

Medium Access Control for Beyond Third Generation Heterogeneous Wireless Networks

by

Mo-Han Fong

Bachelor of Engineering, Nanyang Technological University, Singapore, 1993

Master of Applied Science, University of Victoria, Canada, 1995

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Electrical and Computer Engineering

© Mo-Han Fong, 2006
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by
photocopy or other means, without the permission of the author.

Medium Access Control for Beyond Third Generation Heterogeneous Wireless Networks

by

Mo-Han Fong

Bachelor of Engineering, Nanyang Technological University, Singapore, 1993

Master of Applied Science, University of Victoria, Canada, 1995

Supervisory Committee

Dr. T. Aaron Gulliver, Co-Supervisor (Dept. of Electrical and Computer Engineering)

Dr. Vijay K. Bhargava, Co-Supervisor (Dept. of Electrical and Computer Engineering)

Dr. Panajotis Agathoklis, Departmental Member (Dept. of Electrical and Computer
Engineering)

Dr. Frank Ruskey, Outside Member (Dept. of Computer Science)

Supervisory Committee

Dr. T. Aaron Gulliver, Co-Supervisor (Dept. of Electrical and Computer Engineering)

Dr. Vijay K. Bhargava, Co-Supervisor (Dept. of Electrical and Computer Engineering)

Dr. Panajotis Agathoklis, Departmental Member (Dept. of Electrical and Computer Engineering)

Dr. Frank Ruskey, Outside Member (Dept. of Computer Science)

Abstract

This thesis focuses on one of the key development areas of Beyond IMT-2000 or Beyond 3rd Generation (B3G) systems recommended by ITU-R M.1645, that is, new radio access systems that provide significantly higher performance for different deployment scenarios which may encompass different access technologies while maintaining seamless access and mobility from the user's perspective. Our objective is to develop various Medium Access Control (MAC) solutions for this new B3G access system.

We introduce a novel B3G multi-mode access system framework based on heterogeneous physical layer (PL) modes or configurations, anchored by a common layer 2 and layer 3 protocol stack. Such a system can support a wide variety of physical layer multiple access technologies to target different deployment scenarios and performance targets. The anchor or common layer 2/3 protocols enable seamless handoff and dynamic radio resource/load/spectrum management across the different PL modes to achieve optimum spectrum efficiency and Quality of Service (QoS) support.

As a basic form of the proposed B3G multi-mode access system and the first evolution step from the existing 3rd Generation (3G) cellular systems, we propose the multi-carrier expansion of Direct Spread Code Division Multiple Access, i.e. MC DS-CDMA. MC DS-CDMA supports concurrent transmissions on multiple DS-CDMA carriers anchored by a common layer 2/3 protocol stack. The common layer 2/3 protocol stack supports fully asymmetrical and dynamic Forward Link (FL) and Reverse Link (RL) carrier(s) allocation based on QoS requirements and Service Level Agreement (SLA). MC DS-CDMA also provides backward compatibility to existing single carrier DS-CDMA systems, thus allowing for overlay of legacy and new systems while the network deployment migrates towards B3G broadband support.

We further investigate load balancing schemes across multiple PL modes sharing the same spectrum resource in Time Division Multiplex (TDM), Frequency Division Multiplex (FDM) or Code Division Multiplex (CDM) fashions. For TDM and FDM cases, we propose a new Integrated Load Balancing and Scheduling (ILBS) scheme that maximizes the system capacity while meeting users' QoS and SLA. For the CDM case, we propose a dynamic Walsh code and Base Station (BS) transmit power sharing scheme between power-controlled dedicated traffic channels and rate-controlled packet data channels across multiple CDMA carriers.

An important aspect of MAC for wireless mobile systems is the MAC states management. We develop a universal MAC state machine concept that anchors the heterogeneous PL modes so that when a user switches from one PL mode to another, the MAC state and the associated context information of the MS can be retained, to minimize packet loss and PL mode switching/handoff latency.

We further look into the decision criteria used to transition a user from one MAC state to another. It is an important part of MAC for both the existing 3G systems and the B3G systems. The decision criteria should aim to maximize system capacity while meeting users' QoS and SLA requirements, while at the same time achieving power-saving. We propose a novel priority based state transition algorithm that achieves these objectives.

Overall, our research provides key solutions to the challenges of next generation wireless systems envisioned to encompass heterogeneous characteristics and performance targets.

Table of Contents

Abstract	iii
Table of Contents	vi
List of Tables	ix
List of Figures	x
Glossary	xiii
Acknowledgments	xvii
Dedication	xviii
Chapter 1 Introduction.....	1
1.1 Significance of Research	3
1.2 Thesis Outline	6
Chapter 2 Medium Access Control for Wireless Systems	8
2.1 Contention Based MAC Protocols	10
2.2 Contention Free MAC Protocols.....	13
2.3 Summary.....	15
Chapter 3 Wireless Cellular System Modeling.....	16
3.1 Overview of System Evaluation Methodology	16
3.1.1 Physical Layer Modeling.....	19
3.1.2 Layer 2/3 Modeling.....	20
3.1.3 Upper Layer Protocols Modeling.....	21
3.2 Traffic Modeling	21
3.2.1 Full Buffer	22
3.2.2 FTP	22
3.2.3 HTTP	23
3.2.4 WAP	26
3.2.5 NRTV	27
3.2.6 VoIP	29
3.3 Performance Metric	29
3.3.1 Performance Metrics for FTP, HTTP, WAP and NRTV	30
3.3.2 Performance Metrics for VoIP.....	32
3.4 Summary.....	33
Chapter 4 Beyond 3 rd Generation Multi-mode Wireless Access System.....	34
4.1 Multi-mode Access System Concept for B3G.....	34
4.1.1 Spectrum Allocation Scenarios of the Multimode Physical Layer	36
4.1.2 Hierarchical Cell Layout	39
4.2 Common Layer 2/3 Protocol Operation	41
4.2.1 Mobility Management.....	42
4.2.2 Radio Resource and Load Management.....	44
4.2.3 MAC State Control	45
4.2.4 System Access	46
4.3 An Example of a Specific Multi-mode Access System Configuration	46
4.4 Summary.....	49

Chapter 5 Multi-Carrier DS-CDMA with Dynamic and Service Driven Bandwidth Allocation.....	50
5.1 MC DS-CDMA System Description.....	52
5.1.1 System Overview	52
5.1.2 Service Driven Protocol Design.....	55
5.1.3 System Operation.....	56
5.1.4 Multi-carrier Scheduler	57
5.2 Performance Evaluation.....	64
5.2.1 Multi-carrier Diversity	64
5.2.2 Performance Comparison of MC-DV and 1xEV-DV	65
5.3 Summary	72
Chapter 6 Dynamic Load Balancing Across Physical Layer Modes	87
6.1 Load Balancing Schemes.....	88
6.1.1 TDM Multiplexing of PL modes	89
6.1.2 FDM Multiplexing of PL modes.....	91
6.1.3 CDM Multiplexing of PL modes	99
6.2 Performance Evaluation.....	102
6.2.1 Performance of Load Balancing with TDM and FDM of PL Modes	102
6.2.2 Performance of Load Balancing with CDM of PL Modes.....	104
6.3 Summary	108
Chapter 7 Universal MAC States.....	118
7.1 Overview of MAC States in 3G Systems	119
7.1.1 Active State.....	119
7.1.2 Dormant State	120
7.1.3 Intermediate States	121
7.1.4 State Transition	122
7.2 Universal MAC States Concept for B3G Multi-MODE Access System	123
7.2.1 Universal MAC States Definition	123
7.2.2 MAC State Operation during PL Mode Switching	124
7.3 Performance Evaluation.....	128
7.3.1 System Model	129
7.3.2 Traffic Model.....	130
7.3.3 Performance Metric.....	130
7.3.4 Performance Results.....	131
7.4 Summary.....	132
Chapter 8 MAC State Transition Algorithm.....	135
8.1 Priority Based MAC State Transition Algorithm.....	136
8.1.1 Generic Algorithm Description.....	137
8.1.2 Application of the Algorithm to the Multi-mode Access System.....	141
8.2 Performance Evaluation.....	142
8.2.1 System Level Simulation Configurations.....	142
8.2.2 Performance Comparison of State Transition Algorithms	144
8.2.3 Comparison of State Transition Frequencies.....	148
8.3 Summary	150
Chapter 9 Conclusion	163
9.1 Suggestions for Future Work.....	166

Bibliography 168
Appendix A Calculation of Number of Active State and Control-hold State Users..... 172

List of Tables

Table 1	System Level Simulation Parameters.....	17
Table 2	Fast Fading Channel Models	19
Table 3	FTP Traffic Model Parameters	23
Table 4	HTTP Traffic Model Parameters	25
Table 5	WAP Traffic Model Parameters	26
Table 6	NRTV Traffic Model Parameters.....	28
Table 7	VoIP Packet Sizes	29
Table 8	Fairness Criterion	32
Table 9	Relative Per-Carrier Sector Throughput Gain of MC-DV over 1xEV-DV...67	
Table 10	Summary of Normalized Per-Carrier System Capacity Gain of MC-DV (3x) over 1xEV-DV	72
Table 11	Performance of the Proposed Dynamic Load Balancing Scheme for Mixed Circuit Voice and Packet Data (Full Buffer Traffic) Across 3 Carriers.....	107
Table 12	TPR for R-FCH.....	173
Table 13	Different Combinations of Number of Active State Users and Number of Control-hold State Users	175

List of Figures

Figure 1	Protocol Stack for Wireless System.....	9
Figure 2	Cell Layout in System Level Simulation.....	18
Figure 3	Packet Trace in a Typical FTP Session.....	22
Figure 4	Packet Trace of a Typical Web Browsing Session.....	24
Figure 5	Contents of a Packet Call.....	24
Figure 6	Packet Trace for WAP Traffic Model.....	27
Figure 7	Packet Trace for NRTV Traffic Model.....	28
Figure 8	Anchor Layer 2/3 Protocol Design for Multimode Physical Layer.....	36
Figure 9	An Illustration of Spectrum Allocation Scenarios of the Multi-mode Physical Layer.....	38
Figure 10	Hierarchical Cell Layout.....	40
Figure 11	Common Layer 2/3 Protocol Functional Blocks.....	42
Figure 12	An Illustration of Multi-mode Access System Consists of 1xEV-DO and OFDMA overlay and OFDMA standalone physical layer modes.....	47
Figure 13	Dynamic Allocation of PL Modes on Different Carriers and Time Slots.....	48
Figure 14	Baseband Transmitter Block Diagram of MC-DV.....	53
Figure 15	MC-DV Carriers Configuration (An Illustration).....	54
Figure 16	An Illustration of Different Bandwidth Allocation Modes.....	55
Figure 17	Anchor Layer 2/3 Protocol Design for Different DS-CDMA modes: MC-DV, 1xRTT, IS-95.....	56
Figure 18	Global Multiplexing and Scheduling of 1xEV-DV and MC-DV Users at Time Slot n	60
Figure 19	An Illustration of the Fairness and Delay scheduler.....	63
Figure 20	Correlation Coefficient versus RMS Delay Spread for Carrier Separation of 1.25MHz.....	73
Figure 21	CIR Samples of Three Adjacent Carriers (F1, F2, F3), Each Separated by 1.25MHz, for Geometry of 6.4dB, RMS Delay Spread of $0.01\mu s$, and Time Slot Duration of 1.25ms.....	74
Figure 22	CIR Samples of Three Adjacent Carriers (F1, F2, F3), Each Separated by 1.25MHz, for Geometry of 6.4dB, RMS Delay Spread of $0.05\mu s$, and Time Slot Duration of 1.25ms.....	75
Figure 23	CIR Samples of Three Adjacent Carriers (F1, F2, F3), Each Separated by 1.25MHz, for Geometry of 6.4dB, RMS Delay Spread of $0.1\mu s$, and Time Slot Duration of 1.25ms.....	76
Figure 24	CIR Samples of Three Adjacent Carriers (F1, F2, F3), Each Separated by 1.25MHz, for Geometry of 6.4dB, RMS Delay Spread of $0.5\mu s$, and Time Slot Duration of 1.25ms.....	77
Figure 25	Sector Throughput per Carrier of MC-DV versus Number of Carriers, for Full Buffer Traffic with Mixed Mobile Speeds and 20 Users per Sector.....	78
Figure 26	CDFs of Normalized Per User Throughput of 1xEV-DV (1 carrier), MC-DV with 3 Carriers and MC-DV with 5 Carriers.....	79

Figure 27	Scatter Plot of User Throughput Normalized to Per-Carrier versus Geometry for 1xEV-DV (1 Carrier) and MC-DV (3 Carriers).....	80
Figure 28	Sector Throughput Per Carrier versus Number of FTP Users for 1xEV-DV and MC-DV with 3 Carriers	81
Figure 29	System Outage versus Number of FTP Users Per Carrier for 1xEV-DV and MC-DV with 3 Carriers	82
Figure 30	System Outage and Sector Throughput Per Carrier versus Number of HTTP Users for 1xEV-DV and MC-DV with 3 Carriers.....	83
Figure 31	System Outage versus Number of Video Users (Streaming Video with Source Rate of 32kbps and Delay Bound Criterion of 5 Seconds) for 1xEV-DV and MC-DV with 3 Carriers.....	84
Figure 32	System Outage versus Number of Video Users (Real-Time Interactive Video with Source Rate of 32kbps and Delay Bound Criterion of 100ms) for 1xEV-DV and MC-DV with 3 Carriers	85
Figure 33	Sector Throughput Per Carrier for Streaming Video and Real-Time Interactive Video for 1xEV-DV and MC-DV with 3 Carriers	86
Figure 34	Example of Dynamic Spectrum Sharing between PL Modes: MC-DV and OFDMA.....	93
Figure 35	Example of Dynamic Spectrum Sharing between PL Modes: Different OFDMA modes for fixed, medium speed and high speed users, unicast and multicast services.....	94
Figure 36	Instantaneous BS Transmit Power Consumed by Circuit Voice Users for 50% System Loading of EVRC RC3 Users, and 20% of BS Transmit Power Consumed by Overhead Channels	110
Figure 37	Performance Comparison of Load Balancing Schemes for a System Consists of Two PL modes	111
Figure 38	Performance Comparison of Load Balancing Schemes for a System Consists of Four PL Modes	112
Figure 39	Data Sector Throughput versus Percentage of Voice Loading for Mixed 1xRTT EVRC Voice Traffic and 1xEV-DV Full Buffer Data Traffic	113
Figure 40	Voice Outage versus Percentage of Voice Loading for Mixed 1xRTT EVRC Voice Traffic and 1xEV-DV Full Buffer Data Traffic	114
Figure 41	Data Sector Throughput and Data User Outage versus Percentage of Voice Loading, for Mixed 1xRTT EVRC Voice Traffic and 1xEV-DV FTP Data Traffic	115
Figure 42	Data Sector Throughput per Carrier and FTP User Outage versus Number of FTP Users per Carrier at 17% Voice Loading per Carrier, for 1xEV-DV and MC-DV with 3 Carriers	116
Figure 43	Data Sector Throughput per Carrier and FTP User Outage versus Number of FTP Users per Carrier at 47% Voice Loading per Carrier, for 1xEV-DV and MC-DV with 3 Carriers	117
Figure 44	MAC States of a MS in 3G Systems	119
Figure 45	Universal MAC States (An Illustration of the Subset of MAC States Supported by each PL Mode)	124
Figure 46	Algorithm for MAC State Transition during PL Mode Switching	125
Figure 47	Comparison of User Capacity for Different MAC States Scenarios.....	133

Figure 48	Comparison of Sector Throughput for Different MAC States Scenarios.....	134
Figure 49	Comparison of User Capacity for Different MAC States Transition Algorithm Scenarios with FTP Traffic.....	152
Figure 50	Comparison of Sector Throughput for Different MAC States Transition Algorithm Scenarios with FTP Traffic.....	153
Figure 51	Comparison of User Capacity for Different MAC States Transition Algorithm Scenarios with HTTP Traffic.....	154
Figure 52	Comparison of Sector Throughput for Different MAC States Transition Algorithm Scenarios with HTTP Traffic.....	155
Figure 53	Comparison of User Capacity for Different MAC States Transition Algorithm Scenarios with WAP Traffic.....	156
Figure 54	Comparison of Sector Throughput for Different MAC States Transition Algorithm Scenarios with WAP Traffic.....	157
Figure 55	Comparison of User Capacity for Different MAC States Transition Algorithm Scenarios with Mixed FTP/HTTP/WAP Traffic	158
Figure 56	Comparison of Sector Throughput for Different MAC States Transition Algorithm Scenarios with Mixed FTP/HTTP/WAP Traffic	159
Figure 57	Comparison of MAC States Transition Rate for Scenario 1B (Three-State Timer-Based State Transition) and Scenario 2B (Three-State Priority Based State Transition), where D-to-A is Dormant to Active Transition, CH-to-D is Control-hold to Active Transition, A-to-CH is Active to Control Hold Transition, CH-to-A is Control-hold to Active Transition.....	160
Figure 58	Comparison of MAC States Transition Rate for Scenario 2B (Three-State Priority Based State Transition) with State Transition Priority Update Rate of 800Hz and 8Hz.	161
Figure 59	Comparison of System Outage and Sector Throughput for Scenario 2B (Three-State Priority Based State Transition) with State Transition Priority Update Rate of 800Hz and 8Hz.....	162

Glossary

1xEV-DO	1x Evolution Data Only
1xEV-DV	1x Evolution Data and Voice
1xRTT	1x (Single Carrier) Radio Transmission Technology
2G	2 nd Generation
3G	3 rd Generation
AG	Access Gateway
ARQ	Automatic Retransmission Request
B3G	Beyond 3 rd Generation
BS	Base Station
BSC	Base Station Controller
BTS	Base Transceiver Subsystem
CDM	Code Division Multiplex
CIR	Carrier-to-Interference Ratio
C/P	Channel Quality Indicator Channel to Pilot Ratio
CQI	Channel Quality Indicator
CSMA	Carrier Sense Multiple Access
CTS	Clear to Send
DCF	Distributed Coordination Function
ESN	Electronic Serial Number
EVRC	Enhanced Variable Rate Codec
FDM	Frequency Division Multiplex

FDMA	Frequency Division Multiple Access
FER	Frame Error Rate
FL	Forward Link
F-PDCH	Forward Packet Data Channel
GSM	Global System for Mobile Communications
HARQ	Hybrid Automatic Retransmission Request
HSDPA	High Speed Downlink Packet Access
HS-DSCH	High Speed Downlink Shared Channel
ILBS	Integrated Load Balancing and Scheduling
IMSI	International Mobile Station Identity
IP	Internet Protocol
MAC	Medium Access Control
MACA	Multiple Access with Collision Avoidance
MC CDMA	Multi-Carrier Code Division Multiple Access
MC-DO	Multi-Carrier Data Only
MC DS-CDMA	Multi-Carrier Direct Spread Code Division Multiple Access
MC-DV	Multi-Carrier Data and Voice
MC-HSDPA	Multi-Carrier High Speed Downlink Packet Access
MS	Mobile Station
NRTV	Near Real Time Video
OFDMA	Orthogonal Frequency Division Multiple Access
PCB	Power Control Bit
PCF	Point Coordination Function

PL	Physical Layer
PN	Pseudo-random
PPP	Point-to-Point Protocol
QoS	Quality of Service
R-CCCH	Reverse Common Control Channel
R-CQICH	Reverse Channel Quality Indicator Channel
R-EACH	Reverse Enhanced Access Channel
R-FCH	Reverse Fundamental Channel
RL	Reverse Link
RLC	Radio Link Control
RLP	Radio Link Protocol
ROT	Rise Over Thermal
R-PICH	Reverse Pilot Channel
RS	Relay Station
RTP	Real-Time Protocol
RTS	Request to Send
SLA	Service Level Agreement
SNR	Signal-to-Noise Ratio
TCP	Transport Control Protocol
TDM	Time Division Multiplex
TDMA	Time Division Multiple Access
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunications System

VoIP	Voice over Internet Protocol
WINNER	Wireless World Initiative New radio
WLAN	Wireless Local Area Network
WWRF	Wireless World Research Forum

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisors, Dr T. Aaron Gulliver and Dr. Vijay K. Bhargava, for giving me the invaluable opportunity to pursue my Ph.D. studies at the University of Victoria, under their supervision. Your continuing support and guidance have allowed me to develop my potential and explore new horizons.

My appreciation is also extended to my colleagues at Nortel Networks, for all the fruitful discussions and input, which make my work directly applicable to the on-going standards and technology evolution in the wireless telecommunications industry.

I would also like to thank Dr. Charles Despins for his on-going support, encouragement and tireless proofreading of my papers. All of which have led me through challenging times.

Finally, my heartfelt gratitude goes to my family, especially to my parents Mr. Fong, Chung and Ms. Mark, Kat Yan. Thanks for your unfailing support, encouragement and respect for my chosen path in the pursuit of my dreams. Your love and care have helped me go through each step of my life. You guys are the greatest.

Dedication

~ To my parents ~

for all the sacrifices you have made for me

Mom,

for your selfless and tender loving care that always warms my heart

Dad,

for the strength and inspiration you give to my life

I love you!

Chapter 1

Introduction

The commercial success of wireless and mobile communications services has resulted in a wide variety of wireless networks being deployed around the world, such as Wi-Fi, 2nd Generation (2G) and 3rd Generation (3G) cellular, satellite and Bluetooth. These wireless networks differ in their air-interface designs and access network architectures due to their respective markets, deployment requirements and performance targets. For example, wireless LANs such as 802.11a/g [21][23] provide fixed/nomadic access with cell coverage of less than 100 meters and peak data rates of 54Mbps. Existing 3G cellular systems provide mobile access with cell coverage of a few kilometres and peak data rates of 3Mbps over a 1.25 MHz bandwidth for 1xEV-DO [1], and 14Mbps over 5 MHz for UMTS HSDPA [2]. These existing wireless networks are being deployed as disjoint networks with no or minimum inter-working between them at the network layer. Thus seamless mobility and universal service offerings for the end users are not possible.

Next generation wireless systems or so-called Beyond Third Generation (B3G) wireless systems or Beyond IMT-2000 systems [3] are presently the subject of major research programs around the world, such as the Wireless World Research Forum (WWRF) and the Wireless World Initiative New Radio (WINNER). B3G wireless systems are envisioned to provide ubiquitous and seamless inter-working across a wide variety of wireless access technologies [3][4]. ITU-R approved the Recommendation ITU-R M.1645 “Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000” [3] in June 2003 as a basis for B3G wireless

systems development focus. The two key development areas highlighted in [3] are: 1) integration and inter-working of existing and evolving wireless systems to provide global roaming and inter-operability for end-users; 2) new radio access systems that provide significantly higher performance for different deployment scenarios, i.e. data rates up to 100Mbps for mobile access and 1Gbps for nomadic/local wireless access, which may encompass different access technologies while maintaining maximum commonality among them for seamless access and mobility from the user's perspective.

Thus, the B3G wireless systems will encompass existing 2G/3G radio access systems as well as new B3G radio access systems. B3G systems will encompass fixed and mobile networks; pico-, micro- and macro-cellular networks; relay, mesh and adhoc networks; broadcast/multicast and unicast networks etc. The goal is to achieve universal and ubiquitous wireless access with a wide variety of service offerings. In addition, B3G systems should support dynamic radio resource, load and spectrum management across heterogeneous access systems, in order to maximize spectrum efficiency while meeting the Quality of Service (QoS) requirements of user applications and the user Service Level Agreements (SLAs).

Some of the key challenges to B3G system design include mobility management to provide seamless and context-aware handoff across different networks (i.e. intersystem or vertical handoff [6]), network discovery and access, and dynamic radio resource, load and spectrum management across heterogeneous systems.

To date, B3G research in the literature [5][6] has been focusing on inter-working between existing and evolving access systems with completely different air-interface protocols (physical layer, layer 2 and layer 3), through convergence at the network or IP

layer. In such scenarios, to support seamless mobility across the different access systems, a new and common set of handoff and context transfer signaling/protocols need to be defined across different systems at the network/IP layer. The common signaling set provides high-level context information transfer, such as security, service flows and associated QoS information transfer. Detailed access system dependent context, such as ARQ states, Medium Access Control (MAC) protocol states and layer 2 mobility management cannot be transferred across the heterogeneous access systems. Thus, air-interface packet losses and connection setup latency cannot be avoided during inter-system handoff. In addition, dynamic radio resource and load management across heterogeneous access systems is deemed difficult.

1.1 Significance of Research

Our research focuses on the second development area highlighted in ITU-R M.1645, with the objective of developing MAC solutions for a new B3G access system that support seamless mobility and handoff, dynamic and efficient radio resource, load, spectrum management across heterogeneous radio access technologies. To achieve this objective, we put forward the following contributions in this thesis:

1. A novel B3G access system concept based on heterogeneous physical layer (PL) modes with a common layer 2/3 protocol stack;
2. A Multi-Carrier Direct Spread Code Division Multiple Access (MC DS-CDMA) system concept enabling a backward compatible and smooth network migration strategy towards B3G broadband access;
3. An investigation and definition of novel load balancing schemes across multiple PL modes of the B3G access system;

4. The development of a universal MAC state machine concept anchoring the heterogeneous PL modes;
5. The definition of a novel priority based MAC state transition algorithm to maximum system capacity while meeting users' QoS and SLA.

The novel B3G access system concept that we propose is based on heterogeneous PL modes or configurations, with a common (or anchor) layer 2 and layer 3 protocol stack. Such a system can support a wide variety of physical layer access technologies, e.g. CDMA, Orthogonal Frequency Division Multiple Access (OFDMA), Multi-Carrier CDMA (MC-CDMA) etc, with different configurations to target different deployment scenarios and performance targets, while anchored by common layer 2 and layer 3 protocols. Seamless handoff is enabled by the common layer 2 and layer 3 protocols, while dynamic radio resource/load/spectrum management can be performed across the different PL modes to achieve optimum spectrum efficiency and QoS support. Such a B3G access system concept is complimentary to the IP layer convergence approach which is required when the new B3G access system inter-works with existing 2G/3G heterogeneous access systems.

As the first evolution step towards the proposed B3G multi-mode access system framework, we propose the MC DS-CDMA system concept. MC DS-CDMA supports concurrent transmissions on multiple DS-CDMA carriers, and each carrier transmits an independent DS-CDMA waveform. MC DS-CDMA is a fundamentally different concept than the straightforward bandwidth expansion of current CDMA systems in several aspects, including service-driven dynamic and scalable radio bandwidth allocation that is fully asymmetrical between the Forward Link (FL) and the Reverse Link (RL), and fast

multi-carrier scheduling. MC DS-CDMA can provide backward compatibility to existing single carrier DS-CDMA systems, thus allowing for the overlay of existing single carrier DS-CDMA systems with MC DS-CDMA systems while the network deployment migrates towards B3G broadband support.

We further investigate the load balancing schemes across multiple PL modes of the B3G multi-mode access system. The different PL modes can be Time Division Multiplexed (TDM), Frequency Division Multiplexed (FDM) or Code Division Multiplexed (CDM) on the shared spectrum resource. For the TDM and FDM cases, we propose a new Integrated Load Balancing and Scheduling (ILBS) scheme that dynamically allocates radio resources to different PL modes based on packet scheduling priority. The goal is to maximize the system capacity while meeting users' QoS and SLA requirements. For the CDM case, we propose a dynamic Walsh code and Base Station (BS) transmit power sharing scheme between power-controlled dedicated traffic channels and rate-controlled packet data channels across multiple CDMA carriers.

One big part of the MAC for cellular systems is the MAC states which are different operational states of a Mobile Station (MS) based on traffic activities. The purposes of MAC states are power saving and improved resource utilization to support packet data services. For the proposed B3G system, we develop a universal MAC state machine concept that anchors the heterogeneous PL modes such that when a user switches from one PL mode to another, either within the same or different cell hierarchies, the MAC state and the associated context information of the MS can be retained. In this way, packet loss and PL mode switching/handoff latency can be minimized.

The decision criteria used to transition a user from one MAC state to another is an important aspect of MAC states management for both the existing 3G system and the B3G system. The decision criteria should aim to maximize system capacity while meeting users' QoS and SLA requirements, and at the same time achieve power-saving. We propose a novel priority based state transition algorithm that achieves these objectives.

Overall, our research provides key solutions to the challenges of next generation wireless systems envisioned to encompass heterogeneous characteristics and performance targets that serve a wide variety of deployment scenarios and service offering. The outcome of our research and the new ideas proposed will contribute to the system design, implementation and standardization of the next generation or B3G wireless systems.

1.2 Thesis Outline

This thesis consists of nine chapters. In Chapter 2, an overview of MAC protocols for wireless systems is given. This includes the description and rationale of different MAC protocols for Wi-Fi, 2G cellular and 3G cellular systems.

In Chapter 3, we describe the details of cellular system modeling in order to evaluate the performance of various proposed schemes in this thesis.

Chapter 4 proposes the overall B3G multi-mode wireless access system framework. Different aspects of the proposed framework are presented, i.e. service-driven protocol structure, spectrum allocation across different PL modes, support of hierarchical cells, mobility management, radio resource and load management, MAC state control and system access. A specific configuration of a B3G multi-mode access system is also given as an illustration.

In Chapter 5, we introduce the MC DS-CDMA system concept as the first evolution step towards the proposed B3G multi-mode wireless access system. A cross-carrier layer 2 scheduler was proposed to optimize system capacity while ensuring the QoS of different services are met. We conduct detailed performance studies to evaluate the performance of MC DS-CDMA in the form of Multi-Carrier Data and Voice (MC-DV) MC-DV, a multi-carrier extension of the existing 1xEV-DV system [12][13].

Chapter 6 investigates the load balancing schemes across the different PL modes of the proposed B3G system. Dynamic load balancing schemes for TDM, FDM and CDM of PL modes are proposed with detailed performance studies to quantify the performance benefits.

In Chapter 7, we introduce the universal MAC states concept for the proposed B3G multi-mode access system. A detailed account of different MAC state transition scenarios when a user moves from one PL mode to another is given, followed by a detailed performance studies to demonstrate the performance advantage of the proposed scheme.

Chapter 8 explores the algorithms used to transition a user from one MAC state to another. The conventional timer-based state transition algorithm is first described, followed by the proposal of a novel priority based state transition algorithm applicable to both the existing 3G cellular systems as well as the proposed B3G multi-mode access system. To demonstrate the performance benefits of the proposed algorithm, we conduct detailed performance studies of the proposed algorithm and compare the results with the performance of the conventional algorithm.

Finally, conclusions and suggestions for future work are given in Chapter 9.

Chapter 2

Medium Access Control for Wireless Systems

Wireless transmission medium is naturally a shared medium accessed by multiple users. The common multiple access techniques used today are Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA) used in GSM [7] and IS-136 [8]; CDMA used in 1xRTT [9][10][11] and UMTS [14] and OFDMA used in 802.16e [15]. These multiple access techniques at the physical layer divide the wireless transmission medium into multiple parallel physical channels, where the physical channels have minimum mutual interference to one another. The data rates of the physical channels can either be fixed or variable respectively for homogeneous traffic support such as in voice-only systems, e.g. IS-95 [16]; and heterogeneous traffic support such as in voice and variable rate data systems, e.g. 1xRTT.

To control the use of the physical channels by multiple users, a MAC protocol is defined. The MAC protocol is part of the layer 2 protocol of the wireless protocol stack as shown in Figure 1. The MAC protocol is responsible for determining how physical channels resource is shared among users in the system. This includes multiplexing of users' traffic to the available physical channels resource, maintaining users' required QoS and SLA, controlling the fairness of resource sharing among users, and performing call admission control of users into the system. In an actual wireless system realization, the MAC protocol can be distributed to control the physical channels resource of a BS or centralized to control the physical channels resource across multiple BSs. In the case of centralized MAC, the MAC protocol performs load balancing and mobility/handoff

resource management across BSs. A typical system realization employs a combination of distributed and centralized MAC.

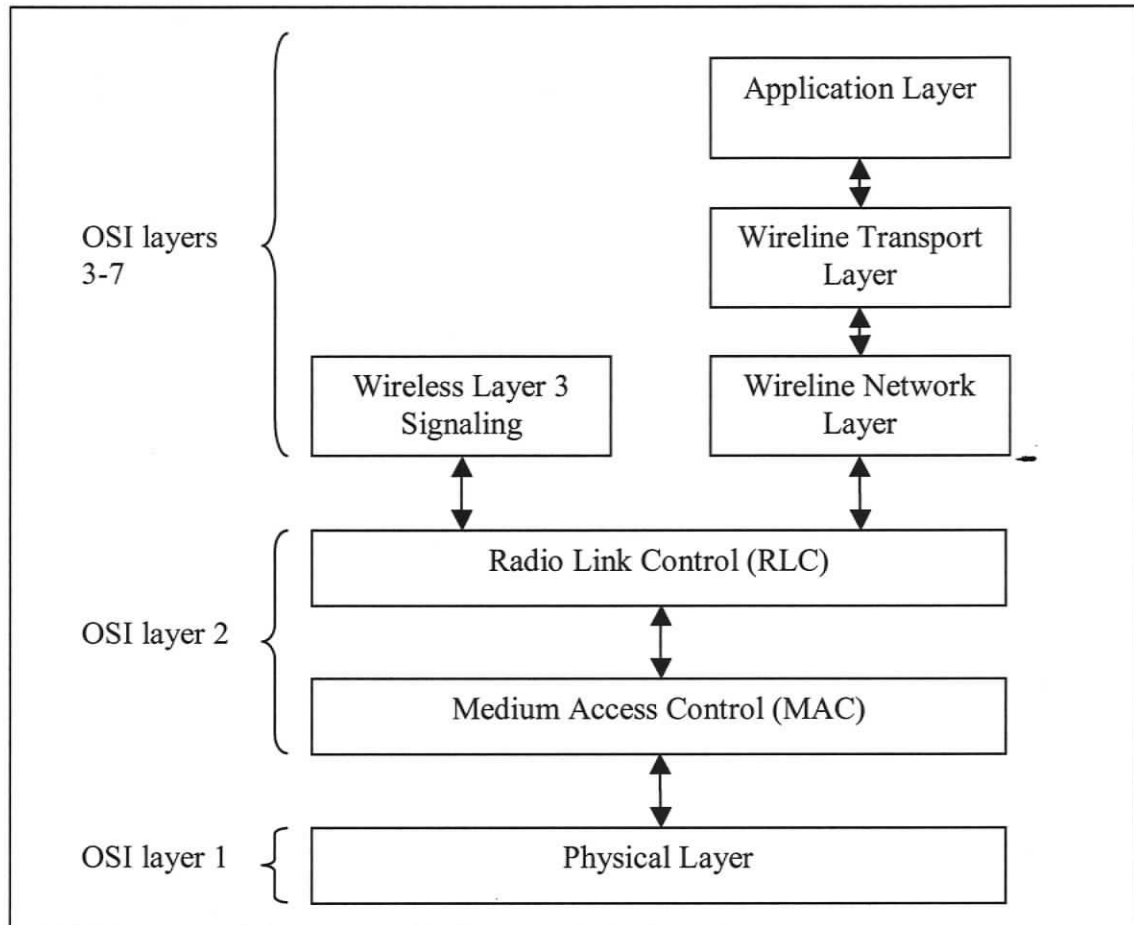


Figure 1 Protocol Stack for Wireless System

There are different variants of MAC protocols aimed for systems with different service offerings, coverage or range, implementation complexity and spectrum availability. MAC protocols can be broadly categorized into contention based protocols and contention free protocols as described in the following sections.

2.1 Contention Based MAC Protocols

Contention based MAC protocols are simpler than contention free MAC protocols since no coordination is required between transmitting stations and a minimal amount of centralized coordination is required at the BS. The simplest form of contention based protocols are the random access protocols, such as ALOHA and slotted ALOHA schemes [19][20]. Due to the free-for-all nature of these schemes, collisions cannot be avoided. When a collision occurs, the transmitting station employs random backoff before retransmitting a frame. A pure ALOHA system with a constant frame size has a maximum bandwidth utilization of 18% whereas a slotted ALOHA system can achieve a bandwidth utilization of about 37%. Such low bandwidth utilization is due to the lack of collision avoidance techniques.

