proc. International Conf. on Communications*, pp.21.3.1-21.3.8, Boston, Jun. 1989.
- [PlaL90] C. Van der Plas and J-P Linnartz, "Stability of Mobile Slotted ALOHA Network with Rayleigh Fading, Shadowing, and Near-Far Effect," *IEEE Transactions on Vehicular Technology*, Vol. 39, No.4, pp.359-366, Nov. 1990.
- [Pras96] R. Prasad, "CDMA for Wireless Personal Communications," Artech House, 1996.
- [PolSil87] A. Polydoros and J. Silvester, "Slotted Random Access Spread-Spectrum Networks" an Analytical Framework," *IEEE Journal in Selected Areas in Communications*, Vol. SAC-5, No.6, July 1987,
- [Purs87] M. Pursley, The Role of Spread Spectrum in Packet Radio Networks, *Proc. of the IEEE*, Vol.75, No.1 January 1997, pp. 116-134

- [QiWyr94] H. Qi and R. Wyrwas, "Markov Analysis for PRMA Performance Study," in *Proc. 1994 IEEE Vehicular Technology Conference*, Sweden, pp. 1184-1188.
- [QiWyr94b] H. Qi and R. Wyrwas, "Effects of Capture on the Stability and Performance of PRMA," in *Proc. IEEE Int'l Conf. on Universal Wireless Access*, Melbourne, Apr 1994, pp 175-178.
- [RACE95] *Proceedings of RACE Mobile Telecommunications Summit*, Cascais, Portugal, Nov. 22-24, 1995.
- [Raati94] K. Raatikainen "Symptoms of Self-Similarity in Measured Arrival Process of Ethernet Packets to a File Server," Report C-1994-4, Department of Computer Science, University of Helsinki, 1994.
- [Rab94] L. Rabiner, "Applications of Voice Processing to Telecommunications," *IEEE Australia Council Distinguished Lecturer Seminar*, Adelaide, 6 July, 1994.
- [Rait91] K. Raith and J. Uddenfeldt, "Capacity of Digital Cellular TDMA Systems," *IEEE Transactions on Vehicular Tech.*, Vol.40, May 1991. pp.323-332.
- [Rap91] S.S. Rappaport, "Modeling the hand-off problem in personal communication networks," in *Proc. 1991 Vehicular Technology Conference*, pp.517-523.
- [RomSi90] R. Rom and M. Sidi, "Multiple Access Protocols," Springer-Verlag, 1990, pp.20-37
- [Rob75] L.G. Roberts, "ALOHA packet systems with and without slots and capture," *Computer Communications Review*, Vol.5, No.2, Apr 1975, pp 28-42.
- [Rub79] I. Rubin, "Access-Control Disciplines for Multi-Access Communication Channels: Reservation and TDMA Schemes," *IEEE Transactions on Inf. Theory*, Vol. IT-25, No.5, Sep 1979, pp. 516-536.
- [RubT89] I. Rubin and Z-H Tsai, "Message Delay Analysis of Multiclass Priority TDMA, FDMA, and Discrete-Time Queueing Systems," *IEEE Transactions on Inf. Theory*, Vol. 35, No.3, May 1989, pp. 637-647.
- [Stall85] W. Stallings, *Data and Computer Communications*, 4th Ed., Maxwell Macmillan International, 1994.
- [Stav91] I. Stavrakakis, D. Kazakos, D., "A Multiuser Random-Access Communication System for Users with Different Priorities," *IEEE Transactions on Communications*, Vol. 39, No.11, pp.1438-1641, Nov. 1991.