To avoid or reduce collision probability, Carrier Sense Multiple Access (CSMA) techniques are introduced. A station in a CSMA system only transmits when it senses that there is no existing transmission from other users. There are several variants of the CSMA scheme based on the action taken by a station when it detects an idle channel. Those variants include 1-persistent CSMA, p-persistent CSMA, non-persistent CSMA and CSMA with random backoff [19][20]. The CSMA protocols have a much higher bandwidth utilization than the ALOHA or slotted ALOHA protocols. The bandwidth utilization can approach 100% if transmission delay is not a concern. A further enhancement to the CSMA protocol is the combination of CSMA and collision detection, i.e. CSMA/CD. In this case, if a station detects a collision in the middle of transmission, it aborts the transmission, thus freeing up the channel for other stations. We can see that the above CSMA protocols require stations to be able to detect each other's transmission.

Therefore, CSMA protocols are suitable for systems with short range coverage such as Wireless Local Area Network (WLAN) and systems that have FL and RL transmission sharing the same spectrum such that a station's transmitter and receiver are locked onto the same carrier frequency. The most typical of such a system is the 802.11 or Wi-Fi system such as 802.11a [21], 802.11b [22] and 802.11g [23]. The distributed coordination function (DCF) of 802.11 employs the CSMA/CD with binary exponential backoff, which is also called the CSMA with Collision Avoidance or CSMA/CA.

Even with a short coverage range, there is no guarantee that a station can detect the transmission of all other stations to the BS. This is called the 'hidden station' problem. To combat this problem, a Multiple Access with Collision Avoidance (MACA) is used where a station sends a short Request to Send (RTS) frame to the destination or the BS. The destination or the BS responds with a Clear to Send (CTS) frame. The station then follows with a data frame transmission. Since the CTS frame will be detected by other stations, the RTS/CTS handshake serves as a channel reservation mechanism to avoid collision from a 'hidden station'. Such a scheme is defined in 802.11 as an optional feature of the DCF, where the RTS frame is sent using CSMA/CA scheme.

As opposed to CSMA, MACA does not rely strictly on a station detecting transmission of another station except for the initial short request; therefore, it is suitable for use in large coverage cellular systems. For many of today's cellular systems such as 1xRTT, UMTS, 802.16e, RL transmission, when no dedicated physical channel is assigned, uses a combination of random access (slotted ALOHA) and random access for sending a short request followed by either detailed request or data frame transmission on dedicated physical channels assigned by the BS. For example, in a 1xRTT system (TIA/EIA-2000-

A [10] and subsequent releases), two access modes are defined for call setup when the MS has no RL dedicated physical channel assigned. In the Basic Access Mode, a MS uses a random access protocol to send the access message on the Reverse Enhanced Access Channel (R-EACH). In the Reservation Access Mode, a MS uses a random access protocol to send a very short reservation request on the R-EACH, followed by a long access message on the Reverse Common Control Channel (R-CCCH) assigned by the BS. In 802.16e, during call setup access or RL bandwidth request when no RL dedicated channel is assigned, a MS transmits a randomly selected pseudorandom (PN) code from a predefined code set on a predefined contention-based OFDMA resource. Collision occurs when there is more than one MS transmitting the same PN code or when the number of MSs transmitting PN codes on the same contention based OFDMA resource exceeds a certain number such that the multi-user interference level becomes higher than that tolerable by the BS receiver. When a BS successfully detects the PN code transmitted by a MS, the BS assigns a dedicated physical channel resource to the MS to transmit a subsequent access message or a bandwidth request message.

Overall, contention based schemes are simpler to implement due to the minimal amount of coordination required among stations and at the BS. However, contention based schemes are mainly suitable for best effort traffic since there is no guarantee that a station can transmit its data with the desired data rate and latency due to collision with other stations. Contention based schemes, however, can be enhanced to support differentiated QoS by setting different random backoff time and allowable transmission time to different classes of stations. Such enhancements are introduced in 802.11e

[24][25][26]. To support assured QoS, a contention free MAC as described in the next section has to be used.

2.2 Contention Free MAC Protocols

To avoid contention and collision, a centralized controller such as the BS assigns one or more dedicated physical channels to a MS for certain duration. Within this duration, no other MSs are allowed to transmit on the same physical channel(s). The assignment can be semi-static or dynamic. A semi-static assignment of dedicated physical channel(s) is typically performed during call setup and the assignment is retained during the entire call duration. It is suitable for circuit-oriented services such as circuit voice and circuit data where the data rate and packet arrival are fixed or vary slowly during the call. Semi-static assignment is employed in 2G cellular systems such as IS-95/A/B [16][17][18] and GSM [7]. The call admission policy is based on users' SLA and system loading. For dynamic assignment, the physical channels resource is dynamically allocated to and de-allocated from a MS during a call. Dynamic assignment allows flexible sharing of the physical channels resource among users in the system based on traffic activity, QoS and SLA of each user. It is typically used to support packet-oriented data services with bursty traffic that comprise most of today's internet based services. Dynamic assignment is employed in 3G cellular systems such as 1xRTT and UMTS.

For FL transmission from a BS to a MS, contention free MAC is easily achieved by the BS deciding when and which physical channel(s) are used to transmit data to a MS. For RL transmission, contention free MAC can be achieved by different methods. In 802.11 [21][22][23], a point coordination function (PCF) at the BS polls one station at a time for RL transmission. A polled station responds with RL data transmission if there is data to

send. If no data transmission is detected from the polled station, the BS polls the next station. A further enhancement is introduced in 802.11e [24] where the polling mechanism of a BS is driven by the QoS requirement of a station's traffic streams. In 3G cellular systems such as 1xRTT and UMTS, contention free RL MAC is used after a MS is admitted to the system through the call setup procedure, i.e. when a MS is in the Active state. More details of the operational states of a MS in 3G cellular systems will be addressed in Section 7.1. A MS in the Active state is assigned a low data rate RL dedicated physical channel, which is used for transmitting low rate data traffic as well as signalling messages associated with RL resource request and other mobility management related control signalling. The RL resource request signalling is used by a MS to request additional RL physical channel resource in a contention free manner on an as-needed basis. In 802.16e, there is no low data rate dedicated physical channel assigned to a user in the Active state. The RL resource request is performed in a contention-based manner as previously described in Section 2.1. The reason for the different approaches used in 1xRTT/UMTS and 802.16e is the different multiple access techniques employed by these systems. In 1xRTT and UMTS which are based on CDMA, a code-multiplexed low rate RL dedicated channel does not occupy any RL resource or cause any interference if the MS has no low rate data traffic or control signalling to transmit on this channel. On the other hand, in 802.16e which is based on OFDMA, a RL dedicated channel is defined based on OFDMA time-frequency resource. Once allocated to a MS, the OFDMA time-frequency resource cannot be used by other MSs even though the MS may not have data or control signalling to transmit for a certain period of time.

Overall, contention free schemes are more complex to implement due to the need for centralized control at the BS or across multiple BSs. However, contention free schemes enable tight QoS control and resource management, thus provide optimum trade-off between QoS and system capacity. Contention free MAC is therefore widely used in 3G cellular systems where spectrum is scarce and QoS support is crucial for multimedia services. It is also envisioned that contention free MAC will be the dominant mechanism used in B3G systems.

2.3 Summary

In this chapter, we presented an overview of MAC protocols and their roles in existing wireless systems. In subsequent chapters of this thesis, we explore the various MAC enhancements and solutions for B3G wireless systems.

Chapter 3

Wireless Cellular System Modeling

In this chapter, we provide an overview of the methodology used to evaluate the system performance of the different MAC solutions proposed in this thesis. In 3G cellular systems, such as 1xRTT [9][10][11], 1xEV-DV [12][13], 1xEV-DO [1], UMTS [14] and HSDPA [2], interaction between the physical layer and MAC layer is intricate and dynamic in order to effectively support a wide variety of voice and packet data services and QoS. For example, the scheduler at the MAC layer performs fast scheduling of packets received from upper-layers based on real-time channel condition at the physical layer and QoS requirements of the traffic. To adequately evaluate the performance of 3G cellular systems, a detailed evaluation methodology [27] has been developed in 3GPP2 standards forum and widely adopted in the cellular industry. Similar methodology based on [27] has also been used in other standards forum, including 3GPP [28] and IEEE [29].

In this thesis, we evaluate the performance of our proposed MAC solutions for B3G systems based on the methodology described in [27]. In subsequent sections of this chapter, we provide an overview and highlight of this methodology as detailed account of the methodology can be found in [27]. Some enhancements we make to the methodology defined in [27] in order to adequately study the MAC performance are also described.

3.1 Overview of System Evaluation Methodology

A system level simulation framework is defined in [27] to model various aspects of physical layer and MAC layer. The system level simulation framework models a cellular

layout with 19 tri-sector cells, each with uniformly distributed mobiles in the cell area. Both long-term propagation loss and fast fading are modeled on each link between a BS and a MS in the 19-cell layout. The system level simulation parameters are summarized in Table 1.

Table 1 System Level Simulation Parameters

Parameters	Value
Number of cells	19 tri-sector cells, 57 sectors (see Figure 2)
Antenna horizontal pattern	70 deg (-3 dB) with 20dB front-to-back ratio (see equation (1)). No loss is assumed on the vertical azimuth.
Antenna orientation	As shown in Figure 2
Propagation model	Modified Hata urban propagation model @ 1.9GHz (COST 231). Minimum of 35 m separation between MS and BS. BTS antenna height = 32 m. MS antenna height = 1.5 m. Equivalent to $28.6 + 35\log_{10}(d)$ dB, where d is the distance between the BS and the MS in meters.
Log-normal shadowing	Standard deviation = 8.9dB. Shadowing correlation between sectors in a BS = 1. Shadowing correlation between BS = 0.5
Mobile location	MSs are dropped uniformly in a sector. Once dropped, a MS' location is fixed during a simulation run. Note that in this thesis, we enhance this by changing the MS location every 30 seconds to evaluate the MAC state performance in Chapter 7 and Chapter 8.

Mobile noise figure	10.0 dB
Thermal noise density	-174 dBm/Hz
Carrier frequency	2 GHz
BS antenna gain with cable loss	15 dB
MS antenna gain	-1 dBi
Other losses	10 dB
Fast fading model	See Table 2.
BS maximum PA power	20 Watts
Site-to-site distance	2.5 km

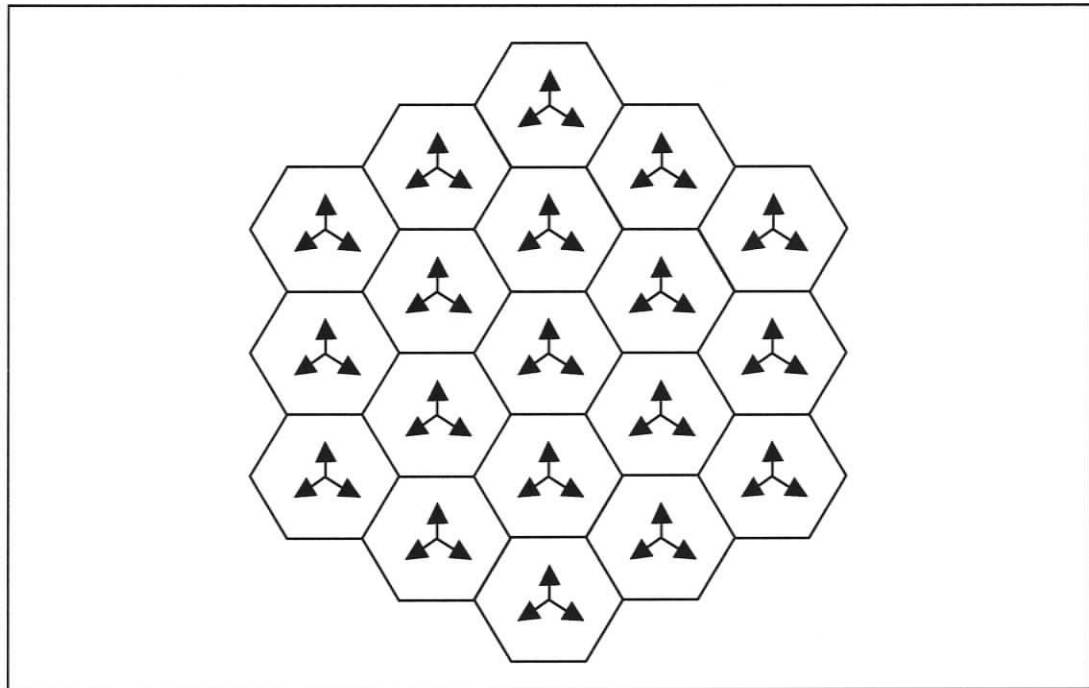


Figure 2 Cell Layout in System Level Simulation

Table 2 Fast Fading Channel Models

Channel Model	ITU Multi-path Model	# of Resolvable Paths	Speed (km/h)	Fading Model	Assignment Probability
Model A	Pedestrian A	1	3	Jakes	0.30
Model B	Pedestrian B	3	10	Jakes	0.30
Model C	Vehicular A	2	30	Jakes	0.20
Model D	Pedestrian A	1	120	Jakes	0.10
Model E	Single path	1	0, $f_D=1.5$ Hz	Rician Factor K = 10 dB	0.10

The antenna pattern is specified as

$$(1) \quad A(\theta) = -\min \left[12 \left(\frac{\theta}{\theta_{3dB}} \right)^2, A_m \right],$$

where $-180 \leq \theta \leq 180$, θ_{3dB} is the 3 dB beamwidth, and $A_m = 20$ dB is the maximum attenuation.

3.1.1 Physical Layer Modeling

Physical layer operation is modeled for each transmission/reception in the system, i.e. for each of the MS in the 19 tri-sector cells.

The physical layer transceiver operation, i.e. modulation and demodulation, encoding and decoding operations are not explicitly modeled in the system level simulation due to the excessive simulation run time. Instead, the transceiver performance of a link is represented by short term frame error rate (FER) versus signal-to-noise-ratio (SNR) curves which are pre-generated by independent link level simulations. At each physical

layer frame instance, the system level simulation calculates the effective short term SNR based on the received signal, interference and noise, and performs table-lookup on the appropriate link level curve to obtain the corresponding FER value. A frame erasure is determined by a random event with probability of the FER value. Typically, one short-term link level curve is required for each unique combination of encoder packet size, modulation, coding, and channel model.

Other physical layer operations including Hybrid Automatic Repeat Request (HARQ), adaptive modulation and coding, power control, channel quality feedback, HARQ acknowledgment are modeled explicitly for each transmission/reception in the system level simulation. HARQ soft combining gain is represented by an increase in cumulative SNR and a decrease in coding rate which are used for table-lookup on the corresponding link level curve. Feedback latency is also modeled explicitly.

3.1.2 Layer 2/3 Modeling

Layer 2 scheduling is modeled explicitly to schedule the transmission of layer 2 packets and HARQ retransmission packets of each user in the system. Radio resource management in terms of power, time slot, carrier, etc., are modeled explicitly. The detailed of such modeling is not given in [27] since it is dependent on the specific algorithms evaluated. In our simulation framework, we perform detailed modeling of the scheduling and radio resource management operations of our proposed schemes.

The modeling of MAC state machine is not described in [27]. To study the performance of the proposed MAC state solution, we enhance the simulation framework to include the MAC state machine modeling for each user and the admission control of users to different MAC states. We also enhance the simulation framework to model the

layer 2 ARQ protocol, which is called Radio Link Protocol (RLP) in 1xRTT, 1xEV-DV and 1xEV-DO and Radio Link Control (RLC) in UMTS and HSDPA.

Layer 3 signalling or call flow is not explicitly modeled in the simulation framework, but its associated overhead on the air interface is accounted for by static resource utilization. For example, in 1xEV-DV, 20% of BS transmit power is consumed by the Pilot channel, the Paging channel and the Sync channel. In 1xEV-DO, 22% of time domain resource is occupied by the Pilot channel and the MAC channel.

3.1.3 Upper Layer Protocols Modeling

Transport Control Protocol (TCP), Internet Protocol (IP) and Point-to-Point Protocol (PPP) overhead are modeled. A simplified modeling of TCP operation is given in [27], which models the slow-start operation of TCP, but not the congestion control operation. To adequately evaluate the performance of TCP-based applications such as FTP, HTTP, we enhance our simulation framework to model full TCP operations.

3.2 Traffic Modeling

In this thesis, we evaluate the performance of our proposed MAC solutions based on the following traffic models defined in [27]: full buffer, FTP, HTTP, WAP and near-real-time-video (NRTV). A brief description of these traffic models are given in subsequent sections as reference. In addition, we also include the modeling of Voice-over-IP (VoIP) traffic which is one of the applications in 3G systems that has received increasing market demand.

3.2.1 Full Buffer

In full buffer traffic model, it is assumed that a user's buffer is always full. This is not a realistic traffic model but is used to represent the upper bound system performance since the system is full loaded and the highest priority or most optimum user selected by the scheduler always has data to transmit.

3.2.2 FTP

As defined in [27], an FTP session consists of a sequence of file transfer, separated by reading time, D_{pc} , as shown in Figure 3. As mentioned earlier, a full TCP is modeled as the underlying transport protocol. Each TCP segment is shown as a packet in Figure 3.

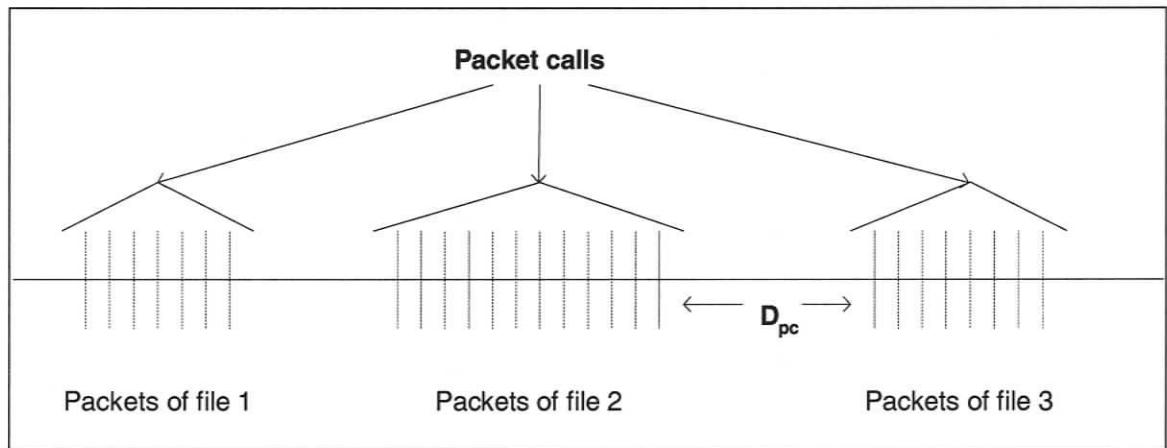


Figure 3 Packet Trace in a Typical FTP Session

The FTP traffic model parameters as defined in [27] are summarized in Table 3.

Table 3 FTP Traffic Model Parameters

Component	Distribution	Parameters	PDF
File size (S)	Truncated Lognormal	Mean = 2Mbytes Std. Dev. = 0.722 Mbytes Maximum = 5 Mbytes	$f_x = \frac{1}{\sqrt{2\pi\sigma x}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], x \geq 0$ $\sigma = 0.35, \mu = 14.45$
Reading time (D _{pc})	Exponential	Mean = 180 sec.	$f_x = \lambda e^{-\lambda x}, x \geq 0$ $\lambda = 0.006$

3.2.3 HTTP

As defined in [27], a web-browsing session consists of a number of web-page downloads and intermediate reading times, as shown in Figure 4. The web-page downloads are referred to as packet calls. As shown in Figure 5, a packet call or web-page download consists of a main object in a page and several embedded objects. As mentioned earlier, a full TCP is modeled as the underlying transport protocol. Each object consists of multiple TCP segments or packets.

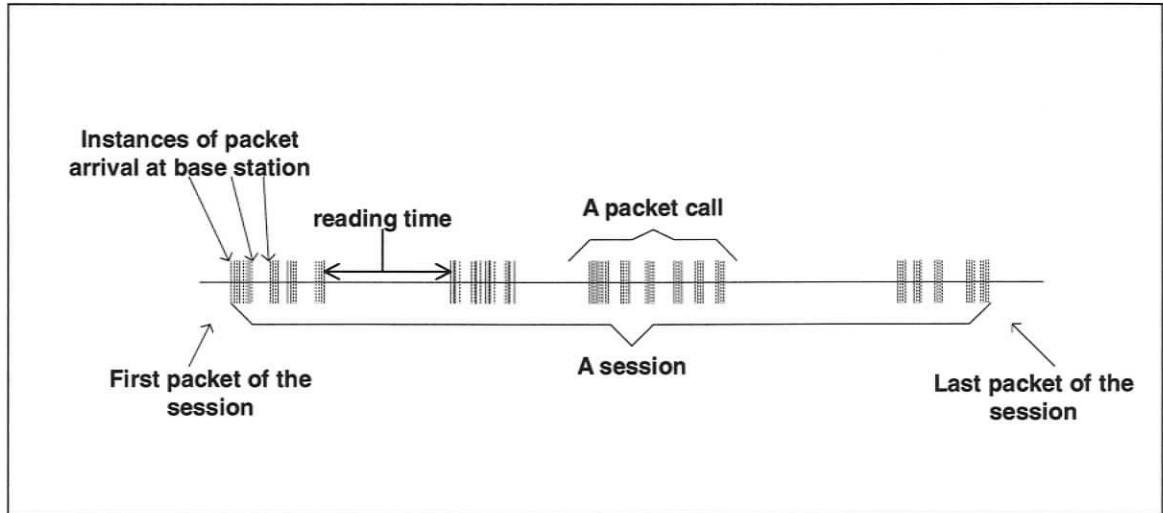


Figure 4 Packet Trace of a Typical Web Browsing Session

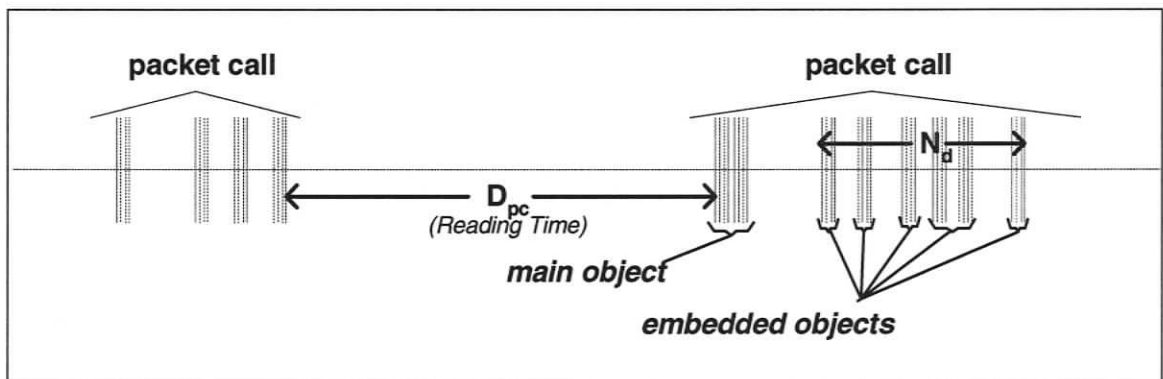


Figure 5 Contents of a Packet Call

The HTTP traffic model parameters as defined in [27] are summarized in Table 4.

Table 4 HTTP Traffic Model Parameters

Component	Distribution	Parameters	PDF
Main object size (S_M)	Truncated Lognormal	Mean = 10710 bytes Std. dev. = 25032 bytes Minimum = 100 bytes Maximum = 2 Mbytes	$f_x = \frac{1}{\sqrt{2\pi\sigma x}} \exp\left[\frac{-(\ln x - \mu)^2}{2\sigma^2}\right], x \geq 0$ $\sigma = 1.37, \mu = 8.35$
Embedded object size (S_E)	Truncated Lognormal	Mean = 7758 bytes Std. dev. = 126168 bytes Minimum = 50 bytes Maximum = 2 Mbytes	$f_x = \frac{1}{\sqrt{2\pi\sigma x}} \exp\left[\frac{-(\ln x - \mu)^2}{2\sigma^2}\right], x \geq 0$ $\sigma = 2.36, \mu = 6.17$
Number of embedded objects per page (N_d)	Truncated Pareto	Mean = 5.64 Max. = 53	$f_x = \frac{\alpha k^\alpha}{\alpha+1}, k \leq x < m$ $f_x = \left(\frac{k}{m}\right)^\alpha, x = m$ $\alpha = 1.1, k = 2, m = 55$ Note: Subtract k from the generated random value to obtain N_d
Reading time (D_{pc})	Exponential	Mean = 30 sec	$f_x = \lambda e^{-\lambda x}, x \geq 0$ $\lambda = 0.033$
Parsing time for the main object (T_p)	Exponential	Mean = 0.13 sec	$f_x = \lambda e^{-\lambda x}, x \geq 0$ $\lambda = 7.69$

3.2.4 WAP

As defined in [27], a WAP session consists of series WAP request, WAP response and reading time, as shown in Figure 6. A WAP response consists of a number of objects or packets. The WAP traffic model parameters as defined in [27] are summarized in Table 5

Table 5 WAP Traffic Model Parameters

Component	Distribution	Parameters
Size of WAP request	Deterministic	76 octets
Object size	Truncated Pareto	Mean= 256 bytes, Max= 1400 bytes K = 71.7 bytes, $\alpha = 1.1$
# of objects per response	Geometric	$P(k) = 0.5^k$, $k \geq 1$ Mean = 2
Inter-arrival time between objects	Exponential	Mean = 1.6 s
WAP gateway response time	Exponential	Mean = 2.5 s
Reading time	Exponential	Mean = 5.5 s

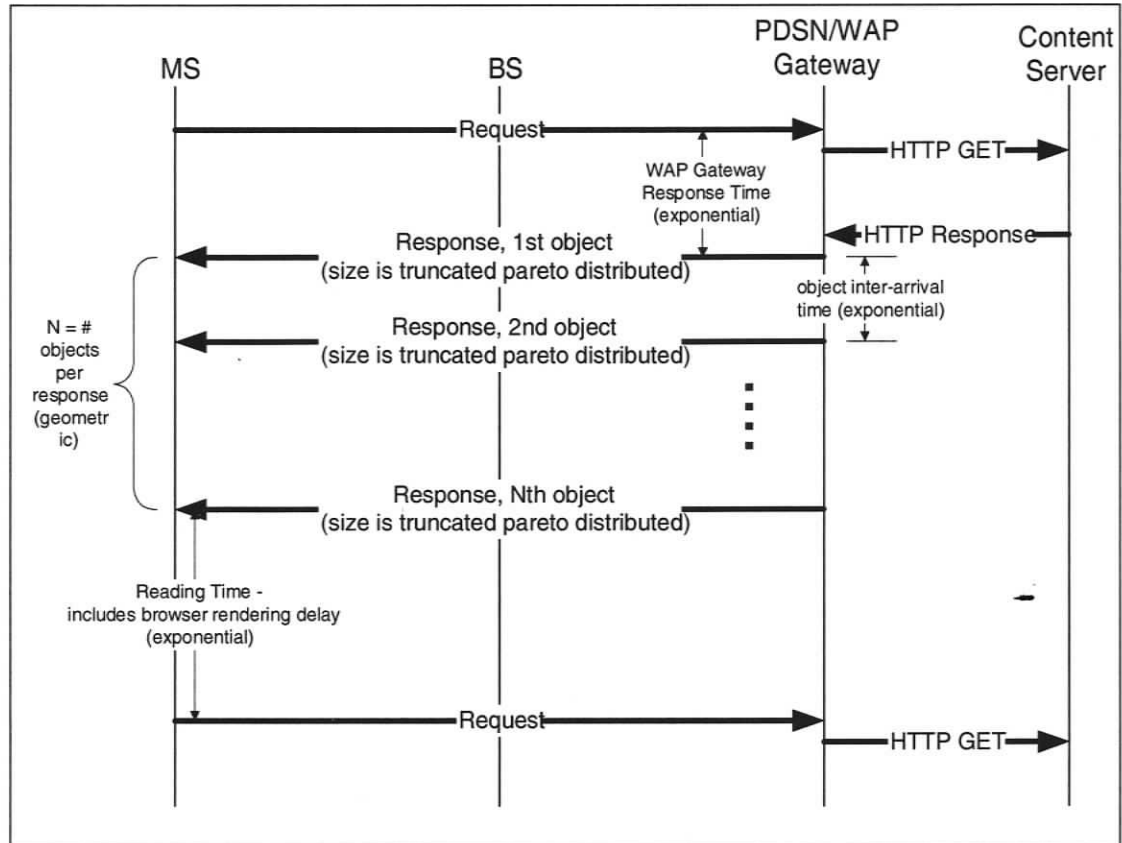


Figure 6 Packet Trace for WAP Traffic Model

3.2.5 NRTV

As defined in [27], an NRTV streaming session consists of regular arrival of video frames spaced by interval T . Each frame is decomposed into a fixed number of slices or packets. This is shown in Figure 7.

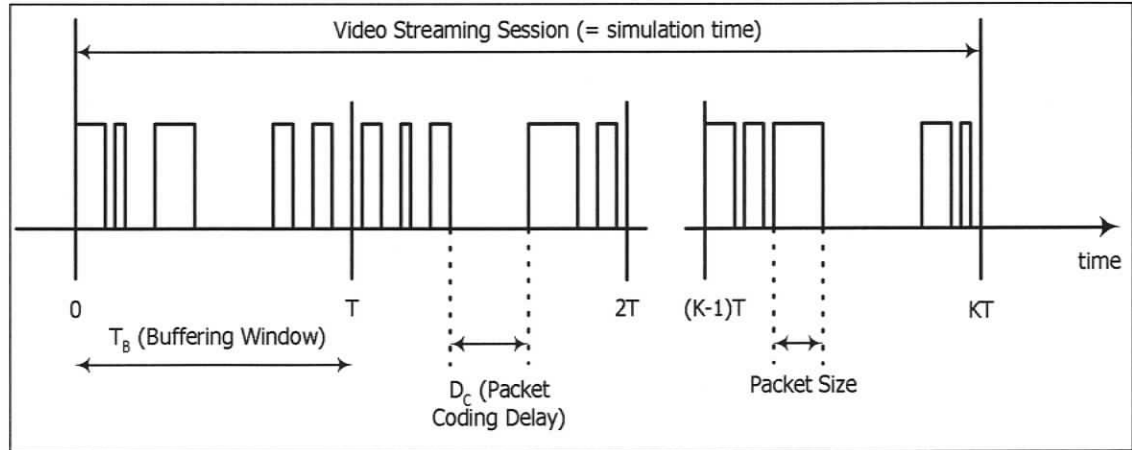


Figure 7 Packet Trace for NRTV Traffic Model

The NRTV traffic model parameters as defined in [27] are summarized in Table 6.

Table 6 NRTV Traffic Model Parameters

Component	Distribution	Parameters
Inter-arrival time between the beginning of each frame	Deterministic (based on 10 frames per second)	100 ms
Number of packets/slices in a frame	Deterministic	8
Packet/slice size	Truncated Pareto	Mean= 50 bytes, Max= 125 bytes K = 20 bytes, $\alpha = 1.2$
Inter-arrival time between packets/slices in a frame	Truncated Pareto	Mean= 6 ms, Max= 12.5 ms K = 2.5 ms, $\alpha = 1.2$

3.2.6 VoIP

The source model for VoIP traffic is based on the standard Markov model defined in [30]. Voice frame is generated by the codec every 20ms. The voice activity factor is 0.403 with 29% full rate, 60% eighth rate, 4% half rate and 7% quarter rate. The packet size is based on the Enhanced Variable Rate Codec (EVRC) [31]. The voice frame payload sizes and the overall packet sizes including the compressed RTP/UDP/IP/PPP header of 8 bytes, i.e. 3 bytes of compressed RTP/UDP/IP header and 5 bytes of PPP header, are shown in Table 7 below.

Table 7 VoIP Packet Sizes

Frame Rate	Voice Frame Payload Size (bits)	VoIP Packet Size (Byte Aligned) (bits)
Full rate	172	240
Half rate	80	144
Quarter rate	40	104
Eighth rate	0 (Eighth rate frame is assumed blanked)	0

3.3 Performance Metric

The performance metrics used to quantify the overall system performance are per-user outage, system outage and fairness. We first describe these performance metrics for FTP,

HTTP, WAP and NRTV as defined in [27], followed by the performance metrics for VoIP used in this thesis.

3.3.1 Performance Metrics for FTP, HTTP, WAP and NRTV

Per-user outage criterion is used to determine which users in the system are in outage. A system outage criterion is used to define the overall system capacity based on the maximum allowable percentage of users in the system that are in outage. In [27], the per-user outage criterion defined for FTP and HTTP is a FTP or HTTP user shall experience a packet call throughput of at least 9600 bps, where packet call throughput for user n , in simulation run m is defined as

$$(2) \quad PCTP(m,n) = \frac{\sum_{k=1}^{K(m,n)} \sum_{l=1}^{L(m,n,k)} B(m,n,k,l)}{\sum_{k=1}^{K(m,n)} (PCTD(m,n,k) - PCTA(m,n,k))},$$

where $K(m,n)$ is the total number of packet calls generated for user(m,n), k is the index of packet calls for a user, $L(m,n,k)$ is the total number of packets generated for the k th packet call of user(m,n), $B(m,n,k,l)$ is the number of information bits contained in the l th packet of the k th packet calls for user(m,n). If the packet is not successfully delivered by the end of the simulation run, $B(m,n,k,l) = 0$. $PCTD(m,n,k)$ is the delivered time of the k th packet call for user(m,n). It is the time when the receiver successfully receives the last packet of the packet call. $PCTA(m,n,k)$ is the arrival time of the k th packet call for user(m,n). It is the time when the first packet of the packet call arrives at the transmitter side and is put into a queue.

The per-user outage criterion defined for WAP is a WAP user shall experience a packet throughput of at least 4800 bps, where packet throughput for user n , in simulation run m is defined as

$$(3) \quad PTP(m, n) = \frac{\sum_{k=1}^{K(m, n)} \sum_{l=1}^{L(m, n, k)} B(m, n, k, l)}{\sum_{k=1}^{K(m, n)} \sum_{l=1}^{L(m, n, k)} (TD(m, n, k, l) - TA(m, n, k, l))}.$$

where $TD(m, n, k, l)$ is the delivered time of the l th packet of the k th packet call for user (m, n) . It is the time when the receiver successfully receives the packet. $TA(m, n, k, l)$ is the arrival time of the l th packet of the k th packet call for user (m, n) . It is the time when the packet arrives at the transmitter side and is put into a queue.

The per-user outage criterion defined for NRTV is the fraction of video frames that are discarded, due to queuing delay at the scheduler exceeds 5 seconds, shall be less than 2% for each user.

The overall system outage criterion for defining system capacity for FTP, HTTP, WAP and NRTV is no more than 2% of users in the system are in per-user outage.

In addition to the per-user outage criteria and system outage criterion, a fairness criterion is also defined in [27] to ensure a fair distribution of system resource to all users in the system. The fairness is defined based on the CDF of the normalized user throughputs with respect to the average user throughput for all users. This CDF shall lie to the right of the curve given by the three points in Table 8.

Table 8 Fairness Criterion

Normalized Throughput with respect to Average User Throughput	CDF value
0.1	0.1
0.2	0.2
0.5	0.5

3.3.2 Performance Metrics for VoIP

We define performance metrics for VoIP similar to that of the NRTV since VoIP is also delay sensitive with a much tighter delay bound compared to NRTV. We define the per-user outage criterion for VoIP as the fraction of voice frames that are dropped, due to queuing delay at the scheduler exceeding 80 ms and over-the-air frame erasure, shall be less than 3% for each user. 80 ms is chosen as the delay bound since the recommended mouth-to-ear delay for acceptable voice quality is no more than 200 ms [32]. 80 ms one way air-link delay will allow sufficient margin for delay at the access network and the core network for the case of mobile-to-mobile call. A 3% per-user outage criterion is chosen for the sum of frame erasure rate typically kept at 1% and voice frame delay outage of 2%.

Further, we define the system outage as no more than 5% of VoIP users in the system shall be in per-user outage.

3.4 Summary

In this chapter, we provided an overview of the modeling and evaluation methodology widely used in the cellular industry to study the performance of 3G systems. We also provided enhancements to the existing evaluation methodology in order to study the MAC performance. Unless otherwise specified, the simulation framework described in this chapter is used in subsequent chapters in this thesis to evaluate the performance of our proposed MAC solutions for the B3G systems.

Chapter 4

Beyond 3rd Generation Multi-mode Wireless Access System

In this chapter, we introduce a new B3G access system concept based on heterogeneous PL modes or configurations, with a common (or anchor) layer 2 and layer 3 protocol stack. Such a system can support a wide variety of physical layer access technologies, e.g. CDMA, OFDMA, MC-CDMA etc, with different configurations to target different deployment scenarios and performance targets, while anchored by common layer 2 and layer 3 protocols. Seamless handoff is enabled by the common layer 2 and layer 3 protocols, while dynamic radio resource/load/spectrum management can be performed across the different PL modes to achieve optimum spectrum efficiency and QoS support. Such a B3G access system concept is complimentary to the IP layer convergence approach which is required when the B3G access system inter-works with existing 2G/3G heterogeneous access systems.

In Section 4.1, we provide a detailed description of the proposed B3G access system concept. Section 4.2 presents the common layer 2/3 protocol operation anchoring the different PL modes. An example of a specific B3G system configuration is presented in Section 4.3, followed by a summary in Section 4.4.

4.1 Multi-mode Access System Concept for B3G

The proposed B3G multi-mode access system concept consists of multiple heterogeneous PL modes where each PL mode corresponds to a specific multiple access

technique, such as CDMA, OFDMA, MC-CDMA etc., with specific configurations. The heterogeneous PL modes are anchored by a common layer 2/3 air-interface protocol stack. Each PL mode targets a specific deployment scenario in terms of mobility conditions, service types, and backward compatibility requirements to existing 3G systems. For example, pseudo-orthogonal multiple access techniques such as CDMA and MC-CDMA are more efficient for low to medium rate, delay sensitive services; whereas orthogonal multiple access technique such as OFDMA is more efficient for broadband high data rate services. PL modes of the same multiple access technique can also be configured differently. For example, in OFDMA, different cyclic prefix lengths are used to support local area unicast, wide area unicast, broadcast and multicast services. Also, fixed and mobile services require different OFDM sub-carriers spacing. Backward compatibility is also an important consideration if the B3G system needs to overlay onto an existing 3G system in the same spectrum. More detailed account of such 3G/B3G overlay scenario is addressed in Section 4.3.

The protocol structure of the proposed multi-mode access system is illustrated in Figure 8. We call this a service-driven protocol design, where the overall goal is to maximize the overall system capacity, while ensuring users' service requirements are met. The common layer 2/3 protocols provide a common interface to wireline upper layer protocols such as IP, PPP, TCP and the application layer. The common layer 2/3 protocols provide dynamic or semi-static radio resource management, load control, mobility management, and spectrum management across the different PL modes. A user is dynamically or semi-statically assigned radio resource from a particular PL mode based on service requirements, i.e., QoS and SLA, mobility conditions, loading of each

PL mode, and the user's location with respect to the coverage area of each PL mode. As shown in Figure 8, the common layer 2/3 can be split into a centralized portion (e.g. at the Base Station Controller (BSC) or the Access Gateway (AG)) and a distributed portion (e.g. at the Base Transceiver Subsystem (BTS) or the BS).

Note that the common layer 2/3 protocols can be an extension of the layer 2/3 protocols in existing 3G cellular systems to support a smooth system evolution from 3G to B3G.

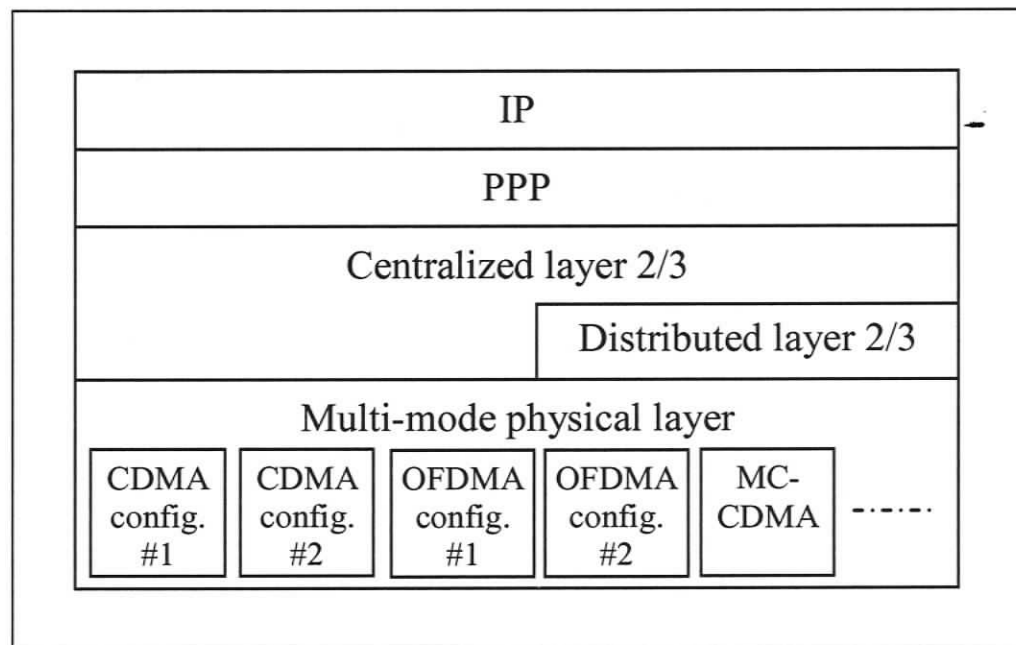


Figure 8 Anchor Layer 2/3 Protocol Design for Multimode Physical Layer

In the following subsections, we describe the different aspects of the multi-mode access system concept, which are, 1) the spectrum allocation scenarios of the multimode physical layer; and 2) the hierarchical cellular layout.

4.1.1 Spectrum Allocation Scenarios of the Multimode Physical Layer

When different PL modes overlap in their coverage areas, spectrum sharing can be done via TDM, CDM or FDM. The TDM approach is used when different PL modes

share the same spectrum. This typically applies to the case where the same spectrum is used to support both legacy terminals and new advanced terminals during the transitional period when the deployment migrates from one PL mode to another. This spectrum sharing scenario is also applicable when the same spectrum is used to support different types of services which require different PL modes, e.g., broadcast/multicast service and unicast service which may have different requirement on physical layer performance (such as coverage, or cyclic prefix length in the case of OFDMA); or fixed service and mobile service which are optimized under different physical layer configurations.

The CDM approach applies only to CDMA based PL modes, and is used when the same spectrum is shared by different CDMA-based PL modes. One example is the overlay of 1xRTT [9][10][11], 1xEV-DV [12][13] and MC-DV, a proposed multi-carrier extension of 1xEV-DV that is described in details in Chapter 5.

The FDM approach is used when different PL modes are assigned non-overlapping frequency spectrum.

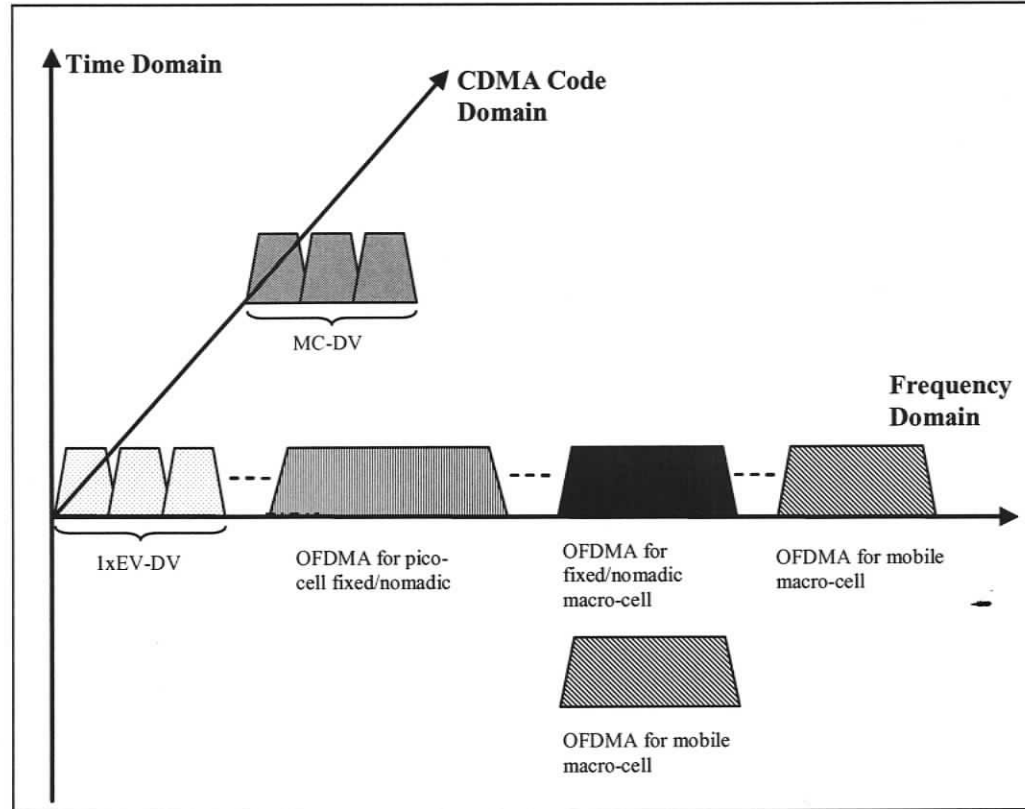


Figure 9 An Illustration of Spectrum Allocation Scenarios of the Multi-mode Physical Layer

Figure 9 provides an example of the spectrum allocation scenarios for illustration purpose. 1xEV-DV and MC-DV are overlaid on the same spectrum using CDM. 1xEV-DV, MC-DV and OFDMA with different configurations are FDM. OFDMA mode with different configurations (e.g. OFDMA for fixed/nomadic macro-cellular and OFDMA for mobile macro-cellular) can also be overlaid on the same spectrum using TDM. This means that certain time slots are used for OFDMA fixed/nomadic transmission/reception, while other time slots are used for OFDMA mobile transmission/reception.

When PL modes are TDM or FDM, the time slot or spectrum partition between PL modes can be static, semi-static or dynamic. For semi-static or dynamic partitioning, the particular time slot(s) or spectrum assigned to each PL mode needs to be signalled to the MS. Either explicit signalling through a common signalling channel or implicit signalling

through blind detection at the MS can be used. There is therefore inherent cost associated with semi-static or dynamic partitioning of TDM or FDM resource across PL modes. However, as we will show in Chapter 6, semi-static and dynamic partitioning of resource across PL modes provide a significant performance gain over static resource allocation. In Section 4.3, we illustrate a specific multi-mode air-interface configuration with three PL modes in TDM and FDM fashion, i.e. 1xEV-DO [1], OFDMA which is TDM with 1xEV-DO and OFDMA which is FDM with 1xEV-DO; and the common signalling channel used to indicate the PL mode of a particular time slot.

4.1.2 Hierarchical Cell Layout

The multi-mode access system can naturally support the hierarchical cell layout envisioned for B3G networks. As the different PL modes used for different cell hierarchies are anchored by the common layer 2/3 protocol, handoff and radio resource/load management between cell hierarchies can be effectively performed.

Figure 10 shows the access network configuration for a hierarchical cell layout. The access network consists of the AG, the BS and the Relay Station (RS). The air-interface physical layer and the distributed layer 2/3 function are contained in the BS and the RS. The RS may contain a full distributed layer 2/3 function or a subset of the distributed layer 2/3 function dependent on the complexity of the RS. The simplest form of a RS is a store and forward repeater. A BS may support multiple PL modes in the same coverage area provided by the BS. A RS may support the same PL mode as the BS it associates to or a different PL mode. The air interface centralized layer 2/3 function is contained in the AG. The AG also serves as a router within the access network and provides an interface to the core network and the public IP network.

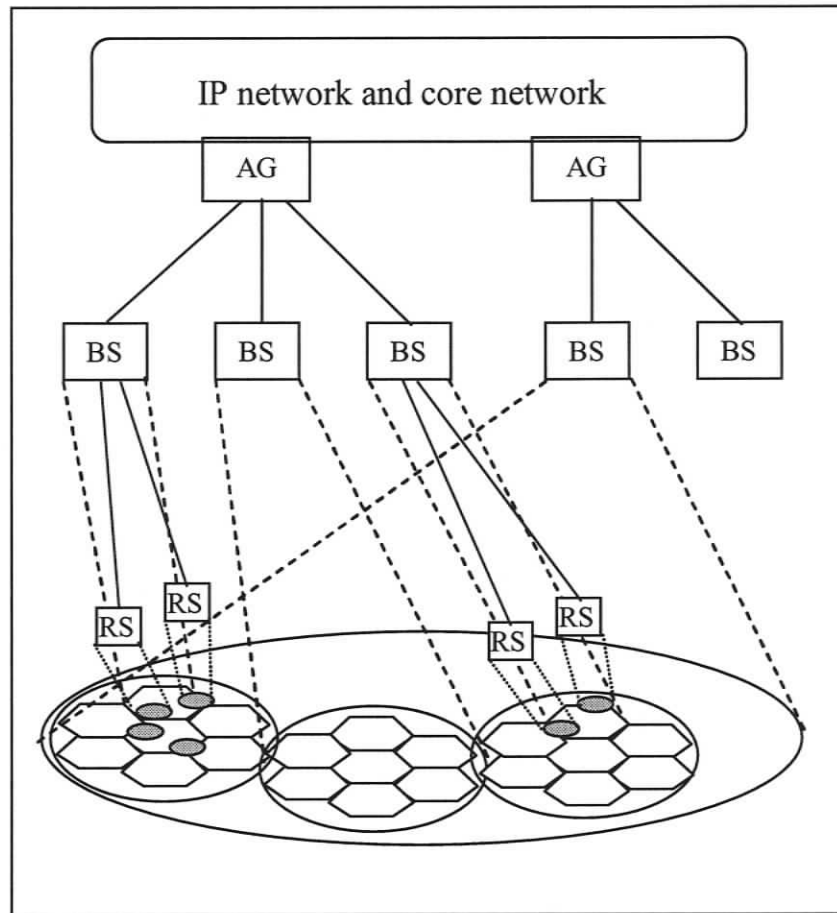


Figure 10 Hierarchical Cell Layout

As illustrated in Figure 10, BSs belonging to different cell hierarchies can be anchored by the same AG or by different AGs. For handoff between BSs (and the corresponding cell hierarchies) anchored by the same AG, the centralized layer 2/3 context of the user is contained within the AG, and thus there is no need for any context transfer. The distributed layer 2/3 context of the user will be transferred from the previous serving BS to the target BS through the AG or through the direct BS-BS link if such a link exists. Since a common layer 2/3 protocol stack is used by all BS PL modes, there will be no loss of layer 2/3 context information during handoff, thus avoiding unnecessary packet loss and latency. For handoff between BSs anchored by different AGs, the layer 2/3

context information will be transferred from the previous serving AG/BS to the target AG/BS. Similarly, handoff between RSs associated to the same BS requires a subset of the layer 2/3 context information to be transferred from the serving RS to the target RS via the BS to RS links. For handoff between RSs associated to different BSs, the required context information for transfer is the same as that of inter-BS handoff.

4.2 Common Layer 2/3 Protocol Operation

As described earlier, the common layer 2/3 protocol stack anchoring the multi-mode physical layer provides mobility management, radio resource management, load control and spectrum management across the different PL modes. A breakdown of the layer 2/3 protocol functional blocks is given in Figure 11. On the data plane, the multiplex/demultiplex sublayer provides dynamic mapping of layer 2/3 data and signaling to/from the resource pools of the various PL modes. On the control plane, the mobility management function provides handoff support, and the radio resource and load management function manages how the multi-mode physical layer resource pools are used as well as the load control between the resource pools. The MAC state control function contains the MAC state machine of each user and controls how a user is transitioned from one state to another.

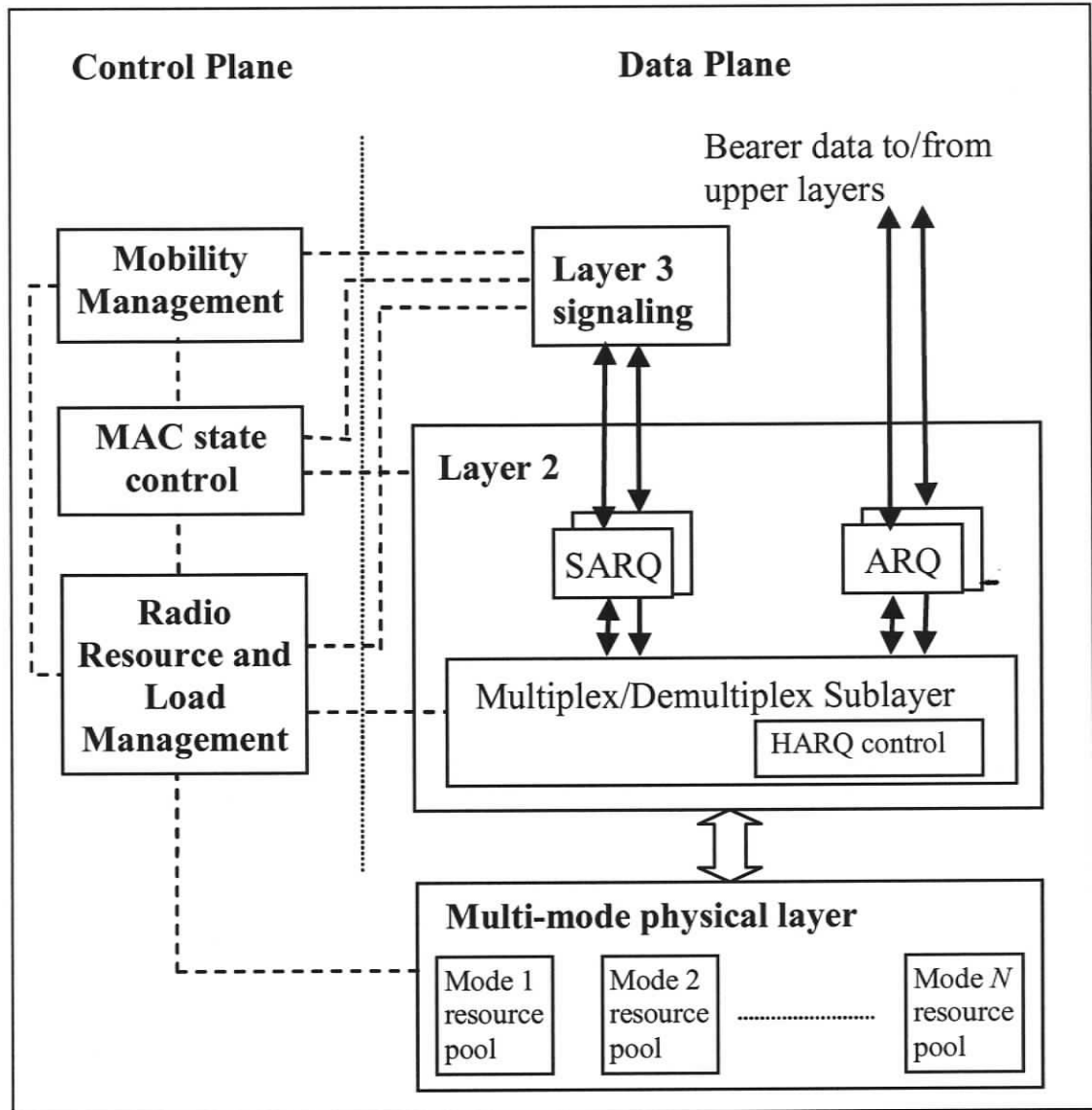


Figure 11 Common Layer 2/3 Protocol Functional Blocks

The following subsections provide a detailed description of the various functional blocks of the common layer 2/3 protocol.

4.2.1 Mobility Management

The mobility management function supports inter-BS handoff, intra-BS handoff between PL modes, inter-AG handoff, and inter-RS handoff. A handoff can either be

initiated by the MS or by the access network. The final handoff decision is made by the access network based on the radio resource and load management policy. Inter-BS and inter-AG handoff occur when the MS moves from the coverage area of one BS to the coverage area of another BS. Intra-BS handoff between PL modes occurs when the mobility conditions, service requirements of the MS change or for the purpose of load balancing between PL modes. Inter-RS handoff occurs when the MS moves from the coverage area of one RS to the coverage area of another RS.

The mobility management function is also responsible for layer 2/3 context transfer during handoff. The required layer 2/3 context information transfer for inter-BS, inter-AG and inter-RS handoffs is described in Section 4.1.2. For intra-BS handoff between PL modes, no layer 2/3 context transfer is necessary. There are two scenarios for intra-BS handoff. The first scenario is the handoff between PL modes occupying different frequency spectrums. An explicit handoff mechanism is required if the terminal is not capable of receiving radio frequency signal from multiple PL modes. An explicit handoff mechanism is not required if the terminal has a wideband receiver or a multi-carrier receiver that can receive signal from multiple PL modes across different frequency bands, and the terminal is able to dynamically switch between PL modes on a frame-by-frame basis based on signalling information from the BS. The second scenario is the handoff between PL modes sharing the same frequency spectrum using either CDM or TDM, as described in Section 4.1.1. In this case, an explicit handoff mechanism is not required if the terminal can dynamically switch between the PL modes on a frame-by-frame basis based on the signalling information from the BS. If the terminal cannot dynamically

switch between different PL modes, an explicit handoff mechanism is required to ‘handoff’ the user from one PL mode to another.

4.2.2 Radio Resource and Load Management

This function is responsible for the following categories of radio resource management:

1) dynamic resource allocation within a PL mode, 2) semi-static and dynamic resource partitioning across PL modes sharing the same spectrum, 3) semi-static and dynamic spectrum allocation and load balancing across PL modes occupying different spectrums and different cell hierarchies.

For the first category, resources within a PL mode (e.g. code domain for the case of CDMA, time domain for the case of CDMA or OFDMA, frequency domain for the case of OFDMA) are dynamically shared among the users based on factors such as the QoS requirements of user traffic, SLA of the users, and the channel conditions experienced by the users.

For the second category, the same spectrum is shared by multiple PL modes using either CDM or TDM. The resource partition, i.e. code space partition for CDM, or time slots partition for TDM, between PL modes can be configured semi-statically based on the loading condition on each PL mode. The resource partition can also be dynamic on a frame-by-frame basis based on the real-time QoS requirements and buffer occupancy of user traffic in each PL mode. The frame-by-frame multiplexing and demultiplexing of user traffic to different PL modes is performed at the multiplex/demultiplex sublayer based on the scheduling policy defined by the radio resource management function.

For the third category, the radio resource and load management function provides load balancing across different PL modes occupying different frequency spectrums. Either the

spectrum allocation among PL modes are adaptive to the loading condition of each PL mode or users are assigned/handoff to different PL modes which have fixed spectrum allocation based on loading condition of each PL mode. The load balancing mechanism aims to maximize the overall spectral efficiency while meeting the QoS requirements of user traffic and mobility conditions (i.e. whether a user is fixed, nomadic or mobile) of users. Detailed accounts of the load balancing schemes between PL modes are given in Chapter 6.

In addition to load balancing, the radio resource management function also makes handoff decisions based on user location and channel conditions with respect to the cell hierarchies of the different PL modes. The radio resource and load management function provides directives to the mobility management function to perform the actual handoff operation.

4.2.3 MAC State Control

For power saving purposes, a MS can be in different operational states, such as the Active state, the Control-hold state and the Dormant state in 1xRTT or 1xEV-DV systems; the Idle state and the Connected state in a 1xEV-DO system; The Idle mode, the Sleep mode and the Normal mode in an 802.16e system [15]. An overview of different MAC states and their corresponding functions are given in Chapter 7. Since a common MAC state machine is used for different PL modes, when a user handoffs or switches from one PL mode to the other, the MAC state machine context can be retained. We introduce a Universal MAC States concept to anchor a user switching/handoff from one PL mode/BS to another. This is described in details in Chapter 7.

4.2.4 System Access

The B3G multi-mode access system supports terminals with different capabilities. For example, a terminal may support one or multiple PL modes. To allow universal system access, a set of primary common channels, i.e. paging, access and synchronization channels are defined for each PL mode to allow terminals that support each PL mode to perform initial access to the system. For each PL mode, designated anchor carrier(s) within the spectrum is defined to carry the primary common channels. In addition to the primary common channels used for initial system access, each PL mode also defines an additional set of common channels called the supplemental common channels used by terminals that have already acquired and operate in the PL mode. This additional set of common channels is used for MAC state transition from the Dormant state or other intermediate states to the Active state. Unlike the primary common channels, the supplemental common channels may not reside in the anchor carrier. The purpose of dividing the common channels into primary set and supplemental set is to allow dynamic resource provisioning of the supplemental set based on loading of each PL mode.

4.3 An Example of a Specific Multi-mode Access System

Configuration

In this section, we provide a specific configuration of the multi-mode access system to illustrate how the proposed multi-mode access system can facilitate evolution of existing 3G cellular systems to the proposed B3G system. This is illustrated in Figure 12, where the multi-mode access system consists of existing 3G 1xEV-DO PL mode, OFDMA overlay PL mode and OFDMA standalone PL mode. The 1xEV-DO mode and OFDMA

overlay mode share the same frequency spectrum using TDM. The OFDMA standalone mode occupies a separate frequency spectrum.

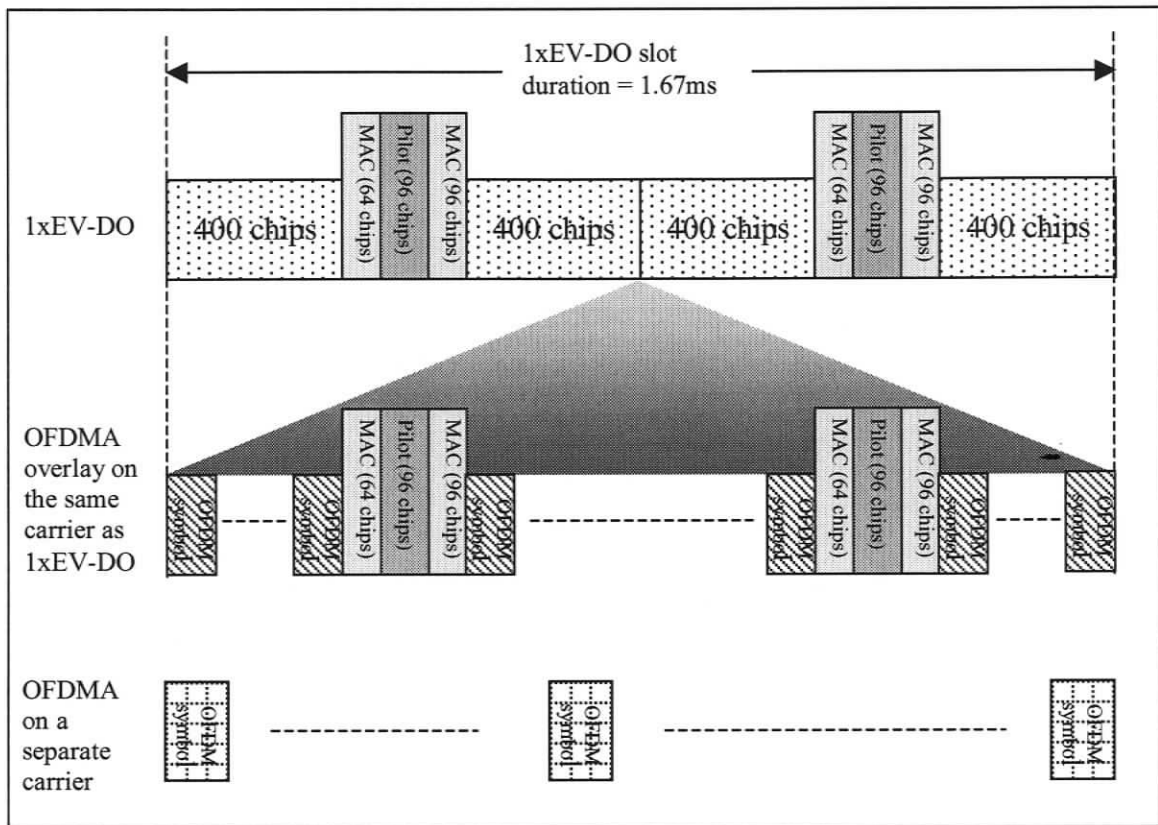


Figure 12 An Illustration of Multi-mode Access System Consists of 1xEV-DO and OFDMA overlay and OFDMA standalone physical layer modes

In this example, the overlay of 1xEV-DO and OFDMA modes onto the same spectrum aims to support a mix of legacy 1xEV-DO terminals and OFDMA terminals while the deployment migrates towards broadband and OFDMA for high speed data services, i.e. the standalone OFDMA mode. The TDM pilot channel and MAC channel in 1xEV-DO are shared by both the 1xEV-DO and the OFDMA overlay modes as common pilot and signalling channels for system acquisition and to carry information regarding the PL mode in use for a particular time slot, as well as the user(s) scheduled on a particular time

slot. The OFDMA standalone mode, on the other hand, does not share common pilot and signalling channels with the 1xEV-DO and the OFDMA overlay modes.

All the abovementioned PL modes share a common layer 2/3 protocol stack based on the existing 1xEV-DO layer 2/3 protocol stack [1]. For the 1xEV-DO and the OFDMA overlay modes, the PL mode and the user(s) scheduled on each time slot can be dynamically changed based on fast layer 2 scheduling across the PL modes. This is illustrated in Figure 13. The fast TDM scheduling at the multiplex/demultiplex sublayer dynamically maps user traffic to different PL modes based on the QoS requirements of the traffic, users' SLA, users' channel conditions, and the PL modes supported by the user terminals.

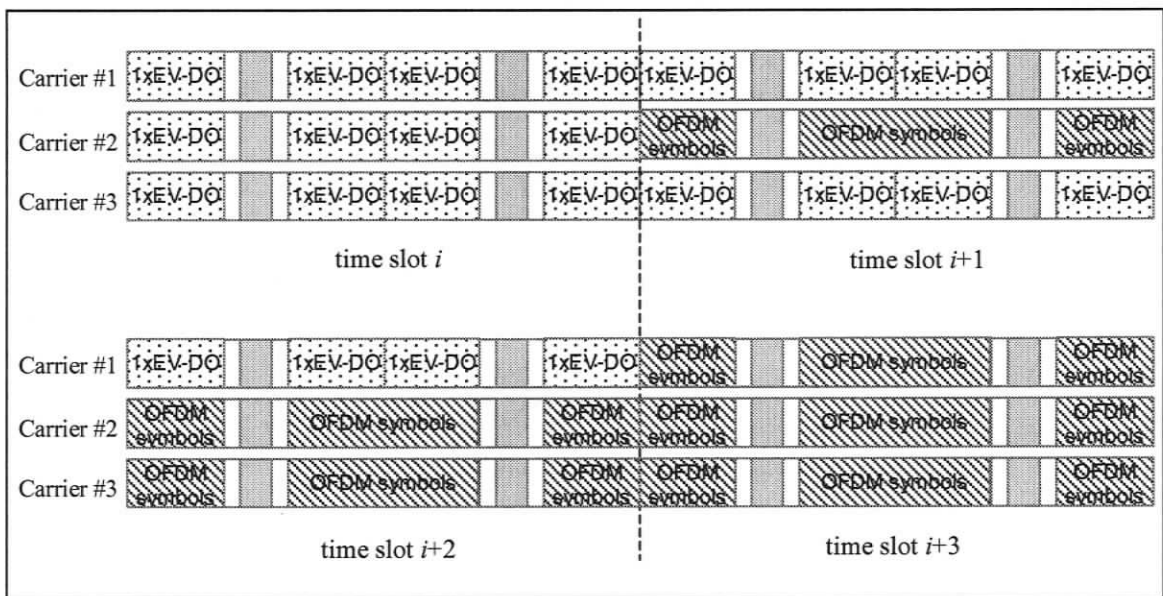


Figure 13 Dynamic Allocation of PL Modes on Different Carriers and Time Slots

4.4 Summary

In this chapter, we proposed a novel multi-mode access system framework for B3G wireless networks. The proposed system framework supports seamless convergence of heterogeneous PL modes optimized for different system deployment scenarios and performance targets envisioned for B3G networks.

In the chapters that follow, we focus on specific components of the proposed system framework, those are, a basic multi-mode access system based on MC DS-CDMA as a first evolution step from 3G towards B3G, dynamic load balancing across PL modes, universal MAC states, and MAC state transition algorithms.

Chapter 5

Multi-Carrier DS-CDMA with Dynamic and Service Driven Bandwidth Allocation

In Chapter 4, we introduced the B3G multi-mode wireless access system framework. In this chapter, explore the evolution of the current 3G DS-CDMA systems towards the proposed B3G system framework.

As an evolution of circuit-oriented 3G wireless systems (e.g. 1xRTT [9]-[11], UMTS Rel. 99 [14]), 1xEV-DO [1], 1xEV-DV [12][13] and HSDPA [2] technologies introduced in the past few years have significantly increased the spectral efficiency of conventional CDMA networks to support packet data services. Much of the spectral efficiency gain of the above systems is obtained through fast TDM scheduling, adaptive modulation and coding, and hybrid ARQ techniques. These techniques exploit the channel variation in the time domain in terms of fast fading and long-term shadowing, to assign the optimum modulation/coding to a user in temporally good channel condition. Thus, they are most beneficial for best effort or delay tolerant data services where opportunistic scheduling can be effectively employed on the TDM fat-pipe channel, e.g. the forward packet data channel (F-PDCH) in 1xEV-DV [12][13] and the high speed downlink shared channel (HS-DSCH) in HSDPA [2]. The sector capacity improvement for best effort services can be as high as three times. However, the above techniques are less effective for delay sensitive services such as VoIP and audio/video streaming since opportunistic scheduling, adaptive modulation/coding assignment and HARQ cannot be effectively

employed when packets with relatively static inter-arrival times have to be delivered within a small delay bound.

To improve the system performance for delay sensitive packet data services and to enable the support of packet data services with highly asymmetrical and variable bandwidth demand envisioned for B3G systems, we propose the MC DS-CDMA. MC DS-CDMA supports concurrent transmissions on N DS-CDMA carriers, where N is scalable based on system bandwidth, and each carrier transmits an independent DS-CDMA waveform. MC DS-CDMA is a fundamentally different concept than the straightforward bandwidth expansion of current CDMA systems in several aspects. First, variable and asymmetrical radio bandwidth (or number of carriers) can be assigned to an MS on the FL and the RL. The number of carriers and the specific carrier(s) assigned to an MS on the FL and the RL can be dynamically changed on a per-scheduling time-slot basis. This flexibility introduces another dimension of fast resource/user multiplexing in the carrier or frequency domain, in addition to the time domain multiplexing as in existing 3G systems. Moreover, MC DS-CDMA also allows for more efficient spectrum sharing and user multiplexing based on service and QoS requirements. Second, the multiple DS-CDMA carriers can be viewed as a basic form of the multiple PL modes anchored by a common layer 2/3 protocol stack as proposed in Chapter 4. The service-driven layer 2/3 protocol design efficiently multiplex user traffic across multiple carriers and to exploit both multi-carrier diversity and multi-user diversity. Third, MC DS-CDMA can provide backward compatibility to existing single carrier DS-CDMA systems, thus allowing for the overlay of existing single carrier DS-CDMA system with

MC DS-CDMA while the network deployment migrates towards B3G broadband support.

In Section 5.1, we provide the detailed MC DS-CDMA system description. Section 5.2 illustrates the performance of the MC DS-CDMA system. This is followed by some concluding remarks in Section 5.3.

5.1 MC DS-CDMA System Description

In this section, we present a detailed description of the proposed MC DS-CDMA system.

5.1.1 System Overview

For illustration purpose, we choose 1xEV-DV as a reference for the single carrier DS-CDMA system. For simplicity, we call the MC 1xEV-DV system as MC-DV. Note that similar multi-carrier enhancement can be applied to other single carrier DS-CDMA systems such as 1xEV-DO and HSDPA.

The baseband transmitter block diagram of MC-DV is shown in Figure 14. Each carrier transmits an independent DS-CDMA waveform at 1.2288Mcps. Adaptive modulation and coding is performed independently on each carrier for each time slot based on the channel condition experienced by the user scheduled on that carrier and the buffer occupancy of the user's traffic. At each scheduling time slot, the cross-carriers layer 2 scheduler / packet multiplexer maps the layer 2 packets of different users onto the N carriers. The mapping aims to optimize the system capacity while meeting the QoS requirement of each user's service. More details of the service driven layer 2/3 protocols are given in Section 5.1.2.