- [Sheik90] A. Sheikh, Y-D Yao and X. Wu, "The ALOHA systems in Shadowed Mobile Radio Channels with Slow or Fast Fading," *IEEE Transactions on Veh. Tech.*, Vol.39, No. 4 Nov 1990, pp 289-298.
- [Srir83] K. Sriram, P. Varshney and J.G. Shanthikumar, "Discrete-Time Analysis of Integrated Voice/Data Multiplexers With and Without Speech Activity Detectors," *IEEE Selected Areas in Communications*, Vol.SAC-1, No.6, Dec. 1983, pp 1124-1132.
- [Steit95] J. Streit and L. Hanzo, "Wireless Videophone Schemes," in *4th International Conference on Universal Personal Communications Record*, Tokyo, Japan, Nov. 6-10, 1995, pp. 692-696
- [Stern90] H.P. Stern, "Design Issues Relevant to Developing an Integrated Voice/Data Mobile Radio System" *IEEE Transactions on Vehicular Tech.*, Vol.39, No.4, Nov.1990, pp 281-288.
- [Stern94] H.P. Stern, et al, "Modelling the On-Off Patterns in Conversational Speech, Including short Silence Gaps and the Effects of Interaction Between Speaking Parties," in *Proc 1994 Vehicular Technology Conference*, Stockholm, Jun 7-10, pp.1296-1300.
- [Tabb92] S. Tabbane and P. Godewlski, "Performance Evaluation of the R-BTMA Protocol in a Distributed Mobile Radio Network Context," *IEEE Transactions on Vehicular Technology*, Vol 41, No.1, pp.24-34, 1992.
- [Tanbm89] A. Tanenbaum, *Computer Networks*, Second Ed., Prentice-Hall, Inc.,1989.
- [Tas83] S. Tasaka, "Stability and Performamnce of the R-ALOHA Packet Broadcast System," *IEEE Transactions on Computers*, Vol.C-32, No.8, Aug 1983, pp 717-726.
- [Tob80] F. Tobagi, "Multiaccess Protocols in Packet Communication Systems," *IEEE Transactions on Communications*, Vol. COM-28, No.4, Apr 1980, pp. 468-488.
- [TobK75] F. Tobagi and L. Klienrock, "Packet Switching in Radio Channels: Part II-The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution," *IEEE Transactions on Communications*, Vol. COM-23, No. 12, Dec 1975, pp. 1417-1433.
- [Torr92] Don, Torrieri, *Principles of Secure Communications Systems*, Artech House 1992.
- [Tsyb85] B.S. Tsybakov, "Survey of USSR Contributions to Random Multiple-Access Communications," *IEEE Transactions on Information Theory*, Vol. IT-31, No.2, pp.143-165, Mar. 1985.

- [Tuck88] R. Tucker, "Accurate Method for Analysis of a Packet-Speech Multiplexer with Limited Delay," *IEEE Transactions on Communications*, vol.36, No.4, April. 1988, pp. 479-483.
- [Urie95] A. Urie, M. Streeton and C. Mourot, "An Advanced TDMA Mobile Access System for UMTS", *IEEE Personal Communications*, Feb. 1995, pp. 38-47.
- [Viter92] A.J. Viterbi and R. Padovani, "Implications of Mobile Cellular CDMA," *IEEE Communications Magazine*, Dec. 1992, pp. 38-41.
- [Vved94] N.D. Vvedenskaya, J.C. Arnbak, B.S. Tsybakov, "Improved Performance of Mobile Data Networks Using Stack Algorithms and Receiver Capture," in *Proc. International Zurich Seminar on Digital Communications*, pp. 464-475, Mar 8-11, 1994,
- [Wang94] G. Wang and N. Ansari, "Maximizing Data Throughput in an Integrated TDMA Communication System Using Mean Field Annealing," in *Proc Globecom'94*, pp 329-333.
- [Wat97] F. Watanabe et al, "Integrated Wireless Systems in Reserved Idle Signal Multiple Access with Collision Resolution," *IEICE Transactions Fund*, Vol.E80-A, No.7 pp.1263-1271, July 1997.
- [Wat95] F. Watanabe et al, "Load Sharing Sector Cells in Cellular systems," in *Proc. 1995 Personal Indoor Mobile Radio Conference*, pp 547-551.
- [Wies95] J. Wieselthier and A. Ephremides, "Fixed- and Movable- Boundary Channel-Access Schemes for Integrated Voice/Data Wireless Networks," *IEEE Transactions on Communications*, Vol. 43, No.1, pp.64-75, Jan. 1995.
- [Wils93] D. Wilson, et al., "Packet CDMA vs Dynamic TDMA for multiple Access in an Integrated Voice/Data PCN," *IEEE Journal on Selected Areas in Communications*, Vol.11 No.6, Aug 1993 pp 870-883.
- [Wu94a] G. Wu, "Performance Evaluation of Random Access Protocols for Wireless Personal Communication Networks," *PhD Thesis, Shizouka University, Japan*.
- [Wu94b] G. Wu, K. Mukumoto and A. Fukuda, " Analysis of an Integrated Voice and Data Transmission System Using Packet Reservation Multiple Access" *IEEE Transactions on Vehicular Technology*, Vol.43, No.2, May 1994, pp 289-297.
- [Wu94c] G. Wu, K. Mukumoto and A. Fukuda, " Performance Evaluation of Reserved Idle Signal Multiple-Access Scheme for Wireless Communications Networks" *IEEE Transactions on Vehicular Technology*, Vol.43, No.3, Aug. 1994, pp 653-658.