Figure 15 shows the overlay of 1xEV-DV/1xRTT operable carriers and MC-DV carriers over the $N \times 1.25\text{MHz}$ spectrum allocated to the system. The 1xEV-DV/1xRTT operable carriers serve the legacy 1xEV-DV/1xRTT terminals, and also serve as the primary carriers for MC-DV terminals for system access from the Null state or the Dormant state, because the 1xEV-DV/1xRTT operable carriers contain the full system access channels (e.g. paging, sync, random access). The primary carriers can also provide voice, circuit data and packet data services to the MC-DV terminals. The MC-DV-only carriers, called the supplemental carriers, serve only the MC-DV terminals in the Active state. The supplemental carriers are optimized for high-speed packet data operation whereby all the Walsh codes and power in the FL excluding those used by the pilot channel are used for the high-speed packet data channels.

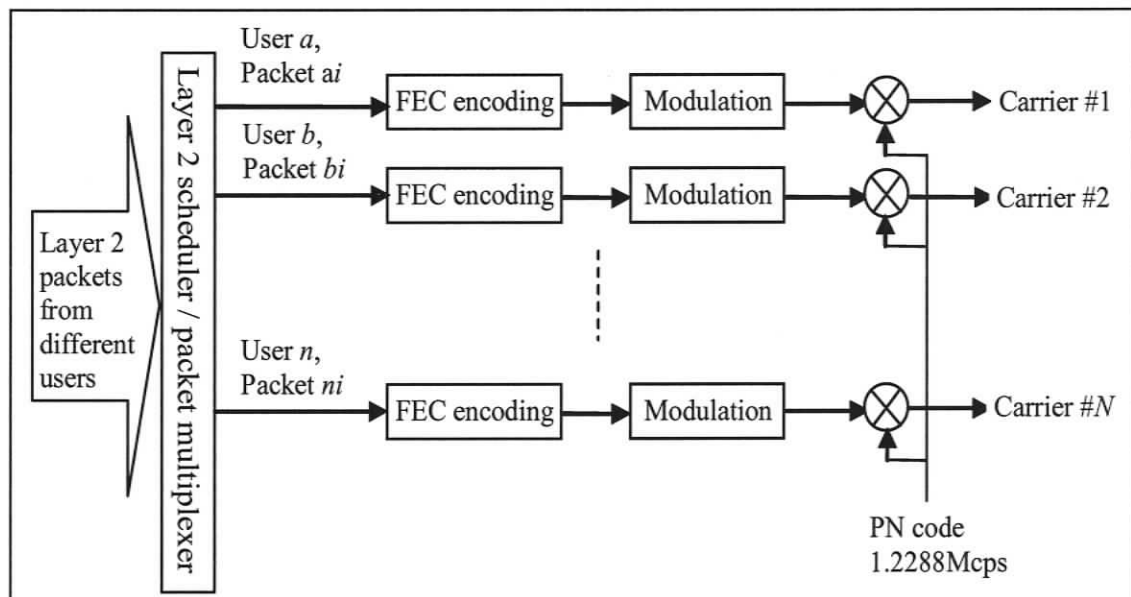


Figure 14 Baseband Transmitter Block Diagram of MC-DV

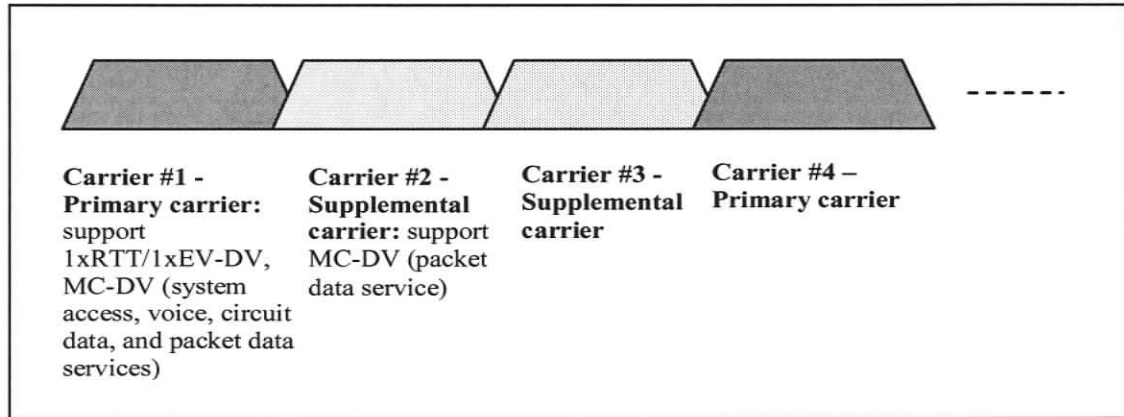


Figure 15 MC-DV Carriers Configuration (An Illustration)

A unique feature of MC-DV is that it supports asymmetric radio bandwidth allocation to a particular MS in the FL and the RL according to the service requirements. This is illustrated in Figure 16. For example, to support FL intensive applications such as real-time movie downloading, three FL carriers and one RL carrier can be assigned to the user. To support RL streaming services, one FL carrier and three RL carriers can be assigned to the user. To support applications that require symmetrical high rate FL and RL such as interactive gaming, three carriers can be assigned to the user on both the FL and the RL. On the other hand, conventional voice users can be assigned a single carrier on both links. The modes can be changed dynamically for each MS based on the service requirements and traffic loading.

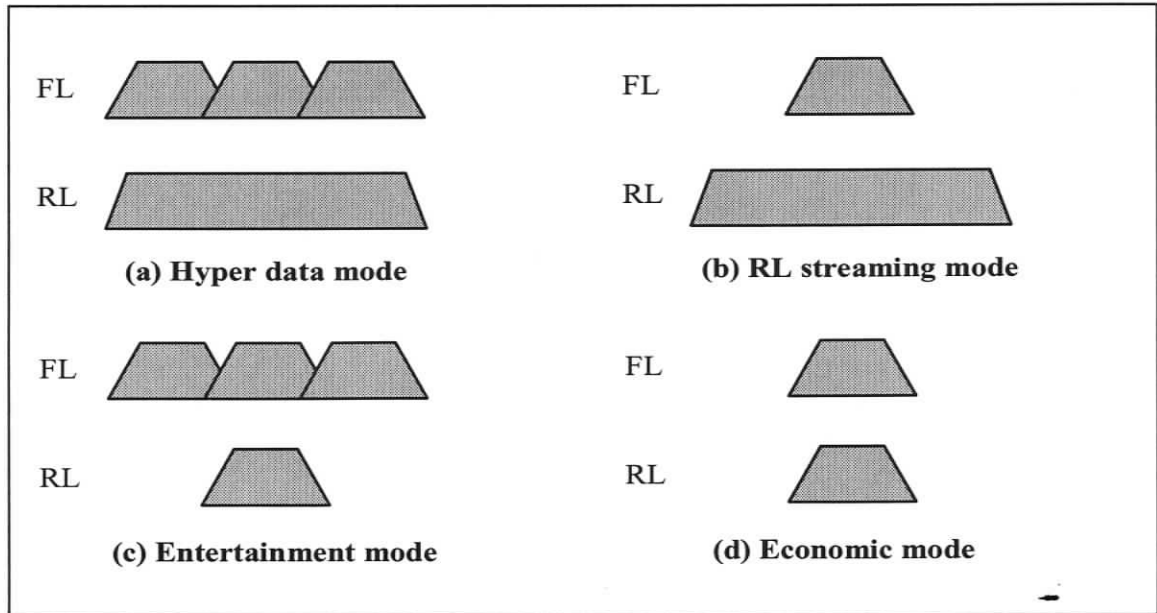


Figure 16 An Illustration of Different Bandwidth Allocation Modes

5.1.2 Service Driven Protocol Design

MC-DV supports fast dynamic resource allocation in the time domain in terms of TDM slots and in the carrier domain in terms of the number of carriers and the specific carrier(s) assigned to each MS. The resource allocation scheme at the layer 2 scheduler is driven by two factors: 1) to meet the service requirements of the MS, and 2) to maximize the overall system capacity.

The protocol structure similar to that of Figure 8 is shown in Figure 17. In this case, the common layer 2 and layer 3 protocol stack supports different DS-CDMA based PL modes, i.e., 1xRTT, 1xEV-DV and MC-DV.

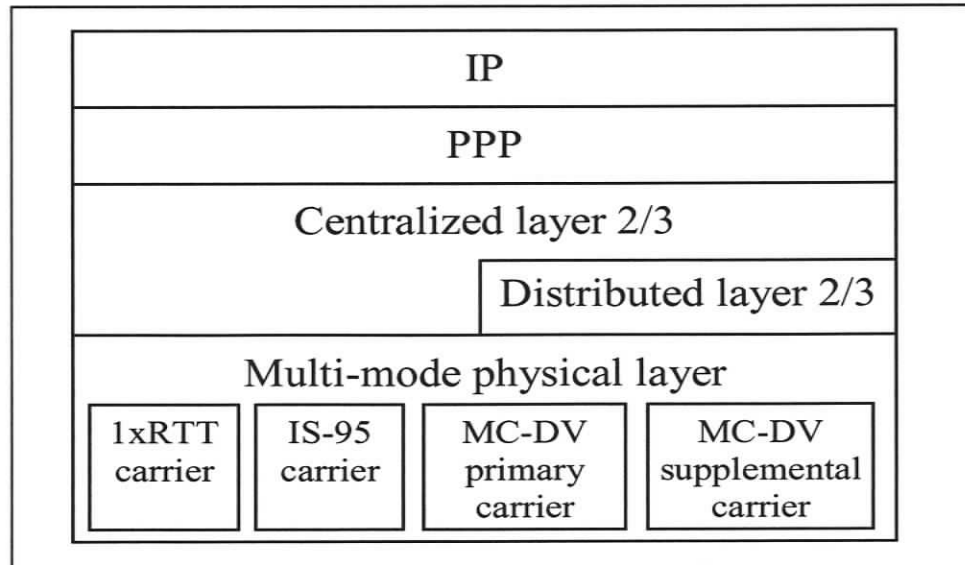


Figure 17 Anchor Layer 2/3 Protocol Design for Different DS-CDMA modes: MC-DV, 1xRTT, IS-95

5.1.3 System Operation

A detailed description of the MC-DV system operation is given below.

5.1.3.1 System Access

When in the power-on state or in the Dormant state, an MC-DV MS monitors the primary carrier(s) for paging information from the network. To initiate a call, an MS performs random access on the primary carrier(s). The particular primary carrier(s) to monitor or perform system access can be determined based on hashing on the MS's International Mobile Station Identity (IMSI) or Electronic Serial Number (ESN).

5.1.3.2 High Speed Packet Data Operation

On the primary carrier(s), 1xRTT, 1xEV-DV and MC-DV MSs share the same code and power space in the FL, and share the same interference or Rise-Over-Thermal (ROT) space in the RL. 1xEV-DV and MC-DV MSs may share the same high-speed packet data channels, i.e. F-PDCH(s) on the primary carrier or they may each have independent F-

PDCH(s) on the primary carrier. The layer 2 scheduler performs fast TDM scheduling on the pools of F-PDCHs across multiple carriers. To select an MS' packet for transmission on a particular time slot, the scheduler prioritizes each of the MSs' packets according to the channel condition experienced by the MS on each of the associated carriers, and the QoS requirements (e.g. delay bound, minimum data rate, etc.) of the packets. The highest priority packet on each carrier is selected for transmission on that carrier. An MC-DV MS can be scheduled to transmit on one or more primary or supplemental carriers on a particular time slot. The prioritization and scheduling of HARQ retransmission packets is treated in a similar way as new layer 2 packets, whereby the HARQ retransmission packets can be scheduled to transmit on any of the associated primary and supplemental carriers. A detailed account of the multi-carrier layer 2 scheduler is given in Section 5.1.4.

In addition to fast scheduling, dynamic load balancing is performed across the carriers to ensure optimum loading of different types of MSs, e.g. 1xRTT, 1xEV-DV and MC-DV MSs, on the primary and the supplemental carriers.

5.1.4 Multi-carrier Scheduler

We propose a multi-carrier scheduler that performs cross carriers scheduling. The objective of the multi-carrier scheduler is to maximize system capacity while ensuring traffic QoS and users' SLA are met. To support different QoS and SLA requirements, we broadly classify traffic into delay sensitive and delay tolerant. For delay sensitive traffic, an air-interface delay bound is defined such that a packet arrives at the layer 2 scheduler buffer will be discarded if it is not transmitted over-the-air within the delay bound. The value of the delay bound is chosen such that the end-to-end network delay experienced by

each packet is within an acceptable value. As an example, for voice traffic, such as VoIP, the acceptable mouth-to-ear delay is around 150ms to 200ms [32]. Given that the total access network delay, core network delay, public internet delay and de-jitter buffer delay is typically ~ 100 ms, the over-the-delay bound is set at less than 100ms. This delay bound is an input to the multi-carrier scheduler when calculating the priority of a packet. For delay tolerant services, there is no need to define delay bound. Instead, per-user average throughput, minimum throughput defined in SLA and overall fairness among users in the system are of importance and should be used as input to the scheduler for users' packets priority calculation.

In summary, the priority of packet j of user i at time slot n and carrier k can be expressed as

$$(4) \quad p_{ij}^{(k)}(n) = f(r_i^{(k)}(n), \bar{r}_i(n), d_{ij}(n), \min_ r_i),$$

which is a function of $r_i^{(k)}(n)$, the instantaneous data rate supported by user i at time slot n and carrier k based on the instantaneous channel condition at carrier k ; $\bar{r}_i(n)$, the average data rate experienced by user i up to time slot n ; $d_{ij}(n)$, the queuing delay experienced by packet j of user i at time slot n (applicable to delay sensitive service); and $\min_ r_i$, the minimum guaranteed data rate of user i based on SLA. The scheduler computes the priority of all the users' packets in the layer 2 buffer for the corresponding carrier(s) supported by the users. For a 1xEV-DV user, k is the designated primary carrier that can transmit data to the user and is configured during call setup based on loading condition of all the available carriers. For an MC-DV user, k belongs to the set of carriers (primary and/or supplemental carriers) that can transmit data to the user. The set of

carriers is configured during call setup based on loading condition of all the available carriers, and the user terminal's capability.

To maximize the system capacity while ensuring traffic QoS and users' SLA are met, the scheduler finds the packet with the maximum priority as follows,

$$(5) \quad [i^1, j^1, k^1] = \arg \max_{i, j, k} \{p_{ij}^{(k)}(n)\}.$$

The selected packet $[i^l, j^l]$ will be transmitted on carrier k^l . The scheduler then finds the next highest priority packet where

$$(6) \quad [i^2, j^2, k^2] = \arg \max_{i \neq i^1, j \neq j^1, k} \{p_{ij}^{(k)}(n)\}.$$

The above process is repeated until all the packets are scheduled or resources in all the carriers are used up. This global scheduling and priority sorting approach allow optimum multiplexing of 1xEV-DV users and MC-DV users that support different number of carriers. This is further illustrated in Figure 18.

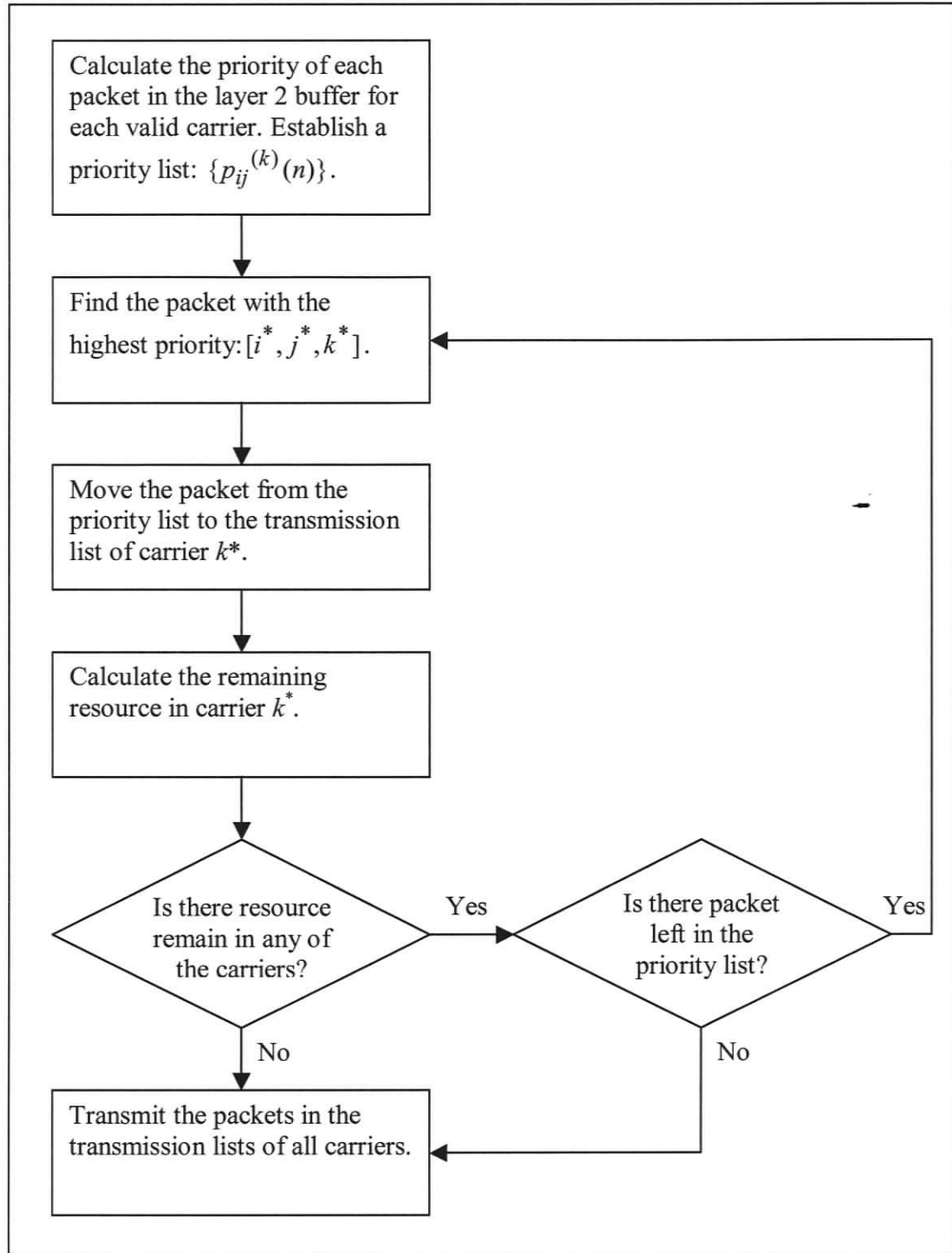


Figure 18 Global Multiplexing and Scheduling of 1xEV-DV and MC-DV Users at Time Slot n .

Some common scheduling priority equations for wireless systems can be found in [36]. One of them is the well known proportional fairness scheduler [36][37] which has the following formulation after expanding into multi-carrier representation:

$$(7) \quad p_{ij}^{(k)}(n) = \frac{r_i^{(k)}(n)}{\bar{r}_i(n)},$$

where $\bar{r}_i(n)$ is a running-average throughput and can be obtained through an IIR filtering operation as follows,

$$(8) \quad \bar{r}_i(n) = \left(1 - \frac{1}{t_c}\right)\bar{r}_i(n-1) + \frac{1}{t_c}s_i(n-1),$$

where t_c is the filter time constant which is related to the maximum amount of time for which an individual user can be starved and $s_i(n-1)$ is the overall data rate transmitted to user i in slot $n-1$.

The proportional fairness scheduler is designed for delay tolerant traffic since it leverages the instantaneous channel condition of each user to exploit multi-user diversity. The relative throughput fairness among users can be adjusted by generalizing the proportional fairness scheduler as follows,

$$(9) \quad p_{ij}^{(k)}(n) = \frac{[r_i^{(k)}(n)]^\alpha}{[\bar{r}_i(n)]^\beta},$$

where α/β ratio defines the relative emphasis on instantaneous data rate versus long-term average throughput. The larger the α/β ratio, the less fair the scheduler, and vice versa.

As discussed earlier, to support delay sensitive services, packet queuing delay has to be used as an input to the priority calculation. An integrated fairness and delay scheduler is

proposed in [38]. The priority equation with multi-carrier expansion is expressed as follows:

$$(10) \quad P_{ij}^{(k)}(n) = F_i^{(k)}(n) + D_{ij}(n),$$

where $F_i^{(k)}(n)$ is the fairness component expressed as

$$(11) \quad F_i^{(k)}(n) = \frac{[r_i^{(k)}(n)]^\alpha}{[1 + \bar{r}_i(n)]^\beta r_{\max}^\alpha},$$

and r_{\max} is the peak data rate of the system, i.e.,

$$(12) \quad r_i^{(k)}(n) \leq r_{\max} \quad \forall n, k.$$

It is used to normalize the $F_i^{(k)}(n)$ term to less than or equal to one.

$D_{ij}(n)$, the delay component of the priority equation is given by

$$(13) \quad D_{ij}(n) = \begin{cases} f(t_n - t_{\min,ij}, t_{\max,ij}) + F_{\max}, & \text{if } t_n \geq t_{\min,ij} \\ 0, & \text{otherwise} \end{cases}$$

where t_n is the system time corresponds to time slot n , $t_{\min,ij}$ is the minimum system time for which delay sensitive packet j of user i has a higher priority than other delay tolerant packets with only the fairness priority component, $t_{\max,ij}$ is the latest time that the j^{th} packet for user i can be sent and F_{\max} is the maximum value of the fairness priority component. The value for $t_{\min,ij}$ is set according the packet arrival time at the layer 2 buffer and depends on the expected number of slots needed to successfully transmit the packet and on the total number of delay sensitive users in the system that have the delay

bound close to the same time as packet j from user i . One simple formulation of the delay component is as follows,

$$(14) \quad f(t_n - t_{\min,ij}, t_{\max,ij}) = \frac{t_n - t_{\min,ij}}{t_{\max,ij} - t_{\min,ij}},$$

$$t_{\min,ij} = t_{\max,ij} - (\text{number of delay sensitive users}) \\ \times (\text{max. number of transmission slots per packet}) \\ \times (\text{max. number of transmissions attempts per packet}) \times \text{slot_duration}$$

and

$$(15) \quad F_{\max} = 1.$$

A graphical representation of the fairness and delay scheduler is shown in Figure 19.

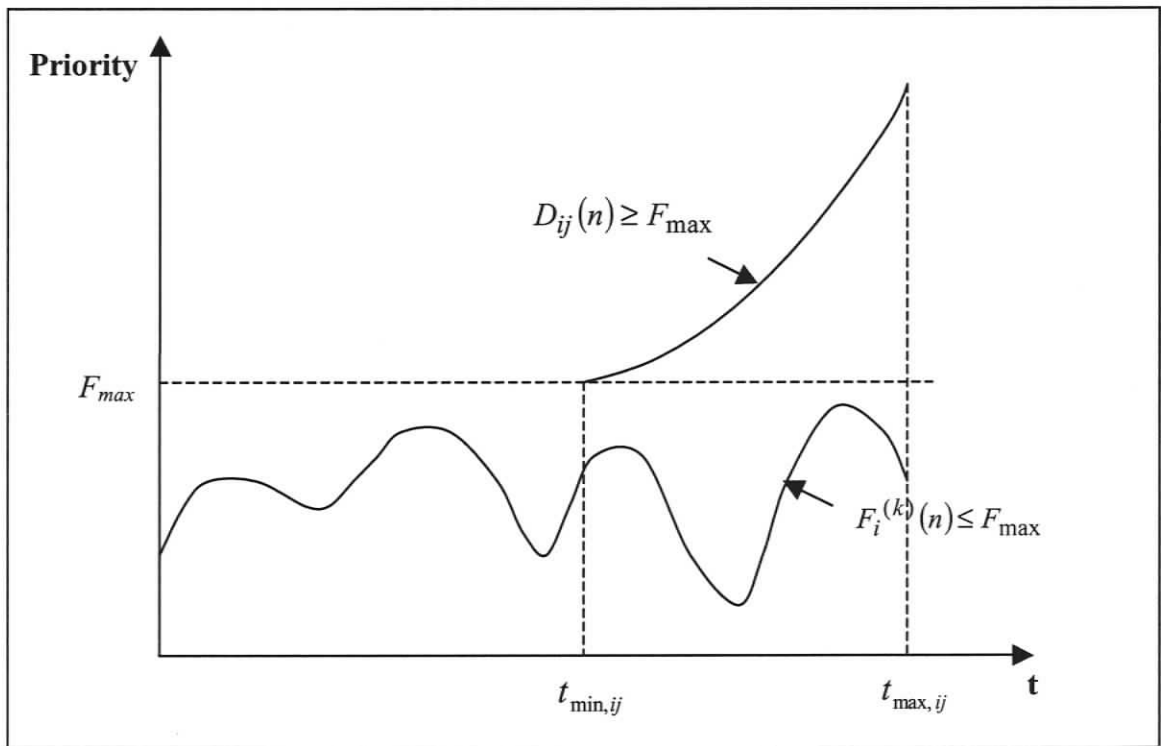


Figure 19 An Illustration of the Fairness and Delay scheduler

In the subsequent performance studies in this chapter and in other chapters, we use the proportional fairness scheduler for delay tolerant services and the fairness and delay scheduler for delay sensitive services.

5.2 Performance Evaluation

In this section, we evaluate different aspects of the MC-DV system performance. In Section 5.2.1, we illustrate the multi-carrier diversity benefit of cross-carrier scheduling. In Section 5.2.2, we compare the system performance of MC-DV and 1xEV-DV to demonstrate the benefit of MC-DV over a simple carriers aggregation of 1xEV-DV carriers.

5.2.1 Multi-carrier Diversity

The fading between carriers is dependent on the carrier separation and delay spread of the channel, as given in the equation below [39].

$$(16) \quad \rho(\Delta f) = \frac{1}{\sqrt{1 + (2\pi\Delta f)^2 \sigma^2}},$$

where $\rho(\Delta f)$ is the correlation coefficient which is a function of Δf , the carrier separation and for a given σ^2 , the RMS delay spread of the channel. The plot of correlation coefficient between two carriers separated by 1.25MHz versus the RMS delay spread is given in Figure 20. The typical range of delay spread for macro-cellular environment is between $0.4\mu s$ and $20\mu s$ [40].

Figure 21, Figure 22, Figure 23 and Figure 24 plot the time domain evolution of the Carrier-to-Interference Ratio (CIR) of three adjacent 1.25MHz carriers at geometry of 6.5dB, for RMS delay spread of $0.01\mu s$, $0.05\mu s$, $0.1\mu s$ and $0.5\mu s$ respectively.

Geometry is defined as the long-term average received CIR at an MS due to path loss, log-normal shadowing and antenna gain, while excluding fast fading. As demonstrated in these figures, the CIR variation between carriers increases as the delay spread increases. With a delay spread of $0.5\mu\text{s}$ or larger, the CIR variation between carriers allows opportunistic scheduling to schedule a user on a carrier that is in up-fade. This form of multi-carrier diversity is analogous to the multi-user diversity in a single carrier 1xEV-DO, 1xEV-DV or HSDPA systems. In the next section, we show the multi-carrier diversity gain of MC-DV on top of the multi-user diversity that is present in single carrier systems.

5.2.2 Performance Comparison of MC-DV and 1xEV-DV

In this section, we evaluate and compare the FL performance of MC-DV with that of 1xEV-DV. The MC-DV system consists of three 1.25MHz carriers. For fair comparison, we normalize the capacity of MC-DV to per 1.25MHz when compared with 1xEV-DV capacity. We assume an RMS delay spread of $0.5\mu\text{s}$ for the performance evaluation.

A full system-level simulation is performed based on the methodology described in Chapter 3, i.e. 19 tri-sector cells, modified Hata path loss model, log-normal shadowing, fast fading with mixed mobile speeds ranging from stationary to 120 km/h. All three MC-DV carriers are configured as primary carriers. Both MC-DV and 1xEV-DV share the same F-PDCH on each carrier. Fast scheduling and rate adaptation per 1.25ms is performed for both 1xEV-DV and MC-DV. Performance is evaluated for both the full-buffer traffic case and the non-full-buffer traffic cases, i.e. FTP, HTTP, streaming video and real-time interactive video. Full TCP/IP protocols are modeled for FTP and HTTP.

5.2.2.1 Full Buffer Traffic

For full buffer traffic, since delay is not a concern and the users' buffers are always full, the multi-user diversity can be fully exploited by the scheduler for a single carrier system. With multi-carrier scheduling, there is further performance gain due to multi-carrier diversity. In this section, we investigate this additional performance gain. The performance metric of interest for full buffer traffic is the sector throughput normalized to 1.25MHz. The proportional fairness scheduler as described in equation (9) with $\alpha = 1$ and $\beta = 1$, is used for both 1xEV-DV and MC-DV.

Figure 25 shows the normalized per-carrier sector throughput of MC-DV with different number of carriers. 20 users are assumed per sector. Note that when the number of carriers equals to one, the system is identical to 1xEV-DV. Table 9 below summarizes the relative gain of MC-DV over 1xEV-DV for different number of MC-DV carriers. We can see that the largest relative gain is obtained by increasing the number of carriers from one to two. The incremental gain decreases as the number of carriers increases and is close to saturation when the number of carriers equals to five.

Table 9 **Relative Per-Carrier Sector Throughput Gain of MC-DV over 1xEV-DV**

Number of MC-DV Carrier	Relative Per-Carrier Sector Throughput Gain of MC-DV over 1xEV-DV
1	1
2	1.08
3	1.11
4	1.13
5	1.14

To ensure the scheduler operates under the same system fairness for a side-by-side comparison, we plot the normalized per-user throughput CDFs of 1xEV-DV, MC-DV with three carriers and MC-DV with five carriers in Figure 26. All three cases exhibit the same fairness and meet the fairness criteria defined in [27].

We further examine the per-user throughput of 1xEV-DV versus MC-DV. As shown in Figure 27 which provides the scatter plot of user throughput (normalized to per-carrier) versus geometry, MC-DV provides per-user throughput gain for different geometry, in particular in high geometry region which the proportional scheduler emphasizes.

5.2.2.2 FTP Traffic

FTP traffic characteristic is very similar to full buffer, except that due to TCP slow start operation, a user's buffer at the layer 2 scheduler may not always be full. In this case, the multi-user or multi-carrier diversity may not be fully exploited. Here, our objective is to evaluate and compare the performance of MC-DV and 1xEV-DV with FTP traffic and investigate how TCP may impact the multi-user and multi-carrier diversity as compared

to the full buffer case. The performance metrics of interest are the sector throughput for a given number of FTP users, and the system outage versus the number of FTP users, where system outage definition for FTP is as described in Chapter 3. Since FTP is delay tolerant, the same proportional fairness scheduler as in the full buffer case is used.

Figure 28 shows the sector throughput per carrier for different number of FTP users per sector. MC-DV provides about 10% gain over 1xEV-DV. This gain is similar to the gain observed in the full buffer case. This means that the multi-carrier diversity gain of MC-DV is not affected by the TCP operation for the FTP case. Comparing Figure 28 and Figure 25, it is also observed that the sector throughput for FTP is a bit lower than that of the full buffer. This is due to the TCP slow-start operation that reduces the overall multi-user and multi-carrier diversity of the system.

Figure 29 shows the system outage for different number of FTP users per carrier. For a 2% system outage criterion defined in [27], the system capacity is 33 users and 38 users per carrier respectively for 1xEV-DV and MC-DV. Therefore, the user capacity gain of MC-DV over 1xEV-DV is about 15%. The reason for the higher user capacity gain compared to sector throughput gain is because the multi-carrier diversity and trunking efficiency of MC-DV improves the overall coverage of the system, i.e. a user has less probability to be trapped in a bad CIR situation.

5.2.2.3 HTTP Traffic

The HTTP traffic is much more bursty than the FTP traffic, due to the nature of web-browsing activity. Our objective here is to evaluate the impact of bursty traffic on multi-user and multi-carrier diversity. To more accurately represent the actual system operation for bursty traffic, we include the modeling of the 1xEV-DV MAC states, i.e., the

Dormant state, the Control-hold state and the Active state, into the system level simulation of 1xEV-DV and MC-DV. A user is put into either the Control-hold state or the Dormant state when its buffer is empty for a predefined period of time. More detailed account of the characteristics of these MAC states and the state transition algorithm is given Chapter 7 and Chapter 8. The following parameters are used in the simulation:

- Dormant state to Active state transition latency: 2 seconds
- Control-hold state to Active state transition latency: 40ms
- Buffer empty timer value to trigger Active state to Control-hold state transition: 200ms
- Buffer empty timer value to trigger Control-hold state to Dormant state transition: 5.8 seconds

The performance metrics of interest are the sector throughput for a given number of HTTP users, and the system outage versus number of HTTP users, where system outage definition for HTTP is as described in Chapter 3. Since HTTP is delay tolerant, the same proportional fairness scheduler as in the full buffer case is used.

Figure 30 shows the system outage and sector throughput per carrier for different number of HTTP users per carrier. For a 2% system outage criterion defined in [27], the system capacity is 72 users and 90 users per carrier respectively for 1xEV-DV and MC-DV. Therefore, the user capacity gain of MC-DV over 1xEV-DV is about 25%. The corresponding sector throughput gain of MC-DV over 1xEV-DV at the capacity point is about 22%. The higher gain of MC-DV over 1xEV-DV compared to the full buffer case and the FTP case is because of the improved trunking efficiency or multi-carrier

multiplexing gain of MC-DV which reduces probability of traffic congestion for bursty traffic.

5.2.2.4 Video Traffic

Here, we investigate the performance of 1xEV-DV and MC-DV with delay sensitive service. We examine the performance of video traffic with two different delay bound criteria, i.e., 5 seconds as defined for NRTV or streaming video in [27] and 100ms to represent interactive real-time video service.

For delay sensitive service, multi-user diversity can not be fully exploited since pure opportunistic scheduling without delay consideration cannot be used. Multi-carrier diversity will therefore play an important role in increasing the likelihood of a user's packet being scheduled under a favourable channel condition. We use the multi-carrier fairness and delay scheduler described in Section 5.1.4 in our simulation.

The performance metric of interest is the system outage versus the number of video users, where system outage definition for streaming video is as described in [27]. System outage for real time interactive video is based on that of streaming video except that the delay bound criterion is reduced from 5 seconds to 100ms.

Figure 31 shows the system outage versus number of streaming video users. For a 2% system outage criterion defined in [27], the system capacity is 16 users and 22 users per carrier respectively for 1xEV-DV and MC-DV. Therefore, the streaming video user capacity gain of MC-DV over 1xEV-DV is about 38%. The higher gain compared to HTTP, FTP or full buffer cases is expected because of the delay bound requirement of streaming video service that reduces the effectiveness of multi-user diversity. Multi-

carrier diversity and trunking efficiency thus plays an increasingly important role in reducing the packet queuing delay at the layer 2 buffer.

For the case of real-time interactive video as shown in Figure 32, the system capacity is 3 users and 8 users per carrier respectively for 1xEV-DV and MC-DV. The corresponding user capacity gain of MC-DV over 1xEV-DV is about 2.7 times. This gain is much larger than that of the streaming video due to the tight delay bound of 100ms. With such a small delay bound, there is no room to take advantage of multi-user diversity for the case of 1xEV-DV. MC-DV, on the other hand, benefits from multi-carrier diversity and trunking efficiency to significantly reduce the packet queuing delay and correspondingly the per-user outage and system outage.

The sector throughput per carrier corresponds to the achievable user capacity for streaming video and real-time interactive video is shown in Figure 33. The sector throughput gain of MC-DV over 1xEV-DV is 41% and 2.7 times respectively for streaming video and real-time interactive video. The gains are in line with the gains of user capacity observed earlier.

5.2.2.5 Summary of Performance Comparison

The normalized per-carrier system capacity gain of MC-DV with three carriers over 1xEV-DV for different services is summarized in Table 10. Overall, the proposed MC-DV system with QoS and service-driven cross-carrier scheduler provides a significant performance advantage over the conventional 1xEV-DV system. We can observe that the burstier the traffic the higher the MC-DV gain. Also, the more delay sensitive the service, the higher the MC-DV gain. Similar performance advantage of other forms of multi-carrier DS-CDMA system can be expected, e.g., multi-carrier DO (MC-DO), an

expansion of 1xEV-DO and multi-carrier HSDPA (MC-HSDPA), an expansion of HSDPA.