- [WuMF94] G. Wu, K. Mukumoto and A. Fukuda, "Slotted Idle Signal Multiple Access Scheme for Two-Way Centralized Wireless Communication Networks" *IEEE Transactions on Vehicular Tech.*, Vol.43, No.2, May 1994, pp 345-352.

- [Yue94] O-C Yue, " Achievable Multiplexing Gains for Wireless PCS with Packetized Speech," in *Proc. Globecom'94*, Nov. 28-Dec 2 1994, San Francisco, pp. 1311-1315.

Bibliography

- [1] Alanko et al, "Mobile computing based on GSM: The Mowgli approach", in *Proc. IFIP'96 (Mobile Communications)*, Sep 2-6, Canberra, Australia, pp. 151-158.
- [2] PG Andermo and G. Brismark, "CODIT, a Testbed Project Evaluating DS-CDMA for UMTS/FPLMTS," in *Proc. 44th Vehicular Technology Conf.*, June 7-10, 1994, Stockholm, pp.21-25.
- [3] Ananasso and F.D. Priscoli, "Technology Challenges in TDMA Approach to 3rd Generation Personal Communications Services, in *Proc. Globecom'94*, Vol. 2 pp. 702-706.
- [4] Benkner, T., " Dynamic Slot Allocation for TDMA-Systems with Packet Access", in *Proc. Workshop on Multiaccess, Mobility and Teletraffic for Personal Communications*, 20-22 May, 1996, Paris.
- [5] Borgonovo, F, et al "Capture-Division Packet Access for Wireless Personal Communications," *IEEE Journal on Selected Areas in Communications*, Vol.14, No.4, May 1996, pp.609-622.
- [6] Bruell, Steven C., "Computational algorithms for Closed Queueing Networks," Elsevier 1980.
- [7] Clymer, J., "Systems Analysis Using Simulation and Markov Models", Prentice-Hall, Inc 1990, pp. 169-179.
- [8] Chakraborty, S.S. , "The eXtended Time Division Multiple Access (XTDMA): A Multiple Access Scheme for Integrating Packet Video and Voice over the Cellular Air Interface," in *Proc. Sixth In'l. workshop on Packet Video*, Portland, Oregon, Sep 26-27, 1994.
- [9] Daigle, J.N. and Langford, J.D. "Models for Analysis of Packet Voice Communications Systems" *IEEE Journal on Selected Areas in Communications*, Vol. SAC-4, No.6 Sep. 1986, pp 847-855.
- [10] S.Deng, "Traffic Characteristics of Packet Voice", in *Proc. 1995 International Conference on Communications*, pp. 1369-1374.
- [11] ETSI STC/SMG2, "SMG2 Tdoc 258/94 "Variable Rate Reservation Access, air interface proposal for GPRS, Sophia Antipolis, France 19-21 December, 1994.
- [12] ETSI SMG2 GPRS Ad-hoc, "Evaluation Criteria for the GPRS Radio Channel," *Source : Motorola*", Uppsala, Sweden, Sep 12-14 1995.

- [13] ETSI SMG3/GPRS, "ETCS traffic input profile for the GPRS",
Source: UIC, 19-21 September 1995, Paris.
- [14] Falconer, D.D. Adachi, F. and Gudmunson, B. "Time Division Multiple Access Methods for Wireless Personal Communications" *IEEE Communications Magazine*, Jan.1995, pp 50-57.
- [15] Fayolle, G., Flajolet, P., Hofri, M and Jacquet, P., "Analysis of a Stack Algorithm for Random Multiple-Access Communication," *IEEE Transactions on Information Theory*, Vol IT-31, No.2, pp. 244-254, March 1985.
- [16] Georgiadis, L. and Papantoni-Kazakos, P., "Limited Feedback Sensing Algorithms for the Packet Broadcast Channel," *IEEE Transactions on Information Theory*, Vol IT-31, No.2, pp. 280-294, March 1985.
- [17] Geraniotis, Y-W Chang and W-B Yang, "Multi-Media Integration in CDMA Networks," in *Proc. IEEE 3rd International Symposium on spread Spectrum Technologies and Applications*, Vol. 1, pp 88-97.
- [18] Georgiadis, L., Merakos, L. and Papantoni-Kazakos, P., "A Method for the Delay Analysis of Random Multiple-Access Algorithms Whose Delay Process is Regenerative," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-5 No. 6 July 1987, pp. 1051-1061.
- [19] ETSI GSM Mobile Network Package Phase 1 Standards
- [20] Goodman, D. J., "Second Generation Wireless Information Networks," *IEEE Transactions on Vehicular Technology*, Vol.40, No.2, pp. 366-374, May 1991
- [21] Grillo, R.J. G. MacNamee, "Towards third generation mobile systems: a European possible transition path," in *Computer Networks and ISDN Systems*, 25 (1993), pp.947-961.
- [22] Hase et al, "R&D Project on Broadband Mobile Communications Using Microwave Band," in *Proc. Multi-Dimensional Mobile Conference*, Seoul, Korea, Sep. 18-20 1996, pp. 158-162.
- [23] Hofmann, J. and Metzner, N., "ATDMA- The Realtime Testbed Realisation," *RACE Mobile Telecommunications Summit*, Cascais, Portugal, Nov. 22-24, 1995, pp. 458-462.
- [24] Ketseoglou, T., "Certain Generalizations on the Slotted Collision Channel Without Feedback," in *Proc. International Zurich Seminar on Digital Communications*, pp. 464-475, Mar 8-11, 1994,
- [25] Kovacs, Laszlo Bela, *Combinatorial Methods of Discrete Programming*, Akademiai Kiado, Budapest 1980.

- [26] Langenbucher, "Efficient Coding and Speech Interpolation: Principles and Performance Characterisation" *IEEE Transactions on Communications*, Vol. COM-30, No.4, Apr. 1982, pp 769-779.
- [27] Lee, J.Y. and Un, C.K. "Performance analysis of an input and output queueing packet switch with a priority packet discarding scheme," *IEE-Proc-Communications*, Vol 142, No.2 April 1995.
- [28] Lindell, J. Skold, P. Willars and Erik Nilsson, "Radio Access Technology Evolution," in *Ericsson Review*, No.3, 1993, pp.83-92.
- [29] Mathys, P and Flajolet, P, "Q-ary Collision Resolution Algorithms in Random- Access Systems with Free of Blocked Channel Access," *IEEE Transactions on Informations Theory*, Vol. IT-31, No.2, pp.143-165, Mar. 1985.
- [30] Malkamaki, E., "Burst-Level ARQ - An Adaptive Low-Delay Error Protection Scheme for Speech Transmission in a TDMA System," in *Proc. of 7th IEE Conf. on Mobile and Personal Communications*, Brighton, U.K., pp.227-232, Dec. 1993.
- [31] Marafih, Y-Q Zhang and R.L. Pickhotz, "Modelling and Queueing Analysis of Variable-Bit-Rate Coded Video Source in ATM Networks," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 4 No. 2, pp. 121-128.
- [32] Mikhailov, V.A. , "Methods of Random Multiple Access", *Candidate Enginering Thesis, Moscow Institute of Physics and Technology*, Moscow, 1979.
- [33] Moore,D. and Rice,M. " Variable Rate Error Control for Wireless ATM Networks,' in *Proc. 1995 International Conf. on Communications*, pp.988-992.
- [34] Mushkin,M. and Bar-David,I., "Capacity and Coding for the Gilbert-Elliott Channels," *IEEE Transactions on Information Theory*, Vol.35, No.6, Nov 1989, pp 1277-1290.
- [35] Nanada,S., Goodman,D. and Timor,U., "Performance of PRMA: A PacketVoice Protocol for Cellular Systems," *IEEE Transactions on Vehicular Tech.*, Vol 40, No. 3, Aug 1991, pp 584-599.
- [36] Nikookar, H. and Prasad, R, "Performance Evaluation of Multi-Carrier Transmission Over Measured Indoor Radio Propagation Channels," in *Proc. 1995 Personal Indoor Mobile Radio Conf.*, Toronto, Canada, Sep. 27-29 1995, pp. 61-65.