Table 10 Summary of Normalized Per-Carrier System Capacity Gain of MC-DV (3x) over 1xEV-DV

Traffic Type	Normalized Per-Carrier Sector Throughput Gain	Normalized Per-Carrier User Capacity Gain
Full buffer	11%	N/A
FTP	10%	15%
HTTP	22%	25%
Streaming video	41%	38%
Real-time interactive video	2.7 times	2.7 times

5.3 Summary

In this chapter, we introduced the MC DS-CDMA system concept as the first evolution step of the current 3G cellular systems towards the proposed B3G multi-mode wireless access system framework. We provided a detailed description of the proposed MC DS-CDMA system including the system configuration, operation and the service-driven common layer 2/3 protocol stack that anchor multiple DS-CDMA carriers. A cross-carrier layer 2 scheduler was proposed to optimize system capacity while ensuring the QoS of different services are met. Detailed performance studies were conducted to evaluate the performance of MC-DV and compared with that of 1xEV-DV. Our studies showed that MC-DV provides significant performance and capacity gain over 1xEV-DV, in particular for bursty traffic and delay sensitive traffic which are key traffic characteristics of 3G and B3G wireless systems.

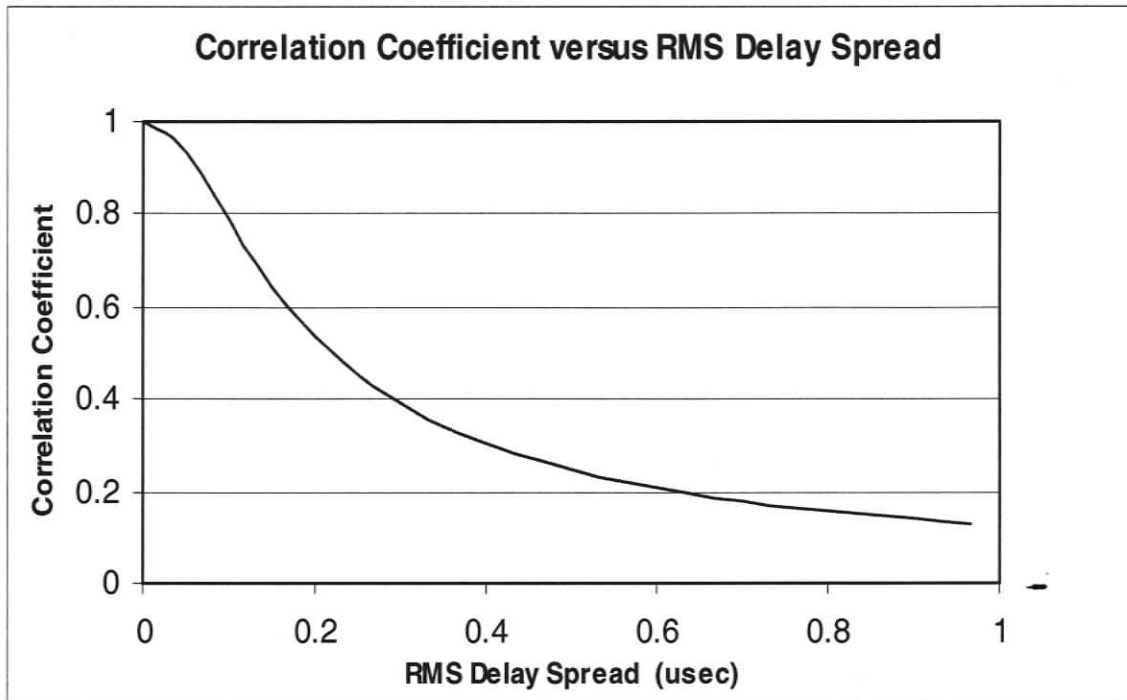


Figure 20 Correlation Coefficient versus RMS Delay Spread for Carrier Separation of 1.25MHz

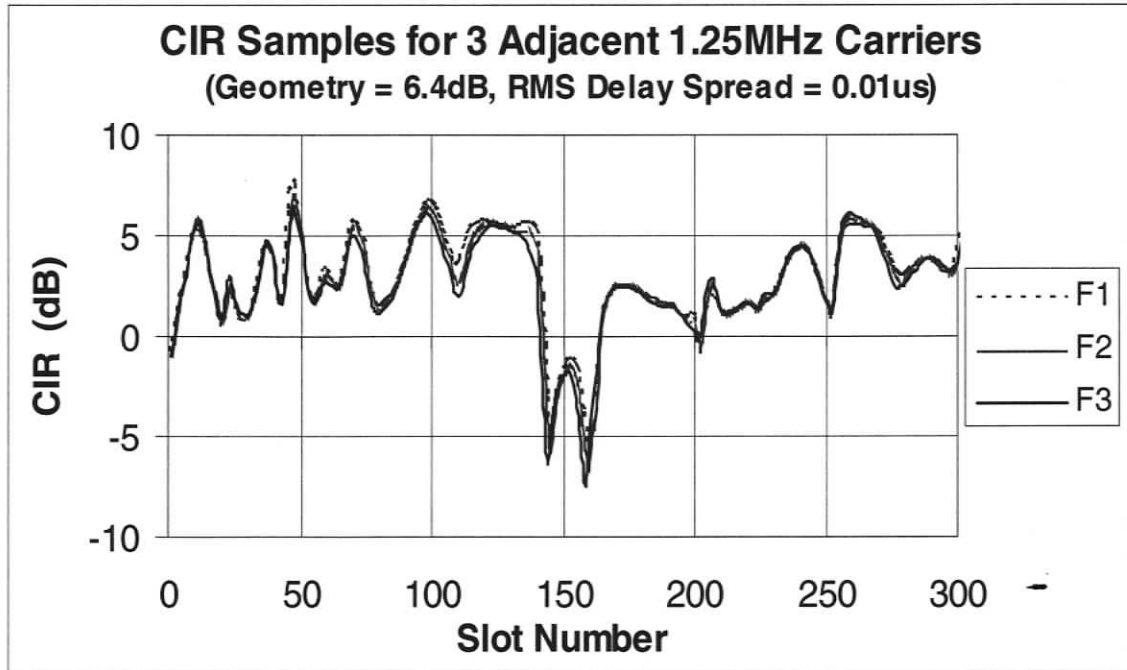


Figure 21 CIR Samples of Three Adjacent Carriers (F1, F2, F3), Each Separated by 1.25MHz, for Geometry of 6.4dB, RMS Delay Spread of 0.01 μ s, and Time Slot Duration of 1.25ms

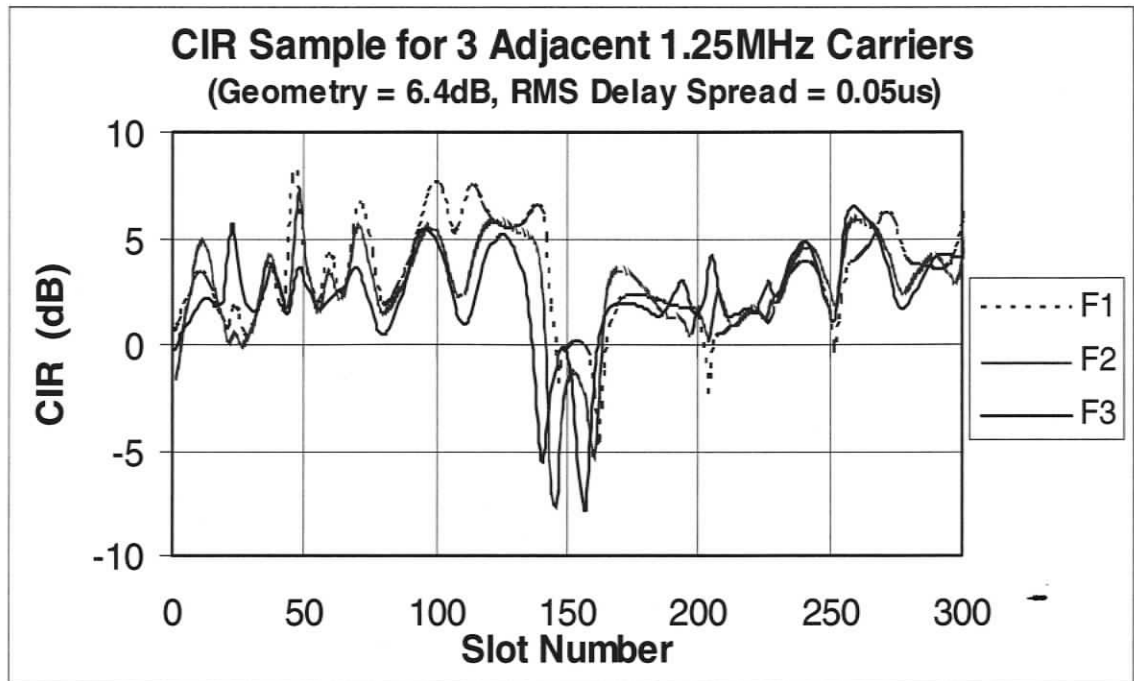


Figure 22 CIR Samples of Three Adjacent Carriers (F1, F2, F3), Each Separated by 1.25MHz, for Geometry of 6.4dB, RMS Delay Spread of $0.05\mu s$, and Time Slot Duration of 1.25ms

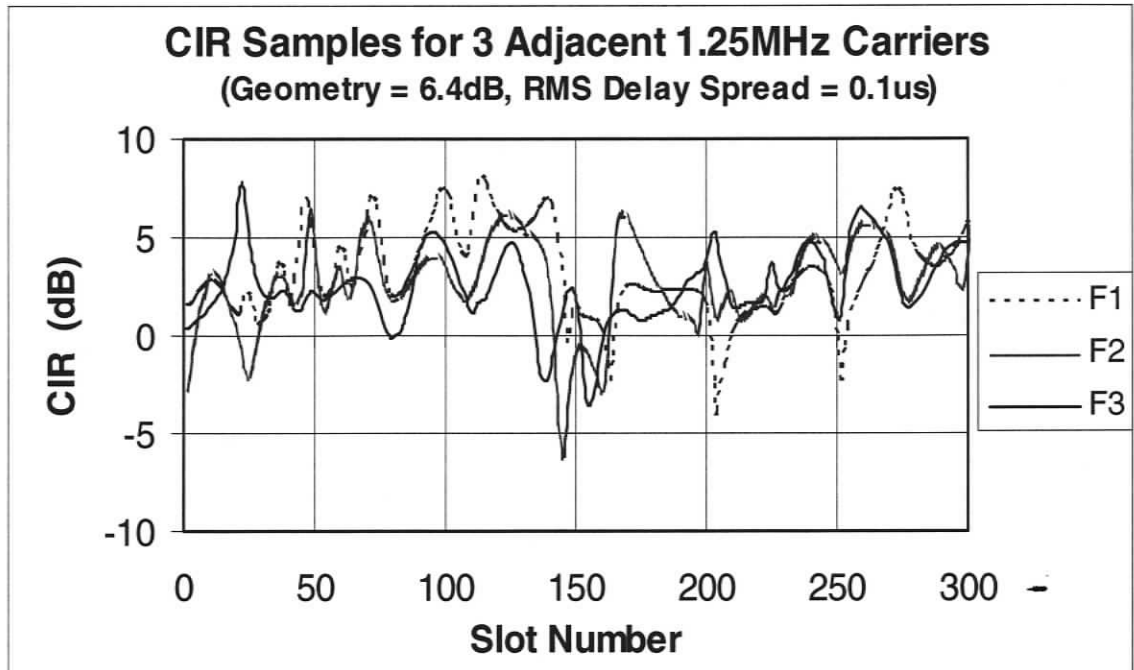


Figure 23 CIR Samples of Three Adjacent Carriers (F1, F2, F3), Each Separated by 1.25MHz, for Geometry of 6.4dB, RMS Delay Spread of 0.1 μ s, and Time Slot Duration of 1.25ms

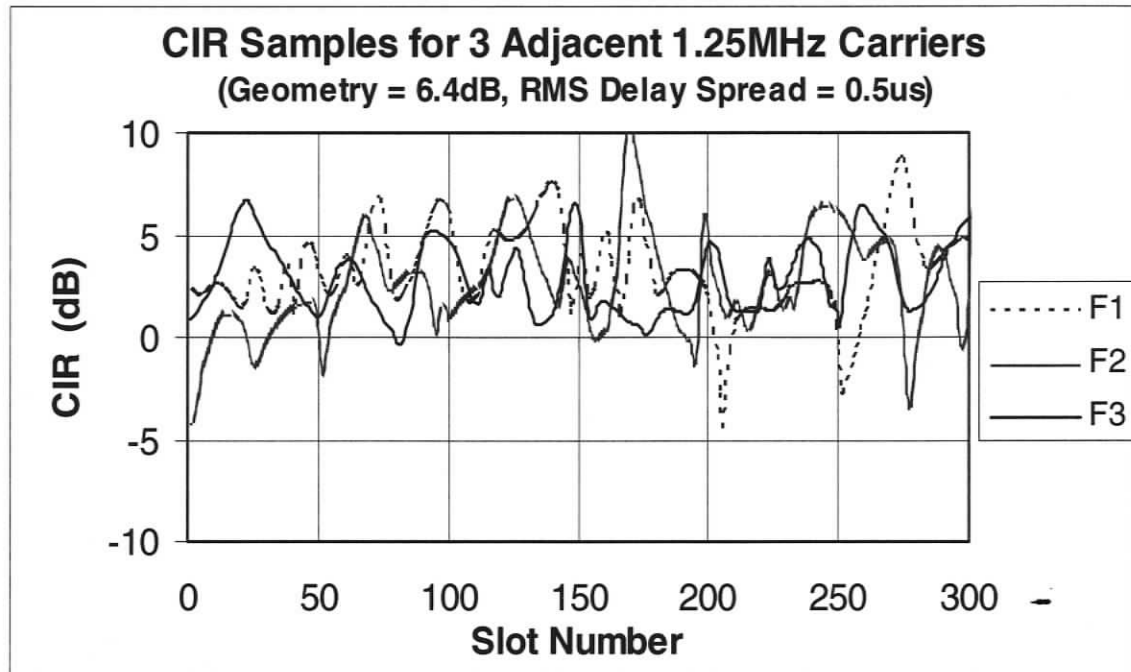


Figure 24 CIR Samples of Three Adjacent Carriers (F1, F2, F3), Each Separated by 1.25MHz, for Geometry of 6.4dB, RMS Delay Spread of $0.5\mu s$, and Time Slot Duration of 1.25ms

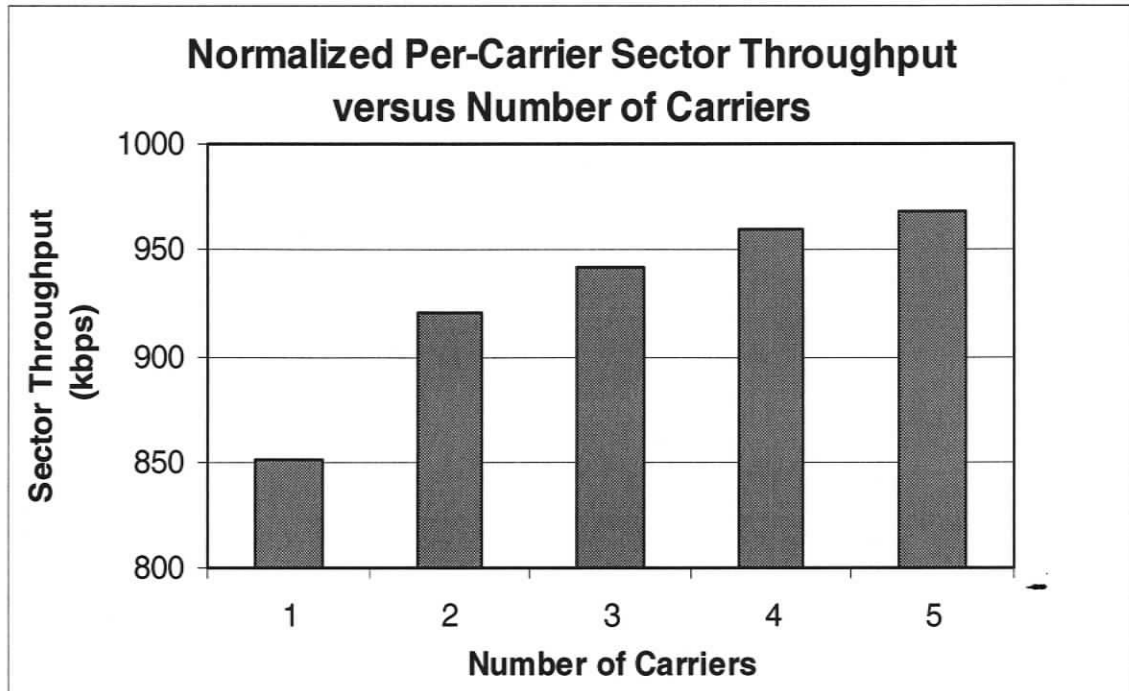


Figure 25 Sector Throughput per Carrier of MC-DV versus Number of Carriers, for Full Buffer Traffic with Mixed Mobile Speeds and 20 Users per Sector

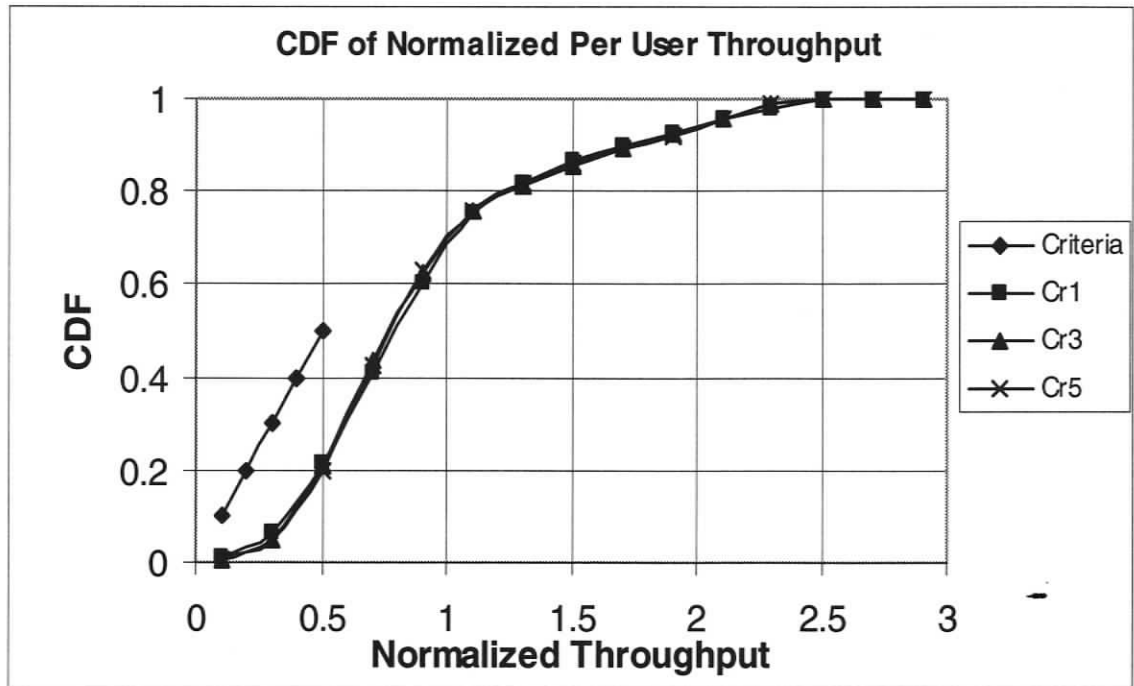


Figure 26 CDFs of Normalized Per User Throughput of 1xEV-DV (1 carrier), MC-DV with 3 Carriers and MC-DV with 5 Carriers

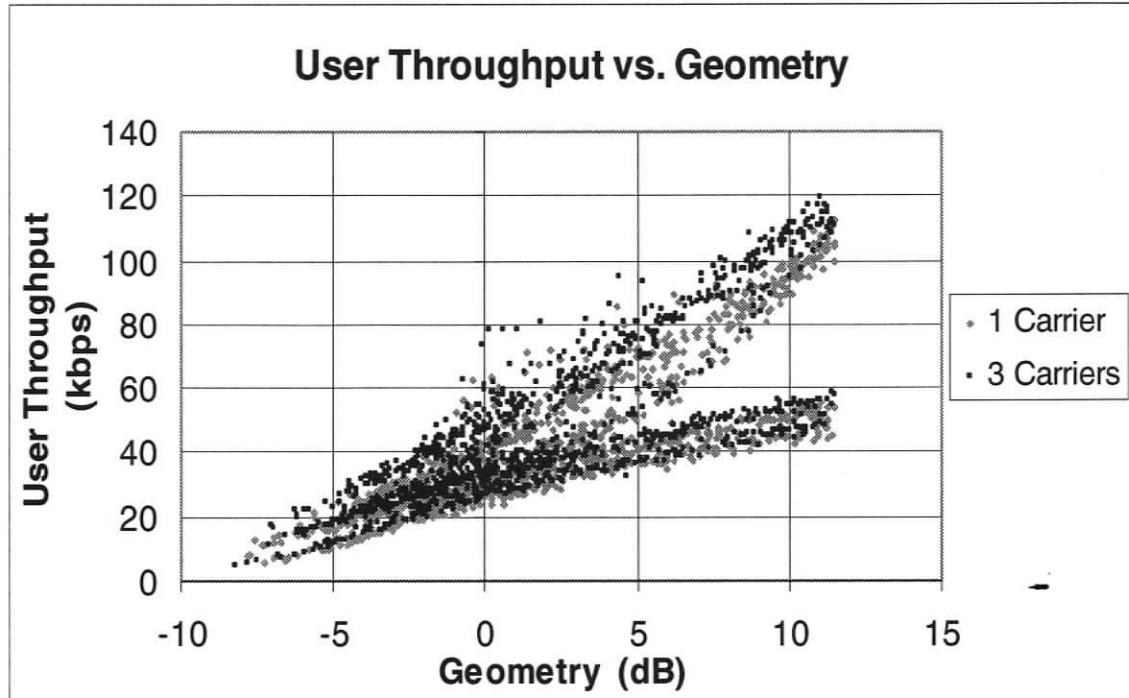


Figure 27 Scatter Plot of User Throughput Normalized to Per-Carrier versus Geometry for 1xEV-DV (1 Carrier) and MC-DV (3 Carriers)

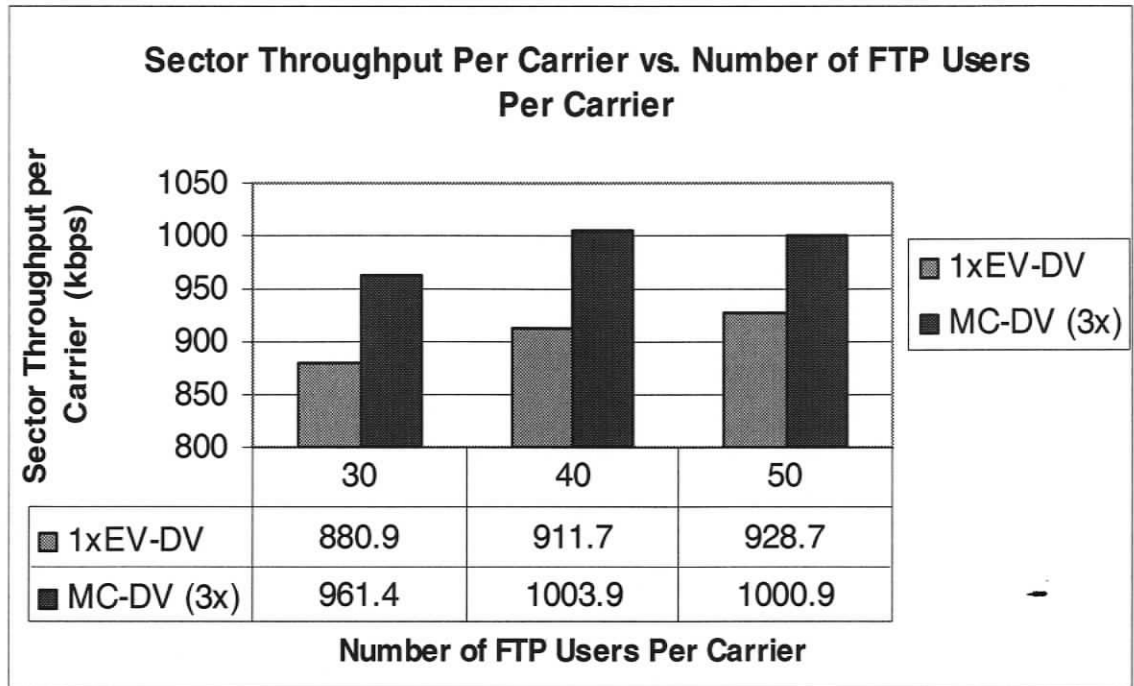


Figure 28 Sector Throughput Per Carrier versus Number of FTP Users for 1xEV-DV and MC-DV with 3 Carriers

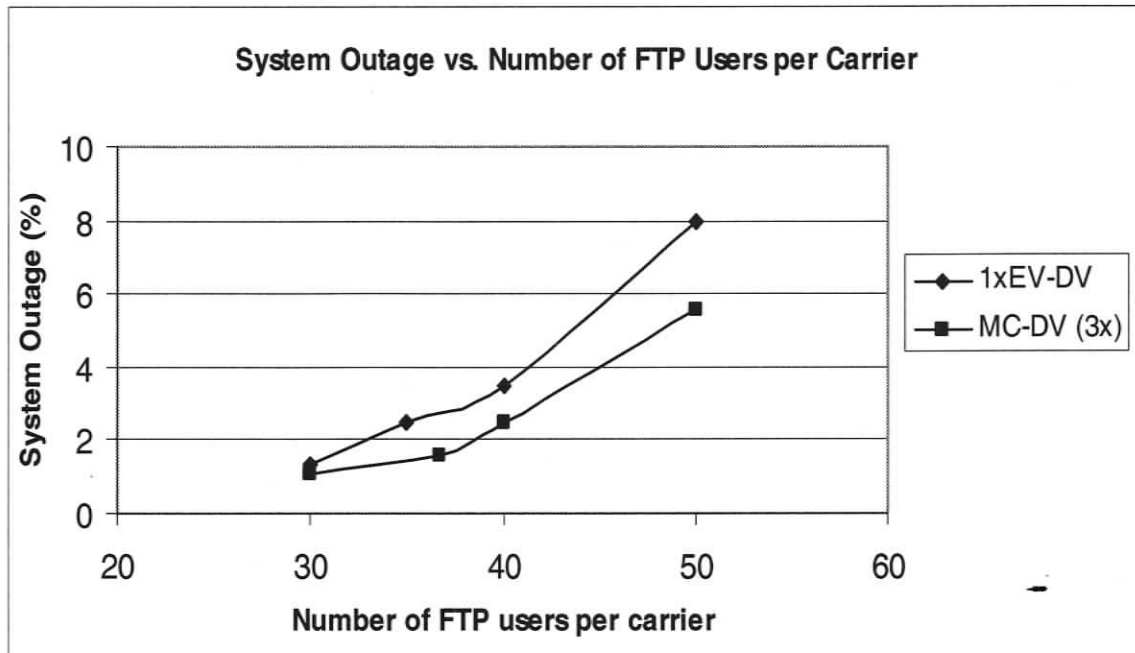


Figure 29 System Outage versus Number of FTP Users Per Carrier for 1xEV-DV and MC-DV with 3 Carriers

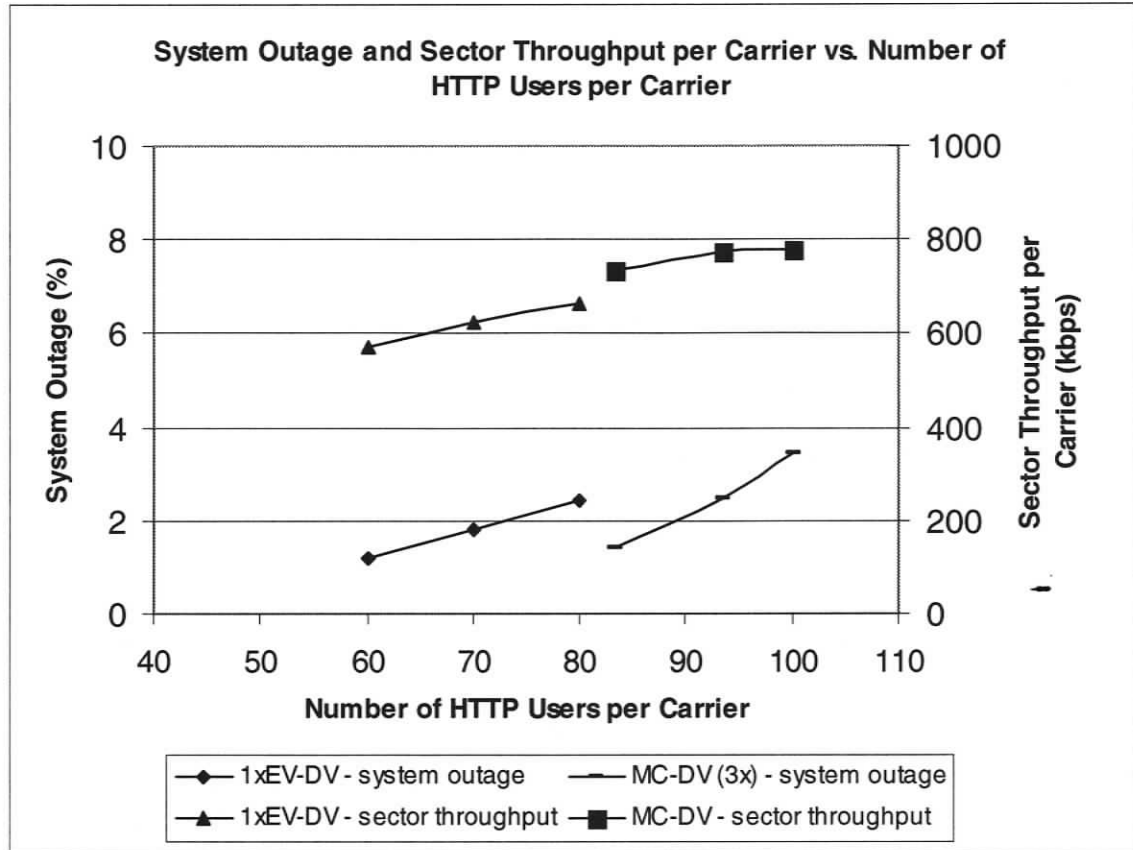


Figure 30 System Outage and Sector Throughput Per Carrier versus Number of HTTP Users for 1xEV-DV and MC-DV with 3 Carriers

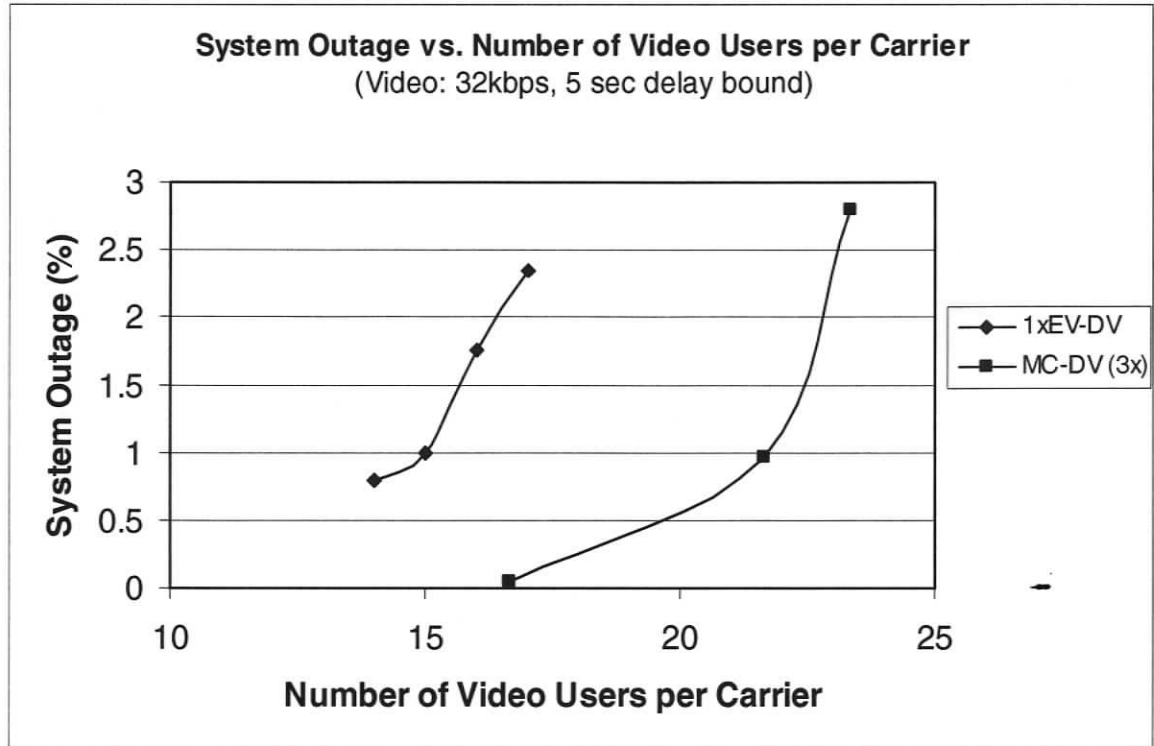


Figure 31 System Outage versus Number of Video Users (Streaming Video with Source Rate of 32kbps and Delay Bound Criterion of 5 Seconds) for 1xEV-DV and MC-DV with 3 Carriers

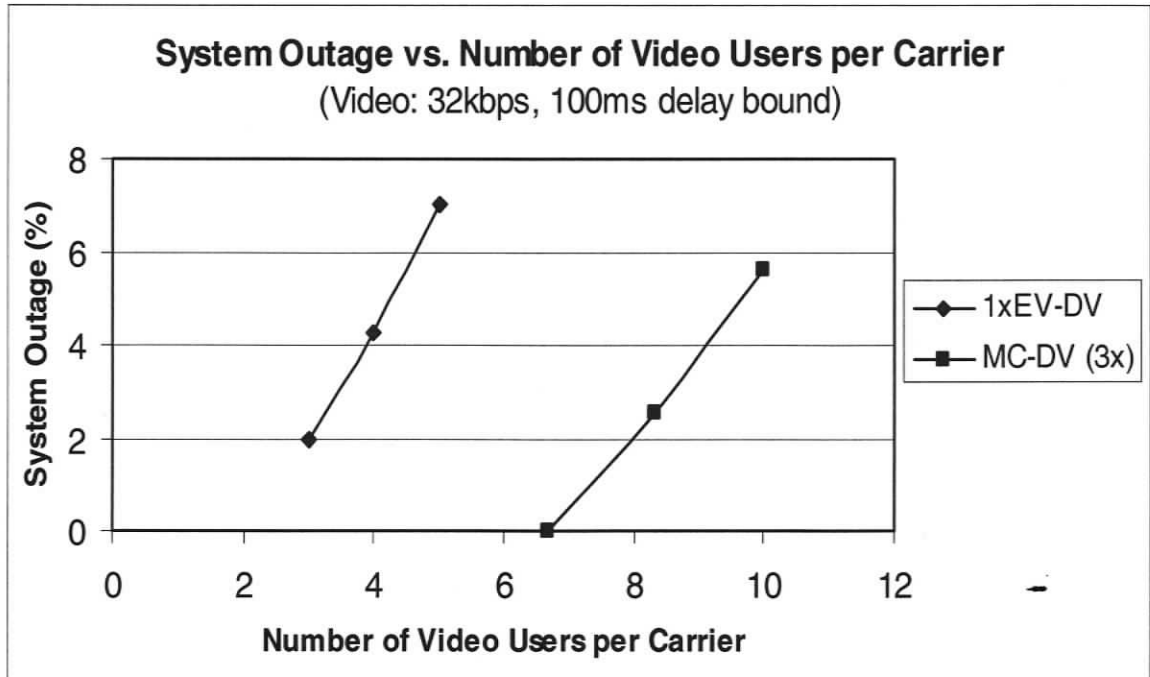


Figure 32 System Outage versus Number of Video Users (Real-Time Interactive Video with Source Rate of 32kbps and Delay Bound Criterion of 100ms) for 1xEV-DV and MC-DV with 3 Carriers

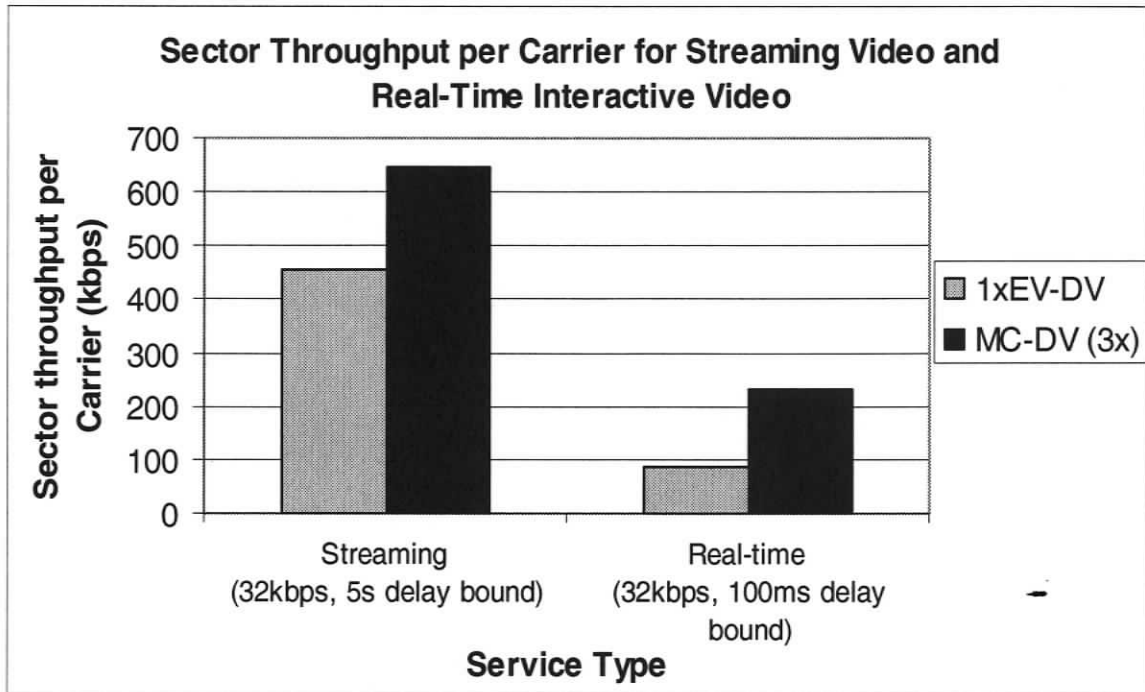


Figure 33 Sector Throughput Per Carrier for Streaming Video and Real-Time Interactive Video for 1xEV-DV and MC-DV with 3 Carriers

Chapter 6

Dynamic Load Balancing Across Physical Layer

Modes

In this chapter, we investigate the load balancing aspect of the B3G multi-mode wireless access system. Load balancing is the mechanism to balance the traffic load across multiple network entities. Its functions are 1) to reduce the congestion probability of a particular network entity and the resulting QoS degradation or call drop; 2) to maximize the overall system capacity while meeting users' SLA and QoS requirements. In the wireless domain, the multiple network entities are typically neighbouring or overlapping sectors/cells as previously investigated [33]-[35]. These studies performed load control through call admission policy, proactive handover, and dynamic allocation of frequency channels to different sectors/cells.