- [37] Paterakis, M., Georgiadis, L. and Papantoni-Kazakos, P., "On the Relation Between the Finite and the Infinite Population Models for a Class of RRA's," *IEEE Transactions on Communications*, Vol.COM-35, No.11, pp.1239-1240, Nov. 1987.
- [38] Pax, T. Hirose and N. Kumahara, "Introduction of Multi-Rate Services into Digital Mobile Communication Systems," in *Proc. 1995 Personal Indoor Mobile Radio Conf.*, Vol. 3, pp 1083-1088.
- [39] "An Overview of the Application of Code Division Multiple Access (CDMA) to Digital Cellular Systems and Personal Cellular Networks," Copyright Qualcomm, March 21,1992
- [40] Ramamurthi, B., Saleh, A.A.M. and Goodman, D.J., "Perfect-Capture ALOHA for Local Radio Communications," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-5 No. 5 June 1987, pp. 806-814.
- [41] Rananand, "Approximating a Variable Bit Rate Source by Markov Process," in *Proc. Globecom 1994*, pp.1107-1112.
- [42] Sarker, J. and Hall, M. "A Novel (MULTI)3 Random Access Model in Packet Broadcast Networks, in *Proc. Multi-Dimensional Mobile Communications, July 18-20, 1996*, Seoul Korea.
- [43] Scholl, M and Klienrock, L. "On Mixed Mode Multiple Access Scheme for Packet-Switched Radio Channels", *IEEE Transactions on Communications*, Vol COM-27, no 6, June 1977, pp. 906-911.
- [44] Shacham, E.J. Craighill and A.A. Poggio, "Speech Transport in Packet Radio Networks with Mobile Nodes" *IEEE Journal on Selected Areas in Communications.*, Vol. SAC-1, No.6, Dec. 1983, pp 1084-1097.
- [45] SMG2 Tdoc 258/94, "Variable Rate Reservation Access, air interface proposal for GPRS", *Source: ETSI STC/SMG2*, Sophia Antipolis, France, Dec 1994.
- [46] Sourour, E. "Time Slot Assignment Techniques for TDMA Digital Cellular Systems," *IEEE Transactions on Vehicular Technology*, Vol 43, No 1 Feb 1994, pp. 121-127.
- [47] Tarkoy, F., "Information-Theoretic Aspects of Spread ALOHA," in *Proc. 1995 Personal Indoor Mobile Radio Conference*, Toronto, pp 1318-1320.
- [48] Tasaka,S., Kusunose, T. and Ohta, M. A , "Slotted ALOHA Channel Losing Slot Synchronization," *IEICE Transactions*, Vol. E62 No. 5 May 1979.

- [49] *A Course in Teletraffic Engineering*, Telecom Australia, 1978.
- [50] Ueda, H Tokizawa, I. and Aoyama, T. "Evaluation of an Experimental Packetized Speech and Data Transmission System," *IEEE Transactions on Communications.*, Vol. SAC-1, No.6 Dec. 1983, pp 1039-1045.
- [51] Viterbi, A.J. "The Evolution of Digital Wireless Technology from Space Exploration to Personal Communications Services," *IEEE Transactions on Vehicular Technology*, Vol. 43, No.3, pp.638-644, August 1994.
- [52] Wang,H.S. and Moayeri,N. "Finite-State Markov Channel-a Useful Model for Radio Communications Channels," *IEEE Transactions on Vehicular Technology*, Vol.44, No.1, Feb 1995, pp 163-171.
- [53] Wu, G., K. Mukumoto, A. Fukuda, M. Mizuno and K. Taira, " A Dynamic TDMA Wireless Integrated Voice/Data System with Data Steal into Voice (DSV) Technique," *IEICE Transactions on Communications*, Vol.E78-B, No.8, August 1995, pp 1125-1135.
- [54] Wu, G., and Mark, J.W., "Capacity Allocation for Integrated Voice/Data Transmission at a Packet Switched TDM," *IEEE Transactions on Communications*, Vol. 40, No.6, pp.1059-1069, June 1992.
- [55] Zhang and K. Pahlavan, "An Integrated Voice/Data System for Mobile Indoor Radio Networks" *IEEE Transactions on Vehicular Technology*, Vol.39, No.1, Feb. 1990, pp 75-82.
- [56] Zorzi, M. and Rao, R.R., "Slotted ALOHA with Capture in a Mobile Radio Environment," *Proc. International Zurich Seminar on Digital Communications*, pp. 453-463, Mar 8-11, 1994.