Here, we investigate a different aspect of load balancing where the network entities of interest are the different PL modes in the proposed B3G multi-mode wireless access system. As described in details in Chapter 4, the heterogeneous PL modes can be co-located in the same sector/cell, where each PL mode is optimized to support specific mobility conditions, service types, and backward compatibility requirements. The different PL modes can be multiplexed in a TDM fashion if they share the same spectrum, or in CDM fashion if they are based on CDMA and share the same frequency spectrum, or in a FDM fashion if they occupy separate frequency bands. We investigate load balancing schemes for TDM, FDM and CDM scenarios.

6.1 Load Balancing Schemes

Load balancing can be classified into two general categories: 1) methods to allocate users to different network entities based on loading conditions; 2) methods to allocate different amounts of channel resources to different network entities based on loading conditions. The user allocation or channel resource allocation described above can be semi-static or dynamic. In the first category, semi-static user allocation is the case where a user is allocated to a particular PL mode during call admission based on the loading condition on each PL mode, as well as whether the PL mode can support the terminal types, service types and mobility condition required by the user. For dynamic user allocation, a user can switch among the supported PL modes on a frame-by-frame basis during an active call session. This category requires user terminals to be multi-mode.

For the second category, semi-static resource allocation is the case where resource allocated to each PL mode changes slowly, in the order of minutes or hours based on the large scale loading condition of the PL modes. For dynamic resource allocation, resource allocated to each PL mode can change as frequent as per frame basis. This category does not require user terminals to be multi-mode.

Here, we focus on the second category of load balancing methods, since the PL mode assigned to a user at a certain period of time is typically fixed as a result of terminal types, service types and mobility conditions required by the user. Therefore it is better to adapt the channel resources allocated to different PL modes based on the loading condition of each PL mode. In addition, this category does not require user terminals to be multi-mode. However, it is important to note that the load balancing algorithms

proposed here can also generally apply to the first category of load balancing methods. We investigate both semi-static and dynamic load balancing schemes.

6.1.1 TDM Multiplexing of PL modes

In this scenario, different PL modes occupy different TDM time slots. Load balancing between the PL modes can be performed either dynamically or semi-statically as described below.

6.1.1.1 Dynamic Load Balancing

We propose an integrated load balancing and scheduling (ILBS) scheme to dynamically assign TDM resources to different PL modes on a slot-by-slot basis. In this scheme, load balancing and fast scheduling are simultaneously performed across the different PL modes. The goals are to perform per time-slot and per packet opportunistic scheduling and load balancing to maximize the system capacity while meeting the traffic QoS. Here is a step-by step description of the proposed algorithm.

At time slot n :

Step 1: Assign priorities to all users of different PL modes based on a certain scheduling priority equation. The same scheduling priority equation as in normal scheduling operation can be used. Equation (4) in Chapter 5 can be generalized as follows. The priority of packet j of user i at time slot n and PL mode m can be expressed as

$$(17) \quad p_{ij}^{(m)}(n) = f(r_i^{(m)}(n), \bar{r}_i(n), d_{ij}(n), \min_{ij} r_i),$$

where $r_i^{(m)}(n)$ is the instantaneous data rate supported by user i at time slot n and PL mode m based on the instantaneous channel condition experienced by the user in PL

mode m . The scheduler computes the priority of all the users' packets in the layer 2 buffer for the corresponding PL modes supported by the users. The set of PL modes supported by a user during an active call session is configured during call setup based on the user terminal capability and the user's requested service types and mobility conditions.

Step 2: Select the user/packet with the highest priority in the list for transmission as follows,

$$(18) \quad [i^1, j^1, m^1] = \arg \max_{i, j, m} \{p_{ij}^{(m)}(n)\}.$$

The PL mode supported by the selected user/packet, m^1 , will be the PL mode used in this time slot.

Step 3: If there are sufficient resources remaining in the time slot to multiplex another user/packet, select the next highest priority user/packet in the list that supports the same PL mode as the user chosen in Step 2, i.e.,

$$(19) \quad [i^2, j^2, m^2 = m^1] = \arg \max_{i \neq i^1, j \neq j^1, m = m^1} \{p_{ij}^{(m)}(n)\}.$$

Step 4: Repeat Step 3 until there are no resources remaining in the time slot or all users/packets have been scheduled.

The above algorithm ensures that the PL mode selection at each time slot is based on the highest priority packet in the layer 2 buffer. With the proper selection of the priority equation, such as the proportional fairness scheduler or the fairness plus delay scheduler described in Chapter 5, the overall system capacity across PL modes can be optimized while meeting users' QoS and SLA.

6.1.1.2 Semi-static Load Balancing

In this method, the TDM resources or time slots are semi-statically partitioned among the PL modes based on the large-scale loading condition of each PL mode. The partitioning changes slowly to adapt to user arrivals/departures to/from the PL modes. For example, a partition ratio of 1:1 across two PL modes means that each PL mode is allocated 50% of the time slots. Load balancing between PL modes is decoupled from the dynamic layer 2 packet scheduling.

At a time slot n , the PL mode is predetermined as $m(n)$ based on the semi-static load balancing decision. For example, in the case of two PL modes with 1:1 resource partition, odd numbered time slots are for PL mode 1 and even numbered time slots are for PL mode 2. The layer 2 scheduler only schedules users' packets that correspond to PL mode $m(n)$ for transmission based on the users' packets prioritization process as described earlier. The semi-static load balancing scheme therefore does not adapt dynamically to short-term variation of users' packets priority due to channel variation, random traffic arrival, queuing delay, etc.

6.1.2 FDM Multiplexing of PL modes

In this scenario, different PL modes occupy different non-overlapping frequency bands. Load balancing between the PL modes can be performed either dynamically or semi-statically as described below.

6.1.2.1 Dynamic Load Balancing

The spectrum allocated to each PL mode is dynamically changed each frame. This scheme requires terminals with wideband receivers capable of receiving the radio frequency signal from the entire spectrum across all the PL modes of interest. The actual

baseband decoding of a portion of the spectrum corresponding to a particular PL mode is based on the signaling received from the dedicated signaling channel in each PL mode. One example of such a dynamic spectrum sharing scheme is the overlay of MC-DO, an expansion of 1xEV-DO [1] based on the Multi-Carrier DS-SS-CDMA concept described in Chapter 5, and OFDMA. In this example, the granularity of spectrum allocation is the bandwidth of each 1xEV-DO carrier, i.e. 1.25MHz. The MC-DO mode can occupy a variable number of 1.25MHz carriers. The OFDMA mode has a fixed size FFT based on the full spectrum but the number of useful sub-carriers varies based on the dynamic spectrum allocation. Figure 34 illustrates the dynamic spectrum sharing between MC-DO and OFDMA. The spectrum allocated to MC-DO and OFDMA can be contiguous or non-contiguous dependent on the scheduling and load balancing algorithms. If frequency selective scheduling is performed for MC-DO and/or OFDMA, it may result in non-contiguous spectrum allocation because a specific MC-DO carrier or OFDMA sub-band may be chosen to transmit a particular user packet. This is further illustrated in later section. In Figure 34, in time slots $n+2$ and $n+3$, the entire spectrum is allocated to MC-DO and OFDMA respectively. Therefore, in these two time slots, the operation is the same as TDM of PL modes described in Section 6.1.1.

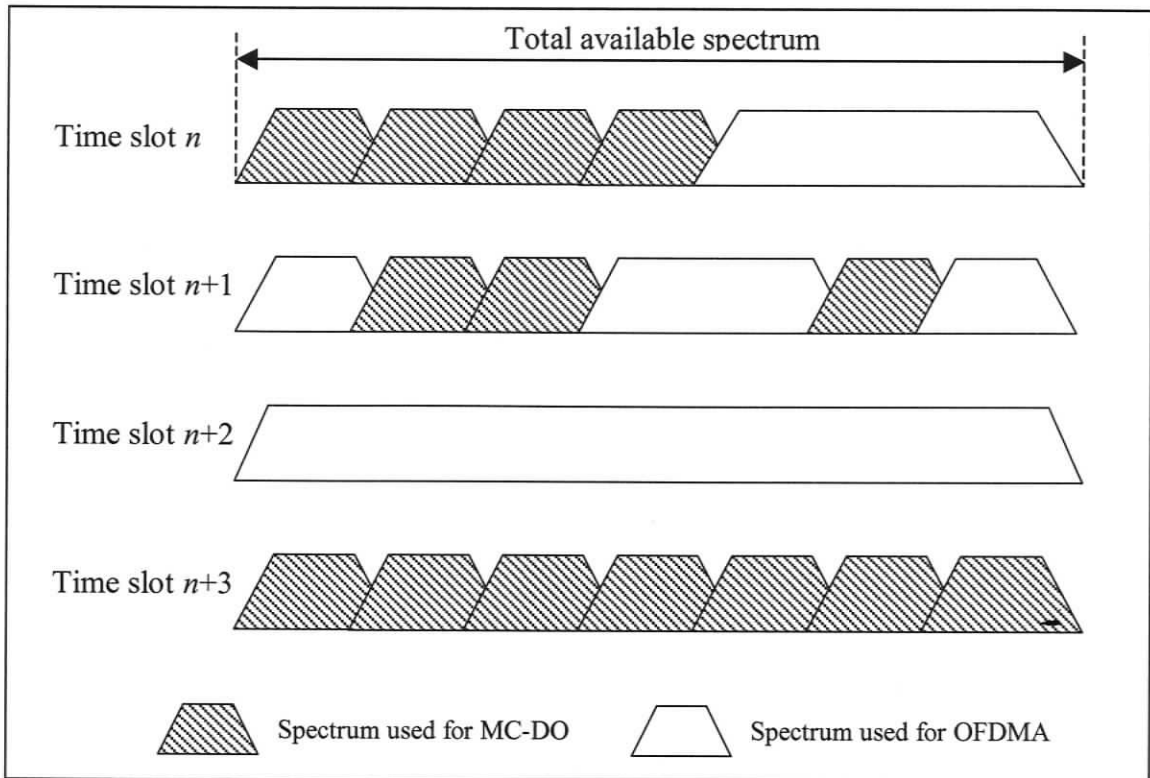


Figure 34 Example of Dynamic Spectrum Sharing between PL Modes: MC-DO and OFDMA.

Figure 35 illustrates the dynamic spectrum sharing between different OFDMA modes for fixed users, medium speed users and high speed users, and for unicast and multicast services. The spectrum or OFDM sub-carriers in this case are dynamically assigned to different OFDMA modes from one time slot to another. To facilitate efficient multiplexing of fixed and mobile users, one approach is to define a fixed FFT size over the entire spectrum, thus the same basic sub-carrier spacing for all OFDMA modes. The basic sub-carrier spacing is defined based on the requirement to support fixed users. For mobile users, null sub-carriers are introduced as shown in Figure 35 to increase the effective sub-carrier spacing to integer number of the basic sub-carrier spacing. Unicast and multicast services are TDM as shown in Figure 35. This is because unicast and multicast transmissions require different cyclic prefix length, thus different OFDM

symbol durations. The specific time slots used for multicast transmission is broadcast through the common signalling channel to all users in the system.

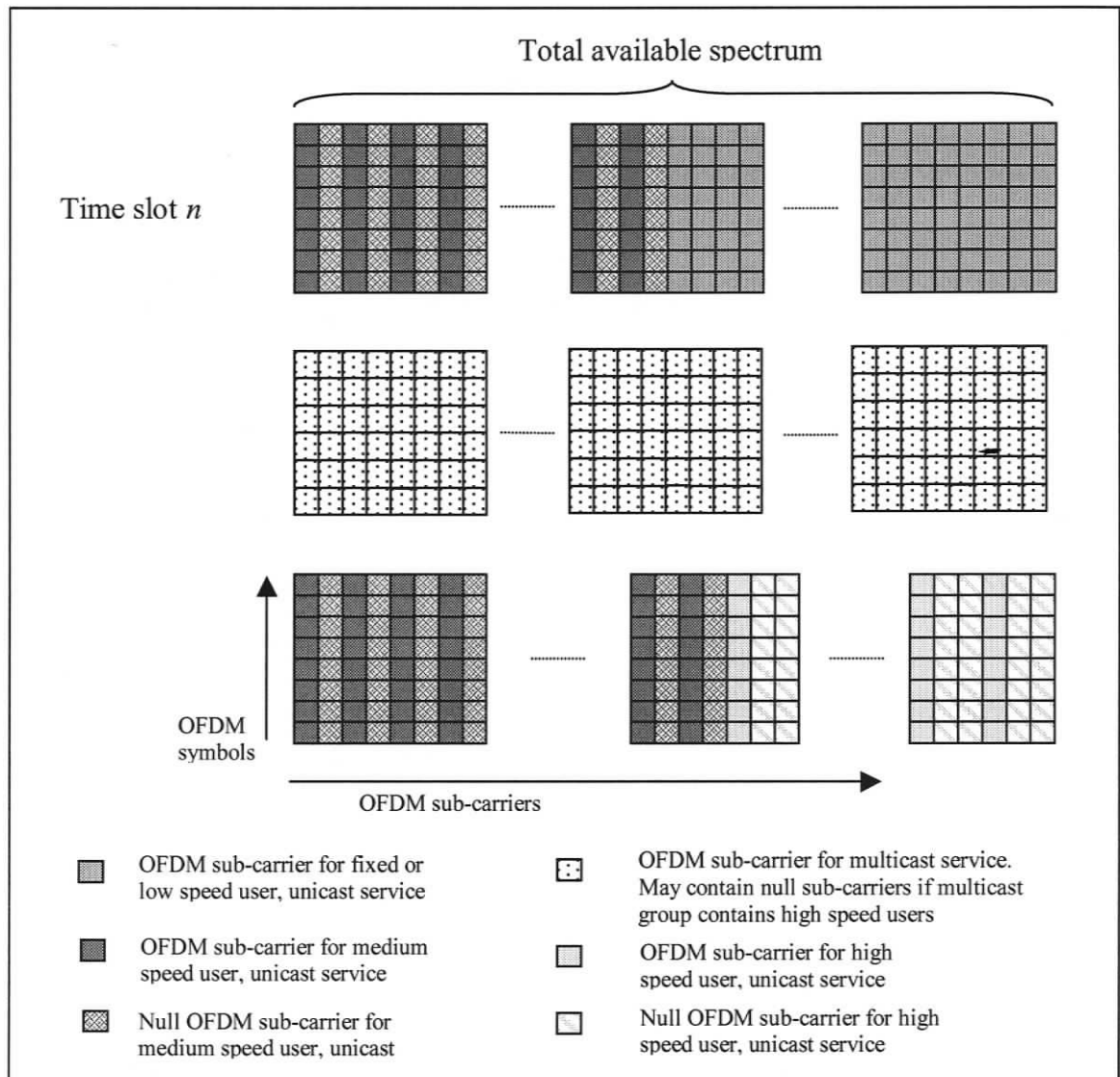


Figure 35 Example of Dynamic Spectrum Sharing between PL Modes: Different OFDMA modes for fixed, medium speed and high speed users, unicast and multicast services

The proposed operation of the ILBS scheme that combines frequency selective scheduling is described as follows.

At time slot n :

Step 1: Assign priorities to all users/packets of different PL modes as in Step 1 in Section 6.1.1.1 and establish a priority list, $\{p_{ij}^{(m)}(n)\}$. Set iteration $l = 0$.

If frequency selective scheduling operation is employed to prioritize user packets across different segments of the spectrum, each 1.25MHz carrier of MC-DO or each sub-band of OFDMA is considered a unique PL mode with a different value of m , where $p_{ij}^{(m)}(n)$ is calculated. For a user packet $[i, j]$, the difference between $p_{ij}^{(m)}(n)$ of different spectrum segments is due to the difference in channel conditions or CIR experienced by user i in different spectrum segments. As an illustration using the example shown in Figure 34, the total available spectrum is divided into seven MC-DO carriers. Therefore, an MC-DO user can have up to seven corresponding PL modes and associated values of $p_{ij}^{(m)}(n)$. Similarly, if the total available spectrum is divided into seven OFDMA sub-bands, an OFDM user can have up to seven corresponding PL modes and associated values of $p_{ij}^{(m)}(n)$. An MC-DO user may support fewer MC-DO carriers than the spectrum allows for reasons such as terminal capability or CIR feedback overhead consideration. Same applies to an OFDMA user where the number of sub-bands supported may be fewer than the spectrum allows.

If frequency selective scheduling is not employed for a particular MC-DO user or OFDMA user due to high mobile speed that renders accurate tracking of channel condition impossible or due to CIR feedback overhead consideration, the user will have one PL mode or associated $p_{ij}^{(m)}(n)$ that corresponds to the wideband channel condition or CIR averaged across the whole spectrum. The user packet can be transmitted on any of the frequency bands.

Step 2: Increment l by 1. This is the l th iteration. Select the user packet with the highest priority in the priority list, i.e. $[i^*, j^*, m^*]$, using equation (18). Calculate the spectrum utilization of the selected user packet.

The spectrum utilization is defined in terms of the frequency band (i.e. the MC-DO carrier, the OFDMA sub-band, or wideband) the PL mode m^* resides, and the amount of resource utilized in the corresponding frequency band. For the case of CDMA based PL mode, the resource is defined in terms of Walsh codes, BS transmit power, multi-user packet (for 1xEV-DO or MC-DO only) and carriers. For the case of OFDMA, the resource is defined in terms of OFDMA sub-carriers and symbols.

For the case of frequency selective scheduling, the frequency band of PL mode m^* is expressed as $B(m^*)$. $B(m^*)$ corresponds to a specific OFDMA sub-band or MC-DO carrier, with integer value ranges from 1 to N_{sb} , where N_{sb} is the total number of sub-bands in the allocated spectrum. The resource utilization of user packet $[i^*, j^*, m^*]$ in $B(m^*)$ is expressed as $R(i^*, j^*, m^*)$, where $R(i^*, j^*, m^*)$ is a function of packet size, the selected modulation and coding scheme and the corresponding required received SNR of user i^* in PL mode m^* .

Step 3: Determine if the packet $[i^*, j^*, m^*]$ can be transmitted in this time slot.

1) **Step 3a:** For frequency selective scheduling, determine if the packet $[i^*, j^*, m^*]$ can be transmitted in frequency band $B(m^*)$. The packet $[i^*, j^*, m^*]$ can only be transmitted on frequency band $B(m^*)$ if the PL mode m^* is compatible with the PL mode of the packets in the current transmission list of frequency band $B(m^*)$. For example, if the existing packets in the transmission list of a particular 1.25MHz carrier are of MC-DO mode, then

a packet of OFDMA mode cannot be transmitted on the same 1.25MHz carrier. On the other hand, a packet of 1xEV-DO mode can be transmitted on the same 1.25MHz carrier.

If the PL mode of packet $[i^*, j^*, m^*]$ is incompatible with the PL mode of the packets in the current transmission list of frequency band $B(m^*)$, discard the entry $p_{i^* j^*}^{(m^*)}(n)$ from the priority list. Otherwise, calculate the new cumulative resource utilization in frequency band $B(m^*)$, as shown in equation (20) below.

$$(20) \quad U_l(B(m^*)) = U_{l-1}(B(m^*)) + R(i^*, j^*, m^*),$$

where $U_l(B(m^*))$ is the updated cumulative resource utilization in frequency band $B(m^*)$ in iteration l after accounting for the resource utilized by the newly selected packet $[i^*, j^*, m^*]$, $U_{l-1}(B(m^*))$ is the cumulative resource utilization in frequency band $B(m^*)$ after $(l-1)$ iterations. Note that $U_0(.) = 0$ for all sub-bands. If $U_l(B(m^*)) \leq U_{avail}(B(m^*))$, where $U_{avail}(B(m^*))$ is the amount of resource available in frequency band $B(m^*)$, move the packet from the priority list to the transmission list of $B(m^*)$. Otherwise, discard the entry $p_{i^* j^*}^{(m^*)}(n)$ from the priority list.

2) Step 3b: For frequency non-selective scheduling, the packet can be transmitted in any part of the spectrum that has compatible PL mode. Therefore, we calculate the cumulative resource utilization of compatible PL mode category as follows:

$$(21) \quad V_l(PL(m^*)) = V_{l-1}(PL(m^*)) + R(i^*, j^*, m^*),$$

where $V_l(PL(m^*))$ is the cumulative resource utilization of PL mode category $PL(m^*)$ after l th iteration. $PL(m^*)$ is a function that categorizes the PL mode m^* into one of the

PL mode categories supported by the system. For example, 1xEV-DO and MC-DO are compatible and in the same PL mode category 1, where $PL(m^*) = 1$. OFDMA is in a different PL mode category 2, where $PL(m^*) = 2$.

We determine if

$$(22) \quad V_l(PL(m^*)) \leq V_{avail}(PL(m^*)),$$

where $V_{avail}(PL(m^*))$ is the amount of resource available to $PL(m^*)$. If the inequality in equation (22) holds, move the packet from the priority list to the transmission list of PL mode category $PL(m^*)$. Otherwise, discard the entry $p_{i^* j^*}^{(m^*)}(n)$ from the priority list.

Update the resource availability of each PL mode category, pl , as follows:

$$(23) \quad V_{avail}(pl) = f\left(S_{total} - \sum_{q, q \neq pl} S(V_l(q))\right),$$

where S_{total} is the total available spectrum in units of frequency sub-channels, a minimum frequency division multiplexing bandwidth among different PL mode categories; $S(V_l(q))$ is the required spectrum in units of frequency sub-channels for the cumulative resource utilization of PL mode category q , $V_l(q)$. The function $f(x)$ maps the available spectrum x to the actual physical resource, i.e. Walsh codes, BS transmit power, carriers for the case of CDMA such as 1xEV-DO and MC-DO; and OFDM sub-carriers and symbols for the case of OFDMA.

Step 4: Repeat Steps 2 and 3 until there is no resource remains in any of the frequency bands or PL mode categories, or there is no packets left in the priority list.

The benefit of the dynamic FDM load-balancing scheme described above over the dynamic TDM load balancing in Section 6.1.1.1 is that the resource allocation to each PL mode is much more flexible since there is no restriction to schedule users with the same PL mode within each time slot. There is however still some restriction in the case of frequency selective scheduling as described in step 3a since the PL modes allocated for a particular sub-band have to be compatible to one another. For the case of frequency non-selective scheduling, there is no such restriction except for the bandwidth granularity of the frequency multiplexing, e.g. 1.25MHz for the case of FDM between 1xEV-DO/MC-DO and OFDMA.

6.1.2.2 Semi-static Load Balancing

In this method, the spectrum resource is semi-statically partitioned among the PL modes based on the large-scale loading condition of each PL mode. The partitioning changes slowly to adapt to user arrivals/departures to/from the PL modes. For example, a partition ratio of 1:1 across two PL modes means that each PL mode is allocated 50% of the spectrum. Load balancing between PL modes is decoupled from the dynamic layer 2 packet scheduling. The semi-static load balancing scheme therefore does not adapt dynamically to short-term variation of user packets priority due to channel variation, random traffic arrival, queuing delay, etc. The performance of this scheme is the same as that of Section 6.1.1.2.

6.1.3 CDM Multiplexing of PL modes

This scenario applies to PL modes that are based on CDMA and have compatible multiplexing techniques using spreading codes and transmit power control. One example is the overlay of 1xRTT, 1xEV-DV and MC-DV in the forward link, where the dedicated

traffic channels (e.g. Forward Fundamental Channel, F-FCH; and Forward Supplemental Channel, F-SCH) of 1xRTT, F-PDCH of 1xEV-DV and F-PDCH of MC-DV are multiplexed using different Walsh codes and consumes different portions of the BS transmit power.

In the case of dynamic load balancing, the resource pools of Walsh codes and the BS transmit power can be dynamically shared among the different PL modes. For the dynamic load balancing of 1xEV-DV and MC-DV F-PDCH transmission, the global scheduling and prioritization of 1xEV-DV and MC-DV packets as described in Section 5.1.4 is used. For the dynamic load balancing of 1xRTT dedicated traffic channels and 1xEV-DV/MC-DV F-PDCH, the followings are performed.

At time slot n :

Step 1: Calculate the required transmit power of all the active 1xRTT dedicated traffic channels, based on the power control bits (PCBs) feedback from the MSs.

Step 2: Compute the BS transmit power for F-PDCH on each carrier.

The 1xRTT dedicated traffic channels are given higher priority than the F-PDCH in terms of BS transmit power and Walsh code usage since they are circuit-oriented channels and are power-controlled to ensure acceptable received SNR. Therefore, the leftover BS transmit power and Walsh codes are allocated to the F-PDCH. The leftover BS transmit power is computed for each carrier as follows:

$$(24) \quad P_{pdch}^{(k)} = P_{\max} - \sum_l P_{dch}^{(k)}(l),$$

where $P_{pdch}^{(k)}$ is the leftover BS' transmit power in carrier k for F-PDCH, $P_{dch}^{(k)}(l)$ is the BS transmit power consumed by the l th dedicated channel in carrier k , and P_{\max} is the

maximum BS' transmit power allocated for traffic channels, i.e. excluding the power allocated to common or overhead channels such as paging channel, common control channel etc.

Step 3: Select the 1xEV-DV or MC-DV packets for transmission using the algorithms described in Section 5.1.4. One enhancement is $r_i^{(k)}(n)$ as shown in equation (4) is now computed based on both the instantaneous CIR in carrier k , as well as $P_{pdch}^{(k)}$.

The above dynamic load balancing scheme allows efficient usage of BS transmit power and Walsh codes among the power controlled dedicated traffic channel and the rate controlled packet data channel. This is illustrated in Figure 36. Due to the varying channel condition on each dedicated channel, i.e. F-FCH for voice user in this case, there is a large fluctuation on the total power consumed by the dedicated channels. Even though additional dedicated channels can be allocated to further consume the remaining BS power, this will cause an increase in power outage rate when the total instantaneous power consumed by the dedicated channels exceeds the maximum available BS transmit power. The increase in power outage rate will cause an undesirable increase in voice or circuit data service outage. The BS transmit power therefore cannot be fully utilized by the power controlled dedicated channels. Rate controlled F-PDCH can be used to water-fill the remaining unused BS transmit power to improve multiplexing efficiency of the system. The water-filling approach does come with a price of increased intra-cell interference due to multi-user interference in a dispersive channel environment and increased inter-cell interference due to overall higher BS transmit power. This in turn increases the voice or circuit data service outage. On the whole, as will be shown in

Section 6.2.2, the dynamic multiplexing of power-controlled dedicated channels and rate-controlled packet data channels provides overall system performance gain.

6.2 Performance Evaluation

In this section, we evaluate the performance of different load balancing schemes described in Section 6.1.

6.2.1 Performance of Load Balancing with TDM and FDM of PL Modes

We compare the performance of the different TDM and FDM load balancing schemes described in Sections 6.1.1 and 6.1.2, i.e.

- Case 1:** semi-static load balancing in Sections 6.1.1.2 and 6.1.2.2;
- Case 2:** dynamic load balancing for TDM scenario in Section 6.1.1.1;
- Case 3:** dynamic load balancing for FDM scenario in Section 6.1.2.1.

A full system-level simulation is performed based on the methodology described in Chapter 3: 19 tri-sector cells, modified Hata path loss model, log-normal shadowing, fast fading with mixed mobile speeds ranging from stationary to 120 km/h. Both the physical layer and the medium access control layer including scheduling and load balancing operations are modeled. To evaluate and compare the performance of different load balancing schemes, we use VoIP as the user application. The VoIP traffic model is as described in Chapter 3. The performance metric used is the system outage versus the number of simultaneous VoIP users per sector, where the system outage definition is given in Chapter 3.

We use 1xEV-DO as the baseline system model to compare the performance of the three cases listed above. Users are partitioned into two groups to represent two PL modes,

with 50% of the users in each group. For Case 1, we statically allocate 50% of the time slots to each PL mode. For Case 2, ILBS is performed on all users belonging to the two PL modes. The PL mode of each time slot is determined by the PL mode of the highest priority user. Once the PL mode is determined, only users belonging to the same PL mode can be multiplexed onto the same time slot. For Case 3, the model is similar to Case 2, except that there is no restriction to multiplexing users belonging to the same PL mode onto each time slot. This is used to emulate the case where users belonging to different PL modes are flexibly sharing the spectrum within a time slot which is representative of the dynamic FDM load balancing with frequency non-selective scheduling.

The performance results in terms of system outage versus total number of VoIP users for the three cases are shown in Figure 37. The performance results for no load balancing cases are also shown in Figure 37. For the no load balancing case, we evaluate two resource partition ratios between the two PL modes, those are, 2:3 and 1:4. Our results show that semi-static or dynamic load balancing provides a significant performance gain or reduction in system outage compared to no load balancing. In addition, dynamic load balancing with the proposed ILBS algorithms provide a significant performance gain compared to semi-static load balancing schemes. The reduction in system outage is in the order of ~5 times. Note that a reduction in system outage corresponds to an overall increase in system capacity given a specific target system outage. For example, at 2% system outage criterion, the capacity gain of dynamic load balancing over the semi-static load balancing is 12.5% and 25% respectively for TDM and FDM cases. As expected, we

also observe that applying dynamic load-balancing to FDM has a small advantage over that with TDM.

We further evaluate the performance for a system consists of four PL modes, where each PL mode contains 25% of the users in the system. The modeling of the 3 cases is similar to what described earlier for two PL modes. Our results in Figure 38 show that the performance difference between dynamic and semi-static load balancing widens compared to the case of two PL modes. The reduction in system outage is 10 to 50 times. The capacity gain of dynamic load balancing using FDM over the semi-static load balancing at 2% system outage criterion is about 2 times. As the number of PL modes increases, the amount of resources available to each PL mode reduces, thus the need for dynamic load balancing becomes increasingly more important. Similarly, the performance difference between FDM and TDM dynamic load balancing also widens compared to the case of two PL modes. Therefore, the larger the number of PL modes sharing the same overall spectrum, the more flexible the loading balancing and multiplexing of PL modes should be to maximize system capacity while ensuring users' QoS and SLA are met. In summary, the dynamic load balancing using FDM is recommended as it has overall superior performance and its performance is insensitive to the number of PL modes. The performance benefit is however at the expense of implementation complexity of the layer 2 scheduler.

6.2.2 Performance of Load Balancing with CDM of PL Modes

We first investigate the multiplexing of power-controlled dedicated channels and rate-controlled packet data channel on one carrier. The system model used is 1xEV-DV and 1xRTT overlay on a 1.25MHz carrier. A full system-level simulation is conducted. We

model the 1xRTT dedicated channels as F-FCH supporting circuit switched voice service, i.e. EVRC vocoder using RC3 [9][10][11]. The 1xEV-DV F-PDCH is used to carry either full buffer data traffic or FTP traffic.

We first obtain the voice capacity of the system by loading only voice users in the system. For the system outage criterion defined in [27], the voice capacity is 30 simultaneous voice users. We then perform the simulation with mixed voice and data traffic, by loading the system with different percentage of voice loading and a fixed number of full buffer or FTP data users. Figure 39 shows the aggregate data throughput per sector versus percentage of voice loading. There are 20 full buffer data users per sector in the simulation. The percentage of voice loading is defined as the percentage of system capacity occupied by voice users. For example, 100% voice loading is referring to 30 voice users in the system, i.e. the previously obtained voice capacity. 50% voice loading is referring to 15 voice users in the system. From Figure 39, we can see that the data sector throughput degrades gracefully with the increase in voice loading. Note that at 100% voice loading, we do not include any data traffic in the system. Figure 40 shows the voice outage with respect to different voice loading in the same mixed voice and full buffer data scenario. We can see that with the introduction of data traffic to water-fill the BS transmit power, the voice outage increases beyond the value observed for voice-only traffic case. The increased voice outage is due to the increase intra-cell and inter-cell interference as previously discussed. To maintain the voice outage at 3% or below as defined in [27], the recommended voice loading for a mixed 1xRTT voice and 1xEV-DV data scenario is up to 50%.

Similarly, we evaluate the performance for mixed voice and FTP data scenario as shown in Figure 41. As in the full buffer data case, the FTP data sector throughput degrades gracefully with the increase in voice loading. The FTP user outage, as defined in [27] and Chapter 3, increases gradually with voice loading up to ~50% voice loading. With voice loading larger than 50%, the FTP user outage increases exponentially. This is due to the small amount and varying resource for F-PDCH which causes frequent TCP timeout and slow ramp-up.

Overall, based on our performance studies, we recommend up to 50% voice loading for a mixed circuit voice and packet data scenario to ensure acceptable voice outage and data outage. We propose the following load balancing scheme for mixed voice and data deployment across multiple carriers:

- During call admission, evenly load each carrier with voice users up to a specific nominal voice loading threshold.
- When the voice loading of all carriers has reached the specific threshold, and if voice service has a higher admission priority, then it is desirable to transition one of the carriers into a voice-only carrier where additional voice calls will be admitted into this carrier.
- As the overall voice loading of the system decreases below the per-carrier nominal voice loading threshold, the designated voice-only carrier will switch back to a mixed voice/data carrier.

We compare the performance of the proposed load balancing scheme with a conventional scheme that segregates voice and packet data into separate carriers in Table 11. Our proposed scheme outperforms the conventional scheme by 12% - 24% for the

scenarios examined. Note that the proposed scheme can generally apply to other mixed dedicated power-controlled traffic channels and rate-controlled packet data channels scenarios.

Table 11 Performance of the Proposed Dynamic Load Balancing Scheme for Mixed Circuit Voice and Packet Data (Full Buffer Traffic) Across 3 Carriers

	Conventional scheme	Proposed scheme
Scenario 1: 30 voice users across 3 carriers	Dedicate one carrier for voice and two other carriers for data. Overall data throughput = 2 x 937kbps = 1.87Mbps	Evenly load each carrier with 10 voice users. Water-fill the left-over resource in each carrier with data. Overall data throughput = 3 x 700kbps = 2.1Mbps Gain of the proposed scheme over the conventional scheme = 12%
Scenario 2: 60 voice users across 3 carriers	Dedicate two carriers for voice and one other carrier for data. Overall data throughput = 937kbps	Dedicate one carrier for voice (30 users). Even load the remaining two carriers with 15 users each. Water-fill the left-over resource in the 2 nd and 3 rd carriers with data. Overall data throughput = 2 x 580kbps = 1.16Mbps Gain of the proposed scheme over the conventional scheme = 24%

We further investigate the mixed circuit voice and packet data scenario for 1xRTT and MC-DV overlay where the packet data traffic is dynamically scheduled across the MC-DV carriers using the algorithms described in Section 6.1.3. Figure 42 and Figure 43 show the 1xEV-DV and MC-DV data performance for mixed voice and packet data

scenario with 17% (i.e. 5 voice users per carrier) and 47% (i.e. 14 voice users per carrier) voice loading respectively. At 17% voice loading and 2% FTP user outage, the FTP user capacity per carrier is 33 users and 23 users respectively for MC-DV and 1xEV-DV. The MC-DV user capacity gain over 1xEV-DV is therefore 43%. The corresponding MC-DV data sector throughput gain over 1xEV-DV is 14%. At 47% voice loading, the MC-DV user capacity gain over 1xEV-DV is 47%. The corresponding MC-DV data sector throughput gain over 1xEV-DV is 11%. The much higher FTP user capacity gain of MC-DV in the mixed voice and FTP traffic scenario compared to the FTP-only traffic scenario in Section 5.2.2.2 is due to the additional multi-carrier diversity gain in terms of the varying left-over BS transmit power for F-PDCH across different carriers. The cross-carrier dynamic scheduling in MC-DV leverages the channel variation and BS transmit power variation across carriers to schedule high priority packets on favourable carriers, thus increases the FTP throughput and reduces the FTP user outage at cell edge.

6.3 Summary

In this chapter, we closely examined the different load balancing schemes across heterogeneous PL modes in the proposed B3G multi-mode access system. In our studies, we assume the PL modes are co-located in the same BS, where dynamic packet based load balancing can be performed. For cases where PL modes are either TDM or FDM, we proposed an integrated load balancing and scheduling scheme to maximize the overall system capacity while ensuring users' QoS and SLA are met. For cases where PL modes are CDM, we proposed a dynamic Walsh codes and BS transmit power sharing scheme between power-controlled dedicated traffic channels and rate-controlled packet data

channels across multiple CDMA carriers. Performance studies show that our proposed schemes outperform the conventional load balancing schemes.

Our proposed schemes can also be applied to the case where PL modes are not co-located in the same cell hierarchy and therefore not co-located in the same BS. Semi-static resource multiplexing in TDM or FDM can easily be employed. Dynamic resource multiplexing in either TDM or FDM can be performed through inter-BS communication either through direct backhaul links between BSs or via the AG. The performance benefit of dynamic load balancing is at the expense of inter-BS signalling overhead. To reduce signalling overhead, the rate of change of resource partitioning between PL modes can be reduced which will in turn degrade the performance of dynamic load balancing to somewhere in between the optimum dynamic load balancing case and the semi-static load balancing case shown in Figure 37 and Figure 38. One scenario of applying the proposed dynamic load balancing scheme to hierarchical cells is the case of relay, where the BS provides master control of resource partitioning between BS and associated relay stations (RSs). The BS communicates the resource partition information to the RSs as part of the existing BS to RS data and signalling flow without incurring much additional overhead. Detailed account of load balancing and resource multiplexing between BS and RS is for future studies and is out of the scope of this thesis.

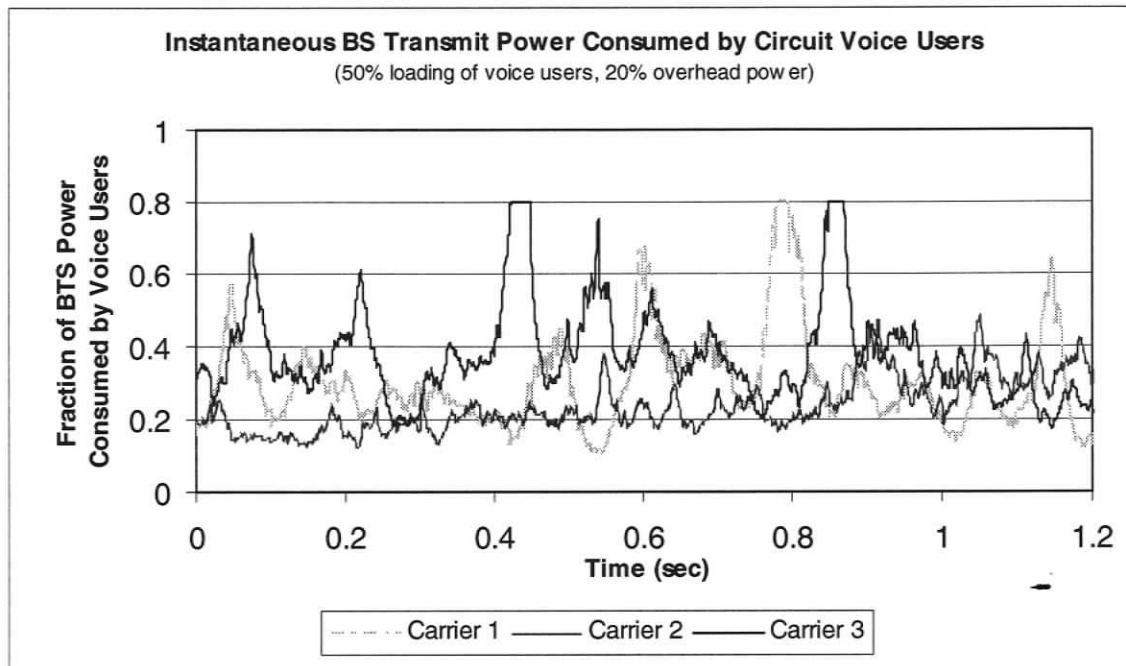


Figure 36 Instantaneous BS Transmit Power Consumed by Circuit Voice Users for 50% System Loading of EVRC RC3 Users, and 20% of BS Transmit Power Consumed by Overhead Channels

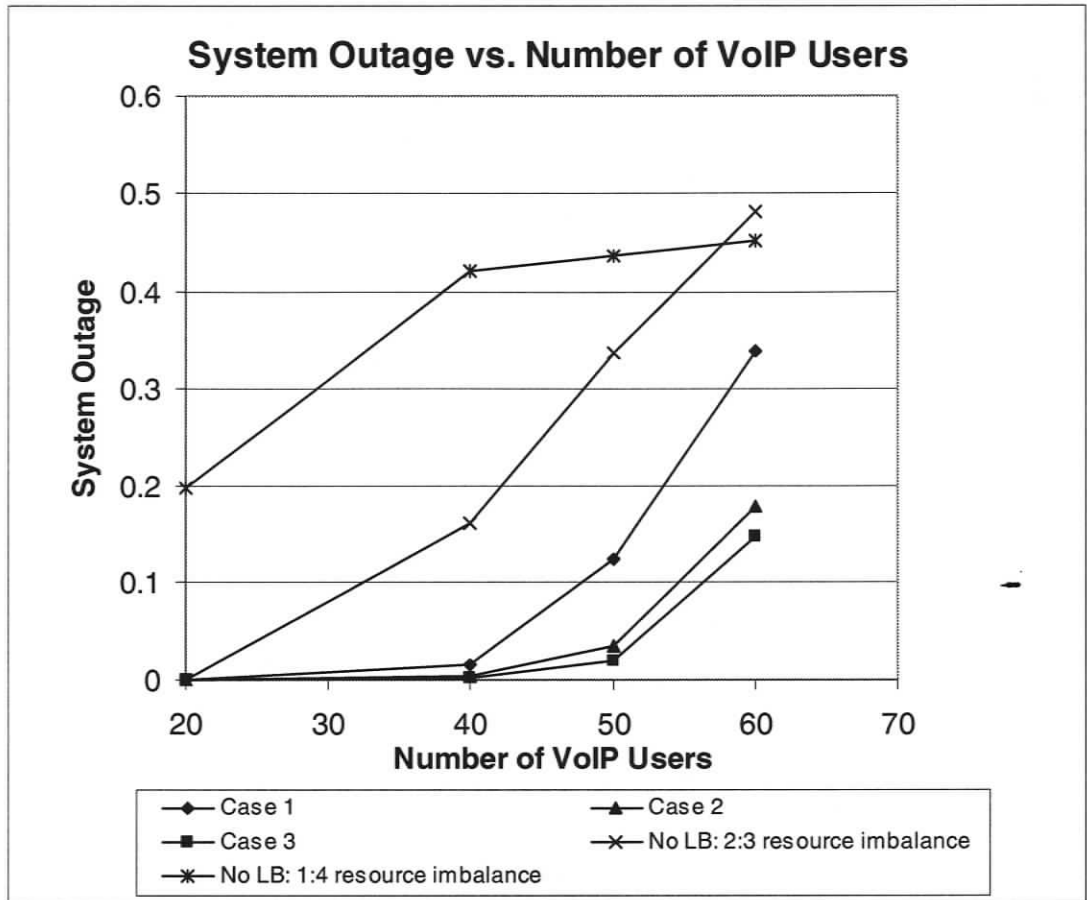


Figure 37 Performance Comparison of Load Balancing Schemes for a System Consists of Two PL modes

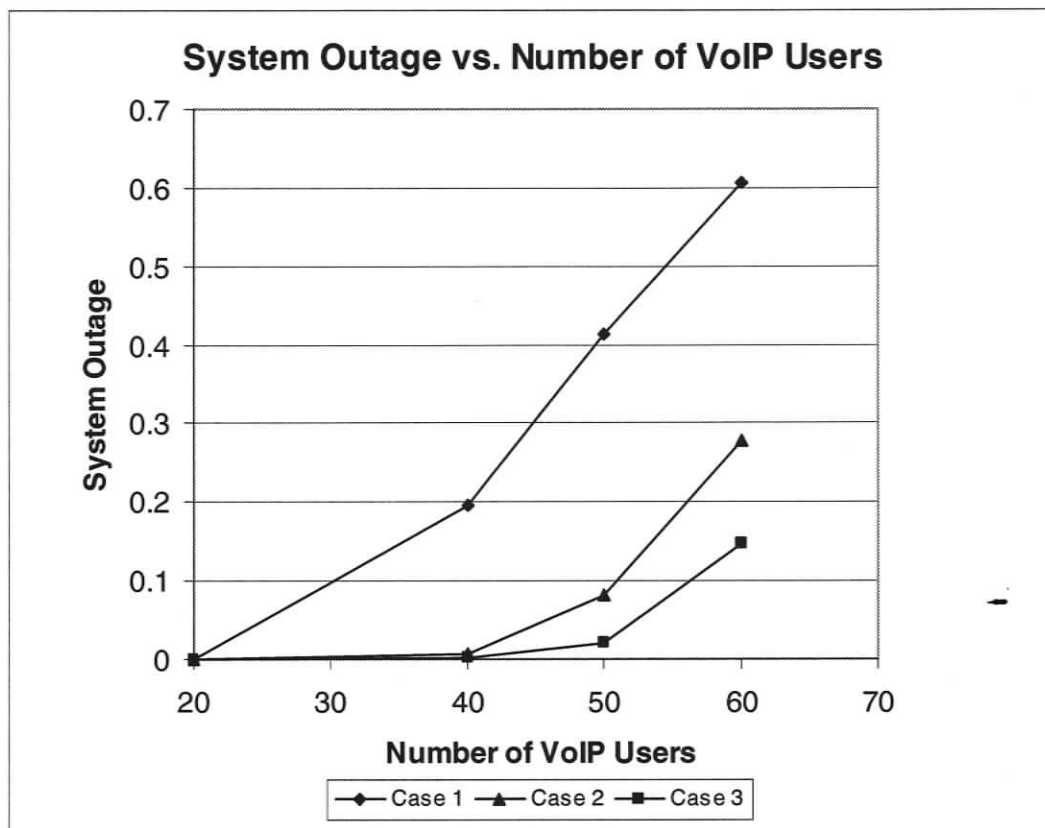


Figure 38 Performance Comparison of Load Balancing Schemes for a System Consists of Four PL Modes

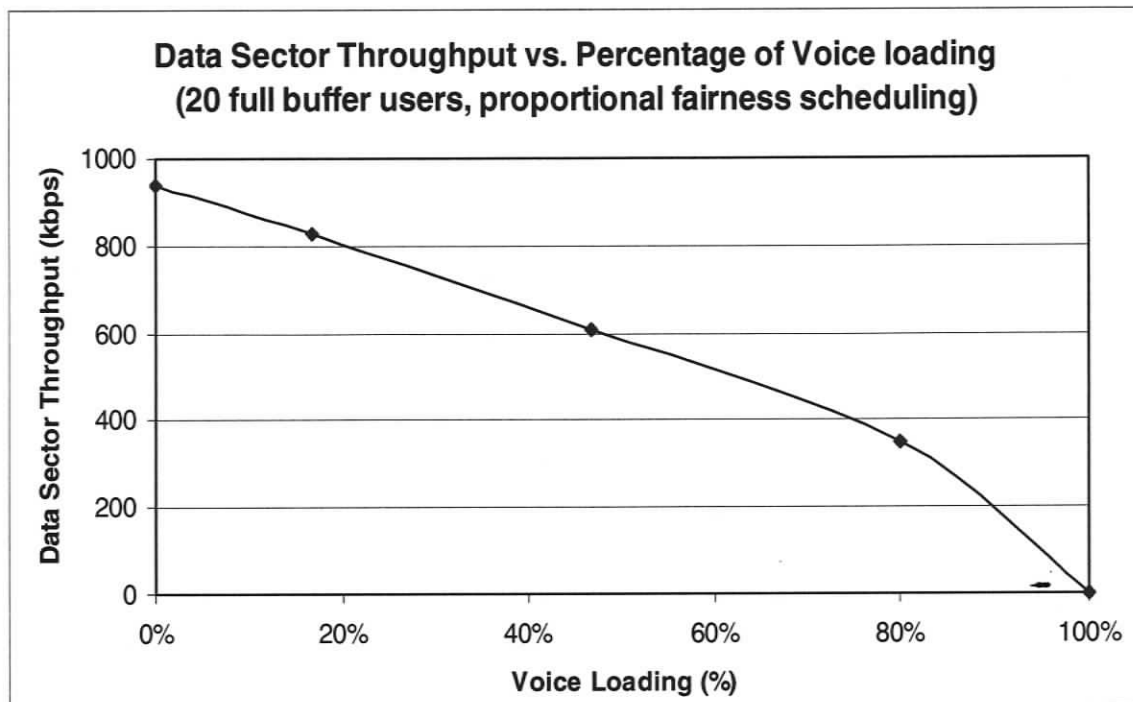


Figure 39 Data Sector Throughput versus Percentage of Voice Loading for Mixed 1xRTT EVRC Voice Traffic and 1xEV-DV Full Buffer Data Traffic

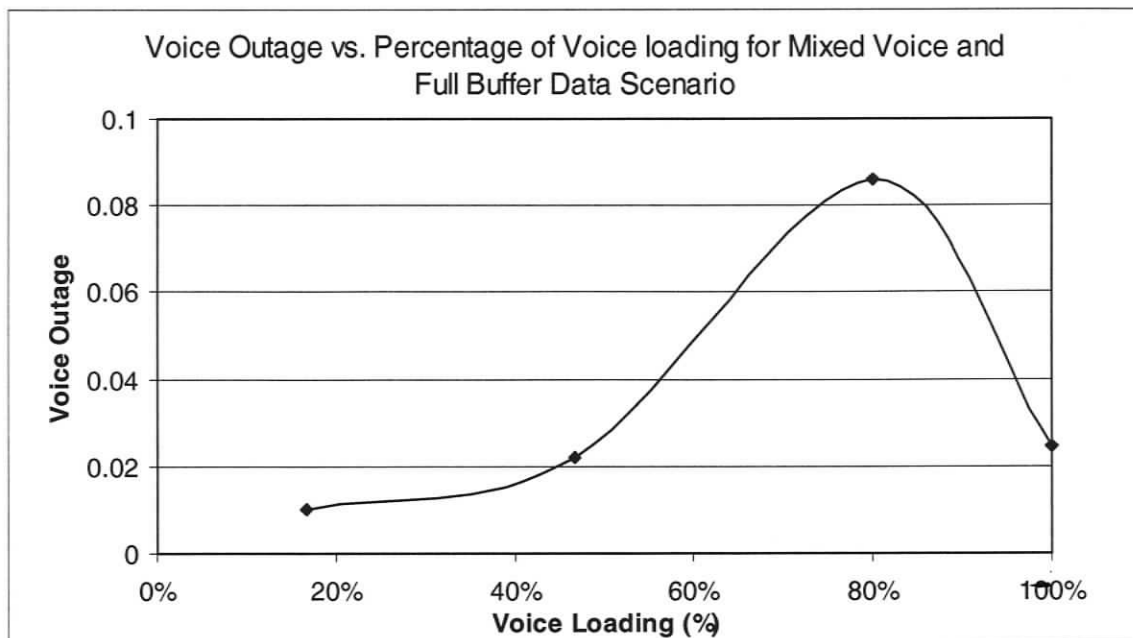


Figure 40 Voice Outage versus Percentage of Voice Loading for Mixed 1xRTT EVRC Voice Traffic and 1xEV-DV Full Buffer Data Traffic

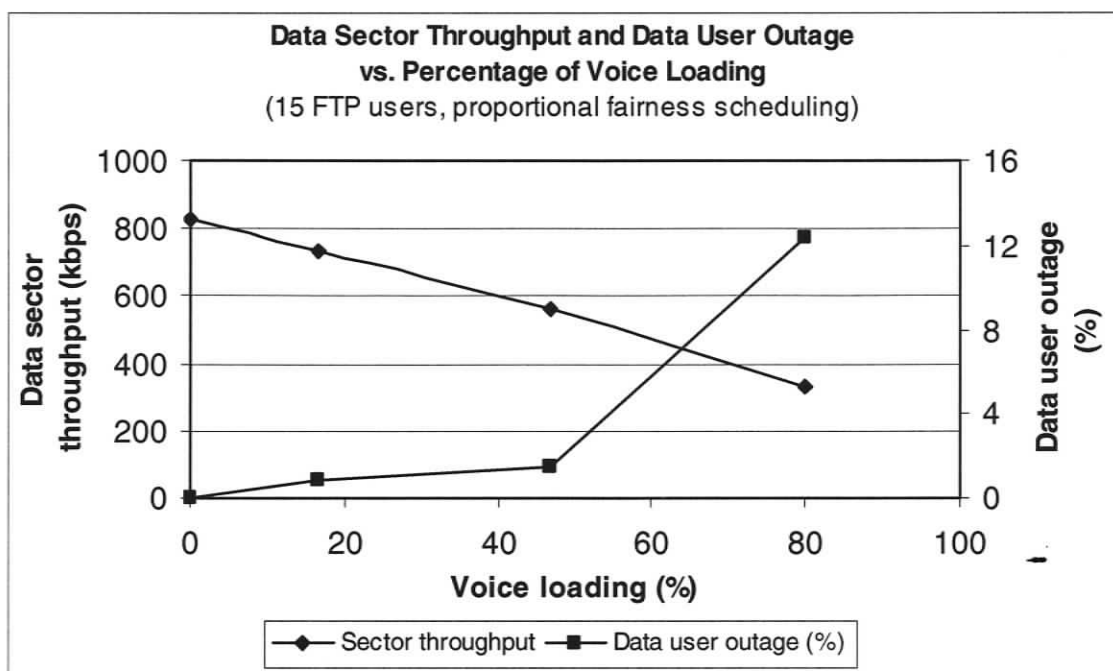


Figure 41 Data Sector Throughput and Data User Outage versus Percentage of Voice Loading, for Mixed 1xRTT EVRC Voice Traffic and 1xEV-DV FTP Data Traffic

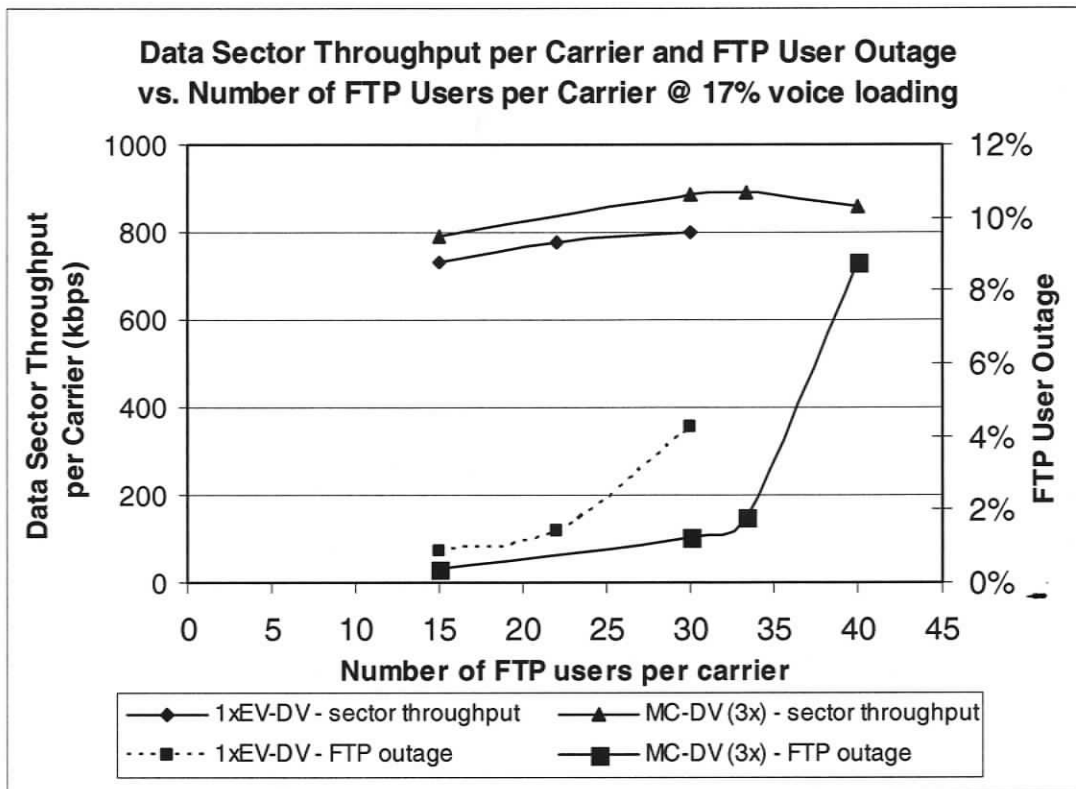


Figure 42 Data Sector Throughput per Carrier and FTP User Outage versus Number of FTP Users per Carrier at 17% Voice Loading per Carrier, for 1xEV-DV and MC-DV with 3 Carriers

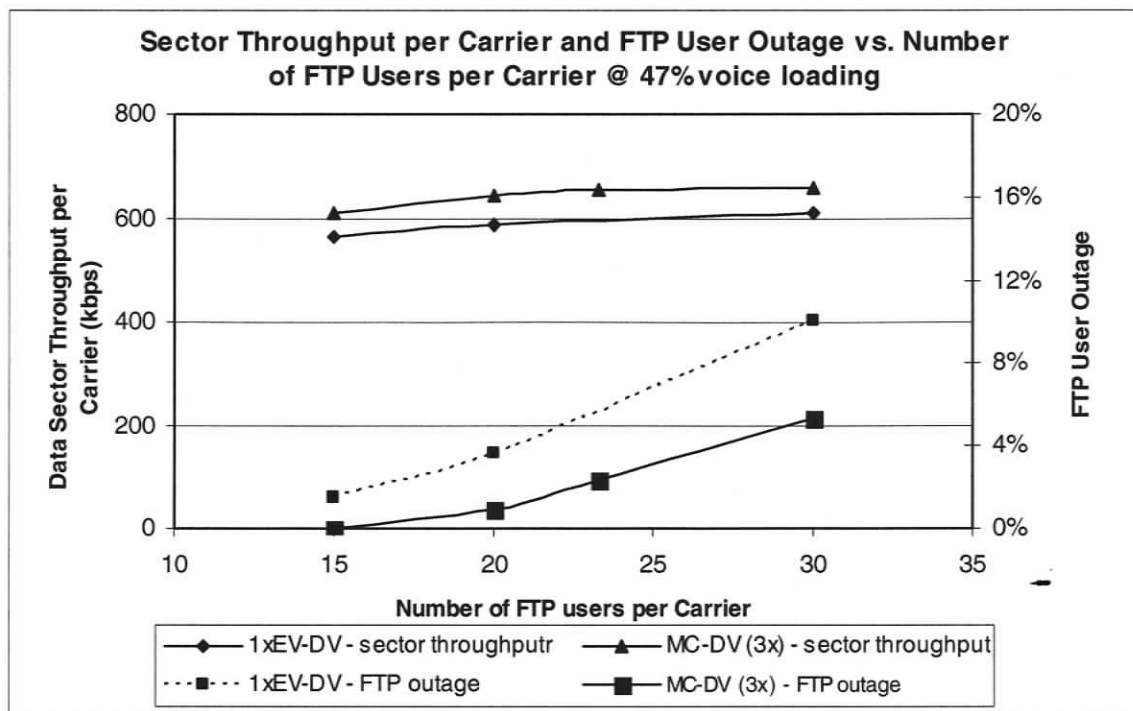


Figure 43 Data Sector Throughput per Carrier and FTP User Outage versus Number of FTP Users per Carrier at 47% Voice Loading per Carrier, for 1xEV-DV and MC-DV with 3 Carriers

Chapter 7

Universal MAC States

The concept of MAC states has been introduced in 3G cellular mobile systems, such as cdma2000 [1][9]-[13], UMTS [2][14] and 802.16e [15]. MAC states serve two main purposes i.e. power saving for the mobile terminals and improved air-interface resource utilization to support packet data services. A user is put in different operational states (or MAC states) based on its traffic activity or other radio resource management policy. Each state differs in terms of its required radio resource and power consumption. Two key MAC states defined in 3G systems are the Active state (also called the Connected state) and the Dormant state (also called the Idle state). There are also intermediate states introduced in between these two states, such as the Control-Hold state defined in [9]-[13], the Suspended state defined in [1] and the Sleep mode defined in [15]. A detailed account of these MAC states is given in Section 7.1.

In this chapter, we investigate the MAC states design for the proposed B3G multi-mode access system. As described earlier, the different PL modes of the B3G multi-mode access system can be co-located within the same sector/cell or located at different cell hierarchies. To support a user moving from one PL mode and/or cell hierarchy to another, we propose a universal MAC state machine concept that anchors the user no matter which PL mode or cell hierarchy it currently operates in.

The rest of the chapter is organized as follows. Section 7.1 provides a description of MAC states defined in the current 3G systems. Section 7.2 presents the proposed universal MAC states design for the B3G multi-mode access system. Section 7.3 presents

the performance evaluation of the proposed scheme, followed by a summary in Section 7.4.

7.1 Overview of MAC States in 3G Systems

As described earlier, MAC states are introduced in 3G mobile systems for the purposes of power saving and improved resource utilization to support packet data services.

Figure 44 below shows a generic MAC states diagram consisting of an Active state and a Dormant state which are the key MAC states, and intermediate states which are the Power-saving state and the Suspended state.

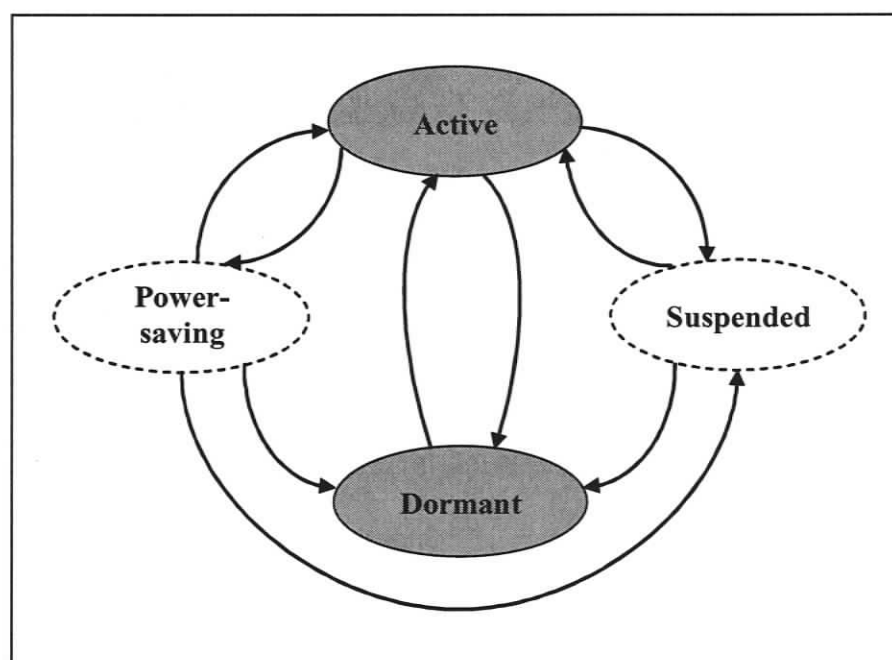


Figure 44 MAC States of a MS in 3G Systems

7.1.1 Active State

The Active state is the state where the MS is readily transmitting and receiving traffic in the FL and the RL respectively when instructed by the BS. The MS is synchronized to the BS in the physical layer in terms of time and frequency, and in layers 2 and 3 in terms

of protocol states. The MS monitors the control signalling from the BS on every time slot. In addition, the MS and the BS maintain a good estimate of the FL channel quality of the MS, while the MS maintains a good estimate of the required transmit power in the RL to achieve the target SNR. In [1], [2], [12], [13] and [15], the knowledge of MS' FL channel condition is maintained by periodic channel quality indication (CQI) feedback from the MS. In [1], [2], [12] and [13], which are based on a CDMA technique, the RL target SNR is maintained by the periodic and fast closed-loop power control operation. In [15], which is based on an OFDMA technique, the RL target SNR and frequency/time synchronization are maintained by the periodic ranging operation which is typically less frequent than the closed-loop power control operation in CDMA systems. As can be inferred from above, an MS in the Active state consumes the most battery power, radio resource and network resource. Putting an MS always in the Active state is therefore not efficient in terms power consumption, radio resource utilization and network resource utilization when supporting packet data applications. This is because packet data traffic is bursty in nature. Most of the time, the BS may not have data to transmit to the MS or the MS may not have data to transmit to the BS. For example, a typical reading time of a web-page is 30 seconds [27]. During this buffer-empty period, the MS can be put into a state where no transmission and reception activities are required. The Dormant state is therefore introduced.

7.1.2 Dormant State

The Dormant state is the state where the MS is unavailable for FL data reception and RL data transmission. BS has no knowledge of the FL channel condition of the MS and the MS does not maintain a good estimate of the required RL transmit power and

frequency/time transmission offset (for the case of OFDMA systems). Layers 2 and 3 protocol states are not retained or synchronized between BS and MS. The MS receiver turns on periodically (in order of hundreds of milliseconds or seconds) to monitor the paging information from the BS. In order for the MS to receive data on the FL or transmit data on the RL, a BS initiated (i.e. paging) or MS initiated call setup procedures need to be performed to transition the MS from the Dormant state to the Active state. Although the Dormant state has the least consumption of power, radio resource and network resource, it incurs undesirable click-response latency to the end-user. Depending on the specific network implementation, the state transition from the Dormant state to the Active state can be in the order of seconds. Therefore, there is a tradeoff between power/resource consumption and the QoS experienced by the end user.

7.1.3 Intermediate States

Intermediate states, i.e. the Power-saving state and the Suspended state as shown in Figure 44, are introduced in between the Active state and the Dormant state to provide different levels of tradeoff between power/resource consumption and QoS. The Power-saving state (also called the Control-Hold state in [9]-[13] and the Sleep mode in [15]) can be viewed as a pseudo-active state. The MS and the BS maintain the same physical layer, layer 2 and layer 3 states synchronization as in the Active state. The main difference is that a sleep interval is introduced in the Power-saving state to allow the MS to disable its transmitter and receiving during this interval. Sleep intervals are interlaced with listening intervals, where during listening intervals the MS operates as in the Active state. Although the Power-saving state consumes less power and radio/network resource than the Active state, it still consumes a relatively large amount of power and

radio/network resource compared to the Dormant state. The advantage of the Power-saving state is the much faster transition back to the Active State, in the order of tens of milliseconds, whenever there is data to be exchanged between the BS and the MS.

The Suspended state is a state that is similar to the Dormant state but provides faster or expedited call setup into the Active state, with latency in the order of hundreds of milliseconds. The key difference between the Suspended state and the Dormant state is that for the Suspended state, some semi-static layer 2 and layer 3 protocols context information is retained in the MS and the BS, such that during call setup, the corresponding signaling handshake between the MS and the BS to establish those context information can be omitted. Layer 2/3 protocol context information that can be retained in the Suspended state includes security context, service flow context, and MS/BS capability and preferred protocol configurations. In addition, location update can be performed by the MS in the Suspended state to allow the BS to closely track the MS' location with respect to a particular cell/sector or paging zone.

7.1.4 State Transition

The possible state transition between the various MAC states is shown by the 'arrows' in Figure 44. It is important to note that not all MAC states need to be supported in a system. In a specific implementation, the BS can decide the MAC state to transition the MS to, based on factors such as buffer status, QoS requirement, SLA and radio resource management policy.

7.2 Universal MAC States Concept for B3G Multi-MODE Access System

In the proposed B3G multi-mode access system, a common layer 2 and layer 3 protocol stack is used to anchor the different PL modes. The common layer 2/3 protocol provides mobility management, radio resource management, load control and spectrum management across the different PL modes. A universal MAC state machine is defined such that when a user switches from one PL mode to another either within the same or different cell hierarchies, the MAC state and the associated context information of the MS can be retained. In this way, packet loss and PL mode switching/handoff latency can be minimized.

7.2.1 Universal MAC States Definition

We propose that a universal MAC states machine be defined to comprise of a superset of MAC states relevant to different PL modes. For example, for PL mode 1 which is used to support fixed services, only the Active state is applicable since power saving is not an issue for fixed devices. Air interface and network resource consumption of an Active state user can be reduced by reducing the FL channel quality feedback rate and the RL power control rate. For PL mode 2 which is used to support mobility services, all four states as described in Section 7.1 are applicable. For PL mode 3 which is used to support nomadic services, only the Active state and the Dormant state are applicable since power consumption is less of an issue for laptop type devices compared to handheld type devices. The universal MAC states machine therefore consists of all the four states: Active, Dormant, Power-saving and Suspended; as illustrated in Figure 45. When a MS

operates in a particular PL mode, only a subset of the MAC states and the corresponding state transition as shown in Figure 45 will be in operation.

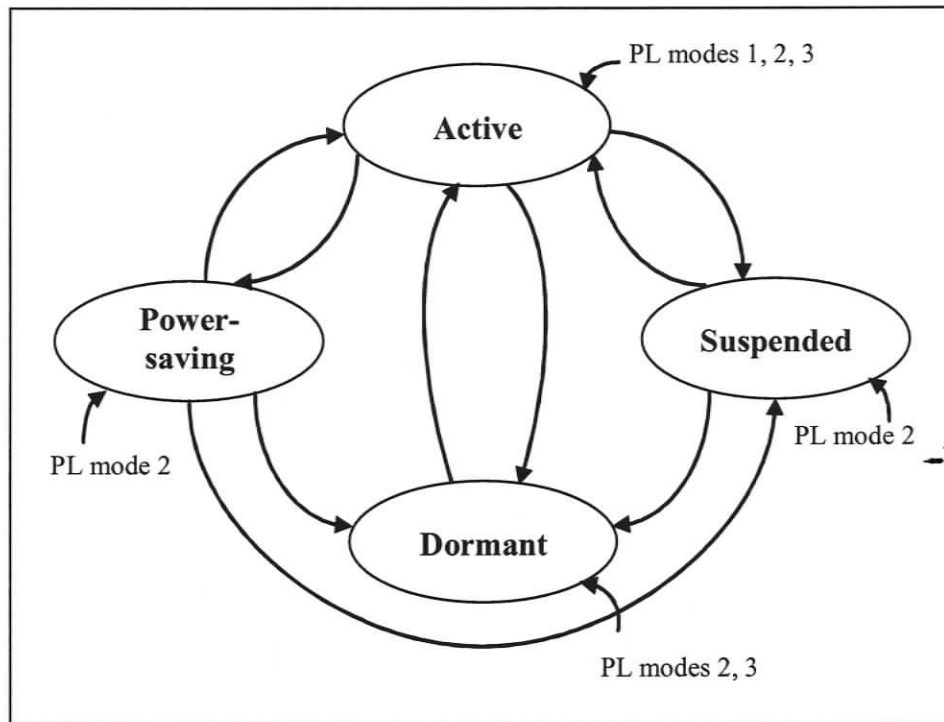


Figure 45 Universal MAC States (An Illustration of the Subset of MAC States Supported by each PL Mode)

7.2.2 MAC State Operation during PL Mode Switching

There are different MAC state transition scenarios when a MS switches from one PL mode to another, depending on the MAC state of the MS in the previous PL mode and the MAC states supported by the new PL mode.

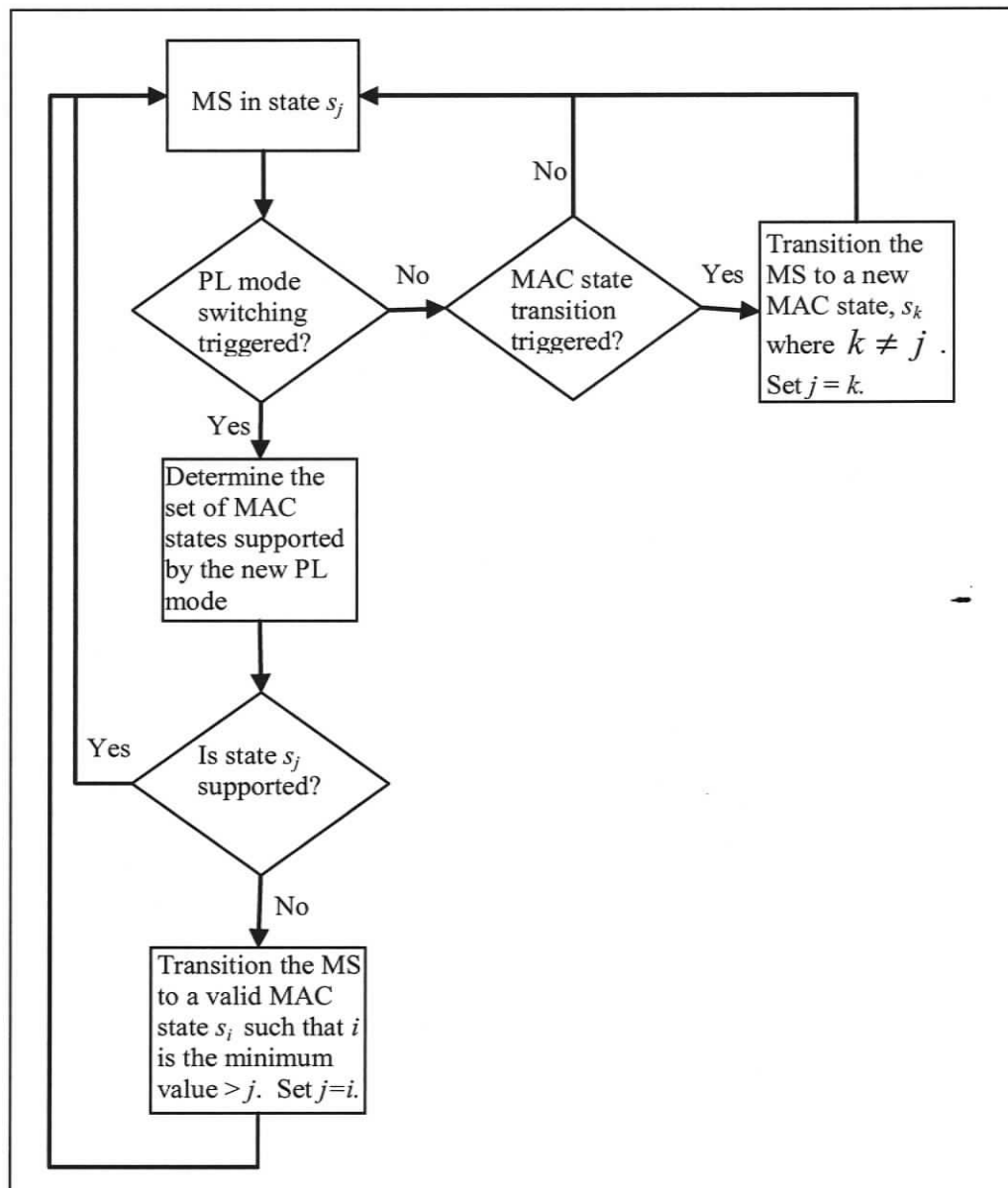


Figure 46 Algorithm for MAC State Transition during PL Mode Switching

The general algorithm is illustrated in Figure 46. We define the set of MAC states as $S = \{\text{Dormant, Suspended, Power-saving, Active}\}$, where s_i is a member of the set for $i = 0, 1, 2, 3$. As shown in Figure 46, when the MS switches from one PL mode to another, the current MAC state at the previous PL mode is retained if the new PL mode supports the same MAC state; otherwise, the MS transitions to the next higher MAC state supported

by the new PL mode and the allowable state transition flow given in Figure 45. These two scenarios are further illustrated in the following sub-sections.

7.2.2.1 Scenario 1: MS Retains the Same MAC State When Switching to a New PL Mode

For illustration purposes, we use the same example as before where PL mode 1 supports the Active state only, PL mode 2 supports all four MAC states, and PL mode 3 supports the Active state and the Dormant state.

While in the Active state, a MS continuously monitors the signal strength of the surrounding BSs and their associated PL modes. PL mode switching can be initiated by either the BS or the MS based on the change in signal strength, mobility condition, service needs or system loading. If the MS switches among the three PL modes while in the Active state, the MS remains in the Active state unless there is insufficient resource in the target PL mode to accommodate the MS in the Active state, in which case, the MS will be downgraded into a lower MAC state supported by the target PL mode. If the PL modes are collocated at the same BS, there is no need for layer 2/3 protocol context information transfer. The only data flow interruption time incurred during PL mode switching is the required time to change the PL configurations at the MS and the BS. The switching latency is in the order of milliseconds to tens of milliseconds. If the PL modes are located at different BSs, the distributed layer 2/3 context information has to be transferred from the previous BS to the target BS. The latency incurred is dependent on the backbone network, which may take tens to hundreds of milliseconds. As a comparison, in B3G heterogeneous networks where there is no convergence at the layer 2/3 [4][5][6], the MS has to perform call setup procedures (which is the same as Dormant state to Active state transition) to re-establish the air interface physical layer, layer 2 and

layer 3 protocol states when switching between heterogeneous systems. The latency can be in the order of seconds. In addition, layer 2/3 packet loss cannot be avoided in this case.

For PL modes 2 and 3, while in the Dormant state, the MS continuously monitors the FL signal strength of the surrounding BSs and their associated PL modes. The MS initiates the PL mode switching, typically due to changes in mobility condition and the received signal strength from different cell hierarchies, by exchanging layer 3 signaling with the target BS associated with the target PL mode. The layer 3 signaling is carried on the FL and RL common channels associated with the target PL mode. The Dormant state related operational parameters of the MS, such as FL paging cycle, RL random access channel parameters, can be updated during the layer 3 signaling exchange. The MS remains in the Dormant state after switching to the new PL mode.

The operation of PL modes switching when a MS is in Suspended state is similar to that of the Dormant state. In addition, when a MS in the Suspended state switches to a new PL mode that supports the Suspended state, the MS remains in the Suspended state with all the stored context information carried forward to the new PL mode. In this way, fast call setup can be performed when the MS transitions to the Active state at a later stage. Likewise, when a MS in the Power-saving state switches to a new PL mode that supports the Power-saving state, the MS remains in the Power-saving state with the layer 2/3 context information intact.

Overall, the universal MAC state machine provides seamless switching from one PL mode to another. It avoids the unnecessary call re-establishment overhead and latency

that would otherwise be incurred when a MS moves between heterogeneous systems that do not have a converged layer 2/3.

7.2.2.2 Scenario 2: MS Transitions to a New MAC State when Switching to a New PL Mode

This scenario occurs when a MS switches to a target PL mode which does not support the MAC state of the MS in the previous PL mode. In this case, the MS is transitioned to the next higher MAC state supported by the target PL mode, if there is sufficient resource at the target PL mode to accommodate the MS in that MAC state. Otherwise, the MS will be downgraded to a lower MAC state supported by the target PL mode. For example, if a MS in PL mode 3 and the Dormant state switches to PL mode 1, the MS will perform state transition to the Active state if there is sufficient resource in PL mode 1 and remain in the Active state while in PL mode 1. Similarly, if a MS in PL mode 2 and the Suspended state switches to PL mode 3, the MS will perform state transition to the Active state. The protocol context information retained in the Suspended state is carried over to PL mode 3, thus reducing the state transition latency from the Suspended state to the Active state. After switching to PL mode 3, the MS may transition between the Active state and the Dormant state according the operational policy of PL mode 3.

Similar to scenario 1, the universal MAC states approach avoids the need to perform call re-establishment when moving from one PL mode to another.

7.3 Performance Evaluation

As described earlier, the goals of the universal MAC states concept are to ensure a user's state is maintained while switching between PL modes, thus minimizing latency and overhead involved in state transitions. In this section, we evaluate the generic impact

of MAC states and state transition latency on packet data performance, in order to demonstrate the importance of the universal MAC states concept.

7.3.1 System Model

We perform a full-scale FL system-level simulation based on the methodology described in Chapter 3. We use mixed mobile speeds with probability as given in Chapter 3. Both the physical layer and the MAC layer of the 1xEV-DV system are modeled. We have chosen 1xEV-DV as the system for evaluation since it contains all the necessary MAC states.

The following MAC states scenarios are simulated, with the goal of demonstrating the impact of intermediate MAC states and therefore their retention during PL mode switching and state transition latency on overall system performance:

1. Dormant state and Active state, with Dormant to Active transition latency of 2 seconds
2. Dormant state with fast call setup (similar to the Suspended state) and Active state, with Dormant to Active transition latency of 1 seconds
3. Dormant state with fast call setup, Control-hold state (similar to the Power-saving state) and Active state, with Dormant to Active transition latency of 1 second, and Control-hold to Active transition latency of 40ms.

For scenarios 1 and 2, the maximum number of Active state users supported by the system is assumed to be 32. This is to account for the RL pilot and CQI feedback overhead incurred by each Active state user. For scenario 3, we assume a maximum of 19 Active state users and 31 Control-hold state users are supported by the system. The reason for the larger number of Control-hold users is users in Control-hold state perform

gated RL pilot and CQI transmission at $\frac{1}{4}$ rate, thus reducing the overhead incurred to the system. These numbers are obtained based on the calculation given in Appendix A.

The decision to transition a user from the Active state to the Dormant state or from the Active state to the Control-hold state is based on an idle timer called T_{active} which starts when a user's buffer becomes empty. When T_{active} expires, the user is transitioned to the Dormant state for scenarios 1 and 2, and to the Control-hold state for scenario 3. In this study, T_{active} is set to 6 seconds for scenarios 1 and 2, and is set to 200 milliseconds for scenario 3. For scenario 3, the decision to transition a user from the Control-hold state to the Dormant state is based on another timer called T_{c-h} , which is set to 5.8 seconds in this study.

7.3.2 Traffic Model

The packet data services used in this study are a mix of web-browsing (27%), FTP (10%) and WAP (63%) services. The relative percentage mix is based on the recommendation in [27].

7.3.3 Performance Metric

The performance metric and outage criteria for FTP, web-browsing and WAP services are as described in Chapter 3.

We compare the performance of scenarios 1, 2 and 3, in terms of their system capacity and corresponding sector throughput, where system capacity is defined as the maximum achievable user loading while maintaining a system outage of less than or equal to 2%..

For each scenario, we average the results across 20 simulation runs. Each simulation run is 450 seconds, with each mobile changes its location in a sector every 30 seconds.

7.3.4 Performance Results

Figure 47 and Figure 48 show the system capacity and sector throughput performance of scenarios 1, 2 and 3.

By reducing the transition delay from the Dormant state to the Active state, scenario 2 has 10.9% higher system capacity and 11.8% higher sector throughput than scenario 1. Since state transition causes degradation in packet or packet call throughput which in turns causes user outage, a reduction in the state transition latency reduces user outage, thus improves the system capacity.

On the other hand, for scenario 3, the use of the Power-saving state, i.e. the Control-hold state, provides further performance gain of 29% in terms of system capacity and 20% in terms of sector throughput compared to scenario 2. The use of the Power-saving state increases the pool of active and pseudo-active users, thus reduces the state transition probability. In addition, the number of users available to the MAC scheduler also increases thus increases the overall sector throughput due to multi-user diversity gain.

Overall, scenario 3, with the combined effect of reducing Dormant state to Active state transition latency and using a power-saving state, provides 43% gain in terms of system capacity, and 34% gain in terms of sector throughput compared to scenario 1.

From the simulation results shown, we can see the importance of retaining a user's MAC state when a user switches from one PL mode to another, in particular if the user is in the Active state, the Power-saving state or the Suspended state. Otherwise, a user needs to perform call re-establishment (which has the same latency and overhead as the Dormant state to Active state transition) when switching from one PL mode to another.

This degrades user performance and overall system performance, as shown in the inferior performance of scenario 1 compared to scenarios 2 and 3.

7.4 Summary

A universal MAC states concept was proposed in this chapter as part of the proposed multi-mode wireless access system framework for B3G. The benefit and importance of the proposed concept was demonstrated through a performance study.

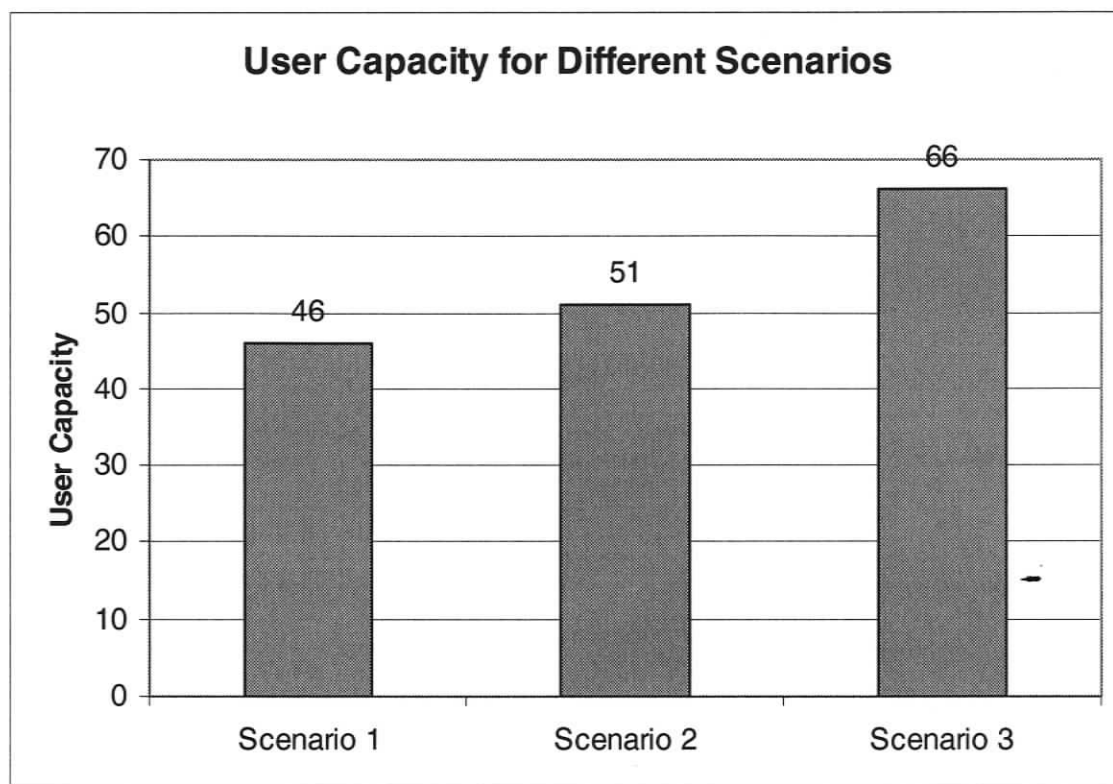


Figure 47 Comparison of User Capacity for Different MAC States Scenarios

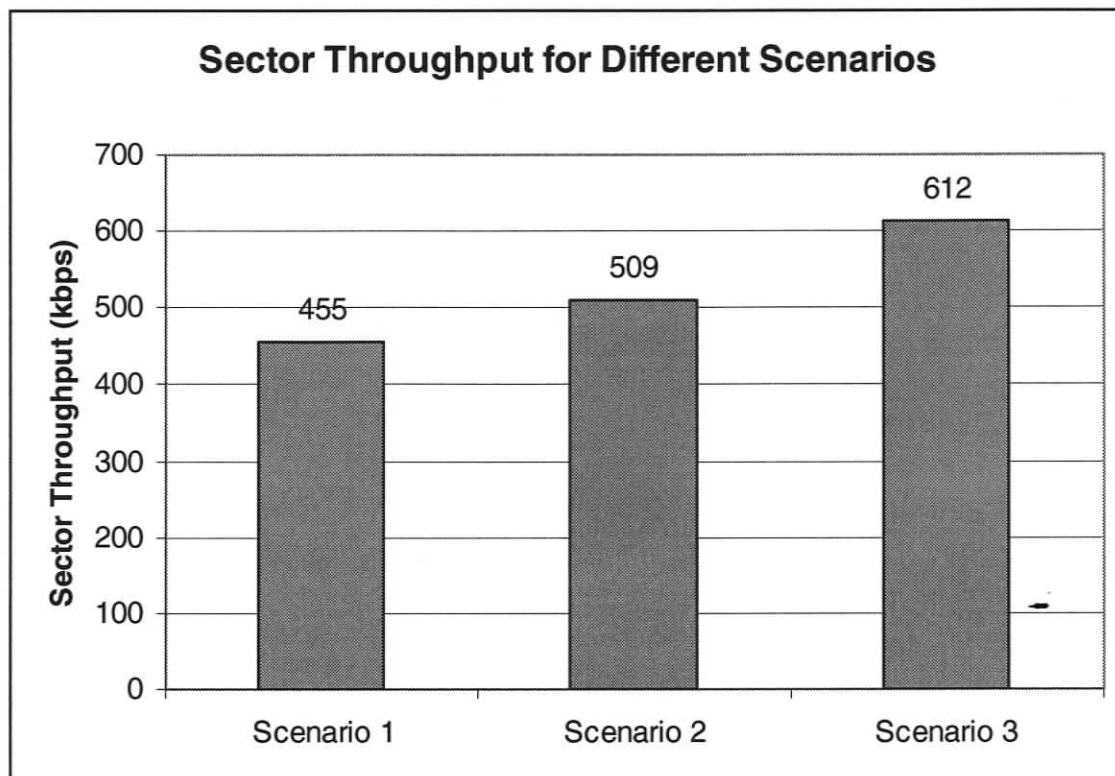


Figure 48 Comparison of Sector Throughput for Different MAC States Scenarios

Chapter 8

MAC State Transition Algorithm

In this chapter, we investigate the details of MAC state transition algorithm design. The MAC state transition algorithm applies to both the conventional 3G system and the proposed B3G multi-mode access system.

As described in Chapter 7, MAC states and MAC state transition aim to reduce terminal power consumption as well as to reduce radio and network resource consumption when the BS has no data to transmit to or receive from the MS. The decision to downgrade a user to a lower state, i.e. from Active to Power-saving, from Power-saving to Suspended etc., is typically based on idle timers which start when the user's buffers at the BS and the MS become empty [41][42]. A user is typically transitioned out of Dormant, Suspended or Power-saving states into the Active state when there is new traffic arrival into the user's buffers at the BS or the MS.

There are several issues with this conventional state transition approach. First, the state transition decision is decoupled from the scheduler operation. A user who is high in the scheduler priority ranking but with larger idle timer value will be transitioned out of the Active state before a user with low priority ranking but smaller idle timer value. Since the scheduler priority ranking is designed to maximize system capacity at a given QoS and fairness constraint, it is desirable to ensure users with high priority ranking remain in the Active state to optimize the scheduling efficiency. Secondly, when the idle timer expires, the user is transitioned out of the Active state even though there may not be other users pending to enter the Active state. Thirdly, once a user enters the Active state, it remains

in the Active state until the idle timer expires regardless of the user's priority ranking. This reduces the scheduling efficiency since a higher priority ranking user with pending data in its buffer may be prevented from entering the Active state due to lack of resource. Fourthly, in the conventional approach, users with pending data in their buffers who are not in the Active state are allowed to transition into the Active state in a first-come-first-served basis when there is available resource in the Active state. This can result in users with lower priority ranking entering the Active state before users with higher priority ranking.

In this chapter, we introduce a novel state transition algorithm, called the priority based MAC state transition algorithm, that integrates with the scheduler prioritization operation, with the objective of maximizing the system capacity at a given the QoS and fairness constraint. In Section 8.1, we provide a detailed description of the proposed scheme. In Section 8.2, we evaluate the performance of the proposed scheme, followed by a summary in Section 8.3.

8.1 Priority Based MAC State Transition Algorithm

The basic principle of the priority based MAC state transition algorithm is to employ similar priority criteria as the scheduler to determine the assigned MAC state of each user in the system. The state transition priority of each user is evaluated periodically by the state transition controller at the BS to determine if a user should be transitioned to another state. We first describe the generic algorithm in Section 8.1.1 that applies to a single PL mode system, like the current 3G system. In Section 8.1.2, we describe how the algorithm applies to the proposed B3G multi-mode system with universal MAC states as presented in Chapter 7.

8.1.1 Generic Algorithm Description

The following procedures are performed at each state transition decision instance:

Step 1: Calculate the state transition priority of all users in the system, i.e. users who are in one of the MAC states, based on a certain priority equation. A similar priority equation as that of the scheduler should be used to ensure overall optimum scheduling efficiency. However, since the state transition priority is per-user rather than per-packet, the scheduler priority equation shown in equation (4) needs to be modified to remove per-packet related parameters. The state transition priority of user i at state transition instance n and PL mode m can be expressed as

$$(25) \quad P_i^{(m)}(n) = f(R_i^{(m)}(n), \bar{r}_i(n), \min\{r_i, \tau_i, B_i\}),$$

where $R_i^{(m)}(n)$ is the short-term average data rate supported by user i at state transition instance n based on the short-term average channel condition experienced by the user at PL mode m . The reason to use the short-term average data rate rather than the instantaneous data rate, $r_i^{(m)}(n)$, as in the case of equation (4) is the fast variation of instantaneous data rate will cause unnecessary and undesirable amount of state transitions. τ_i is the amount of time the buffer of user i is empty. B_i is the buffer occupancy of user i .

If a proportional fairness scheduler or a fairness and delay scheduler is used as shown in equations (9) and (10) respectively, the state transition priority equation becomes

$$(26) \quad P_i^{(m)}(n) = \frac{[R_i^{(m)}(n)]^\alpha}{[1 + \bar{r}_i(n)]^\beta r_{\max}} + \text{sgn}(B_i).$$

The parameter τ_i is not part of the above state transition priority equation but is used to decide if an Active state user should remain in the Active state in Step 3. The above state transition priority equation always ranks a user with non-empty buffer higher priority than a user with empty buffers. For users who are in the Dormant state or the Suspended state, there is no periodic CIR feedback sent to the BS. Therefore the value of $R_i^{(m)}(n)$ is unavailable at the BS state transition controller. For those users, the state transition controller assigns a predetermined constant value C to replace $R_i^{(m)}(n)$. The state transition priority equation for users in the Dormant state and the Suspended state is therefore

$$(27) \quad P_i^{(m)}(n) = \frac{C}{[1 + \bar{r}_i(n)]^\beta r_{\max}} + \text{sgn}(B_i).$$

C is chosen such that it represents the average CIR experienced by users in the system.

Step 2: Calculate the number of users allowed in each state, i.e. the number of users allowed in the Active state, N_a ; the number of users allowed in the Power-saving state, N_p ; the number of users allowed in the Suspended state, N_s ; and the number of users allowed in the Dormant state, N_d . The values of N_s and N_d are typically fixed based on available network resource in the system. Their values are typically very large compared to the air-interface capacity. The values of N_a and N_p can be semi-statically changed based on the radio resource partition between users in the Active state and users in the Power-saving state.

Step 3: Determine the existing Active state and Power-saving state users who should remain in either the Active state or the Power-saving state, regardless of their state transition priority ranking. This is to prevent frequent state transition and moving a user

out of the Active state or the Power-saving state in the middle of a packet call where the user's buffer can become temporarily empty due to TCP ramp-up or HTTP request-response delay. We introduce a buffer empty timeout period, $T_{timeout}$, for users who are in the Active state or the Power-saving state. Users who are already in the Active state or the Power-saving state and with $\tau_i \leq T_{timeout}$ will remain in either the Active state or the Power-saving state. These users can move between the Active state and the Power-saving state based on their state transition priority. The value of $T_{timeout}$ is chosen to provide a good trade-off between overhead associated with state transition, latency and packet call throughput degradation caused by unnecessary state transition and scheduler efficiency or overall system capacity. Setting $T_{timeout}$ to 0 provides the best scheduler efficiency at the expense of state transition signalling overhead and increased latency or decreased packet call throughput caused by unnecessary state transition. On the other hand, setting $T_{timeout}$ to a large value reduces the scheduler efficiency, thus the system capacity.

Next, rank all the users from high to low priority based on the priority values calculated in Step 1. Determine the set of users, X , that are not the top $(N_a + N_p)$ users in the priority ranking, but are currently in either the Active state or the Power-saving state with $\tau_i \leq T_{timeout}$. The number of users in X is N_x . Assign a MAC state to each user in X as follows:

- If $N_x \leq N_p$, assign all users in X to the Power-saving state. Calculate the remaining room available in the Power-saving state as $N_p' = N_p - N_x$. Set the remaining room available in the Active state as $N_a' = N_a$.

- Otherwise, assign N_p lowest priority users in X to the Power-saving state and $(N_x - N_p)$ remaining users to the Active state. Set the remaining room available in the Power-saving state as $N_p' = 0$. Calculate the remaining room available in the Active state as $N_a' = N_a - (N_x - N_p)$.

Step 4: Determine users who should be moved to the Suspended state or the Dormant state regardless of their state transition priority ranking. This is to put those users with buffer empty for a long duration to the Suspended state or the Dormant state for power-saving purpose. We define the idle timers to move a user to the Suspended state and the Dormant state as $T_{to-suspended}$ and $T_{to-dormant}$, respectively. When $\tau_i > T_{to-suspended}$, user i who is in either the Active state or the Power-saving state is transitioned to the Suspended state. Update the remaining resource available in the Suspended state to N_s' . When $\tau_j > T_{to-dormant}$, user j is moved to the Dormant state regardless of its current state. Update the remaining resource available in the Dormant state to N_d' .

Step 5: Rank the remaining users in the system who are not assigned a MAC state in Steps 3 and 4. Assign the top N_a' users in the priority ranking to the Active state, the following N_p' users to the Power-saving state, the following N_s' users to the Suspended state and the following N_d' users to the Dormant state. If state transition from the current MAC state to the newly assigned MAC state is not valid, e.g. Dormant to Suspended, Suspended to Power-saving and Dormant to Power-saving, as shown in Figure 45, the user remains in its current MAC state.

The above state transition decision process is performed at a predetermined rate called the state transition update rate. The higher the state transition update rate, the higher the scheduler efficiency and the higher the state transition frequency or overhead. The converse is true.

The proposed priority based state transition algorithm couples closely with the scheduling priority for the same objective of maximizing the system capacity while meeting users' QoS and SLA requirements.

8.1.2 Application of the Algorithm to the Multi-mode Access System

As described in Chapter 7, all the PL modes are anchored by a universal MAC state machine. When a MS moves from one PL mode to another, it either retains its current MAC state if such a state is supported by the target PL mode, or it transitions to the next higher MAC state supported by the target PL mode if its current state is not supported by the target PL mode. The description of state transition during PL mode switching in Chapter 7 is based on the conventional state transition approach, i.e. no prioritization among users in the system to determine the MAC state assigned to each user.

Here, we apply the priority based state transition algorithm to the universal MAC state operation as follows:

- Since the resources available for each MAC state, i.e. N_a , N_p , N_s and N_d , is specific to each PL mode, each PL mode manages the state transition priority for users within the PL mode.
- When a user switches to the target PL mode, the state transition controller at the target PL mode ranks the user's state transition priority among all users in the target PL mode to determine the MAC state to assign to the user. Since the

layer 2/3 context information of the user is retained when switching to the target PL mode, the state transition controller at the target PL mode uses those information, such as $R_i^{(m)}(n)$, $\bar{r}_i(n)$, τ_i and B_i , to calculate the user's state transition priority.

The above scheme allows a distributed MAC state control in each PL mode to reduce implementation complexity and facilitate network scalability, while at the same time ensures coherence and integrated approach of MAC state transition and scheduling across different PL modes.

8.2 Performance Evaluation

In this section, our objective is to evaluate several performance aspects of the priority based state transition algorithm. First, we compare its performance with the conventional timer-based state transition algorithm. Secondly, we evaluate its performance with different number of MAC states. Thirdly, we examine the state transition frequency of the proposed scheme versus the conventional scheme, which has direct implication on signalling overhead.

8.2.1 System Level Simulation Configurations

The same 1xEV-DV system model as described in Section 7.3 is used here. We evaluate and compare the performance of the following scenarios:

Scenario 1: Conventional timer-based state transition algorithm

- A. Two MAC states: Dormant with fast call setup, and Active.
- B. Three MAC states: Dormant with fast call setup, Control-hold and Active

Scenario 2: Priority-based state transition algorithm

A. Two MAC states: Dormant with fast call setup, and Active.

B. Three MAC states: Dormant with fast call setup, Control-hold and Active

As in Section 7.3, for the two MAC states cases, the Dormant to Active transition latency is 1 second. For the three MAC states cases, the Dormant to Active transition latency is 1 second, and the Control-hold to Active transition latency is 40ms. As assumed in Section 7.3, for the two MAC states cases, the maximum allowable number of Active state users is 32, whereas for the three MAC states cases, the maximum allowable number of Active state users is 19 and the maximum allowable number of Control-hold state users is 31.

The parameters setting used for Scenario 1 are as follow:

- $T_{active} = 6s$ for Scenario 1A, $T_{active} = 200ms$ for Scenario 1B
- $T_{c-h} = 5.8s$ for Scenario 1B.

The above idle timers to transition from the Active state or the Control-hold state to the Dormant state are set to relatively large values in order to cater for traffic with different levels of burstiness, e.g. HTTP, FTP and WAP. If the idle timer value is set too small, a user may transition into Dormant state if its buffer becomes temporarily empty during a packet call or packet transmission, e.g. in between WAP request and WAP response, or in between web-page embedded objects download. For the conventional timer-based algorithm, once a user enters the Dormant state, it has to wait until there is resource available in the Active state before it can transition back to the Active state.

The parameters setting used for Scenario 2 are as follow:

- For a user in the Active state, $R_i^{(m)}(n)$ is calculated based on instantaneous CIR feedback from the MS. For a user in the Control-hold state, $R_i^{(m)}(n)$ is

calculated based on an IIR filtered CIR with filtering length of 128 slots, where the slot duration of 1xEV-DV is 1.25ms. For a user in the Dormant state, C is set to 1, which is equivalent to an average CIR of 0dB.

- State transition priority update frequency: 800Hz
- For Scenario 2A, $T_{timeout} = 200ms$. For Scenario 2B, $T_{timeout} = 0$. This is because with three MAC states, the allowable number of Active state and Power-saving state users is large enough to support the number of concurrent users who have data in their buffer, for the traffic models considered.
- $T_{to-dormant} = 6s$.

8.2.2 Performance Comparison of State Transition Algorithms

The performance of the priority based state transition algorithm is compared with that of the timer-based state transition algorithm under FTP only traffic, HTTP only traffic, WAP only traffic and mixed FTP/HTTP/WAP traffic with the percentage mix as given in Section 7.3. The performance metric and outage criteria for FTP, HTTP and WAP services are as described in Chapter 3.

We compare the performance of scenarios 1A, 1B, 2A and 2B in terms of their system capacity and corresponding sector throughput, where system capacity is defined as the maximum achievable user loading while maintaining a system outage of less than or equal to 2%. For each scenario, we average the results across 20 simulation runs. Each simulation run is 450 seconds, with each mobile changes its location in a sector every 30 seconds.

8.2.2.1 FTP Traffic

Figure 49 and Figure 50 show the user capacity and the sector throughput respectively for the four MAC state transition scenarios with FTP traffic. For the timer-based state transition algorithm, a two-state MAC state machine has better performance than a three-state MAC state machine. This is because the FTP traffic is not bursty. A user in the Control-hold state is unlikely to transition back to the Active state due to the long reading time with mean of 180 seconds. Therefore, an intermediate state between Active state and Dormant state does not provide any benefit. On the contrary, with the Control-hold state, the number of allowable Active state users is reduced from 32 to 19, thus reducing the scheduling efficiency.

For the priority based state transition algorithm, the performance of the two-state case is similar to that of the timer-based state transition algorithm. Although the priority based state transition algorithm enables flexible transition of users between the Active state and the Dormant state, it does not benefit the FTP traffic since the traffic is not bursty. The three-state priority based state transition, on the other hand, significantly outperforms the two-state priority based state transition as well as the two-state and three-state timer-based state transition. With priority based state transition, users can be flexibly moved between the Active state and the Control-hold state. With a small state transition latency between the Control-hold state and the Active state, users in these two states form a virtual Active user pool that the scheduler can effectively schedule. With three states, the maximum number of users in the virtual Active user pool is 50 compared to two states case where the maximum number of users in the Active user pool is 32. Therefore, the three-state priority based state transition case has improved scheduling efficiency, thus

the superior performance. Overall, the three-state priority based state transition provides 48% user capacity gain over the best case (i.e. two-state case) timer-based state transition.

8.2.2.2 HTTP Traffic

The user capacity and sector throughput performance of the four MAC state transition scenarios with HTTP traffic are shown in Figure 51 and Figure 52 respectively. For the timer-based state transition algorithm, a two-state MAC state machine has better performance than a three-state MAC state machine. This observation is similar to the FTP case. Although the HTTP traffic is burstier than the FTP traffic, the reading time in between web-page download is still large with a mean of 30 seconds. During this time, a user in the Control-hold state will eventually transition to the Dormant state before having a chance to go back to the Active state. Although for the three-state case, T_{active} can be set to a very small value to aggressively move a user from the Active state to the Control-hold state once its buffer is temporary empty in between embedded web-page objects download, this doesn't provide performance benefit in the case of the timer-based state transition because once a user is moved out of the Active state, it has to wait in a first-come-first-served basis until there is available resource in the Active state before it can transition back to the Active state. Moreover, for the three-state case, the maximum number of Active state users that the scheduler can schedule is limited to 19.

For the case of the priority based state transition, the two-state case and the three-state case have similar performance and they significantly outperform the timer-based state transition. For the two-state case, $T_{timeout}$ of 200ms is chosen such that a user does not transition out of the Active state within a packet call or during web-page download, but is aggressively moved into the Dormant state once the packet call is completed, i.e. when its

buffer is empty for more than 200ms. This makes room for other higher priority users in the Dormant state that have data in their buffer to enter the Active state. This flexibility in state transition improves the scheduling efficiency and users' packet call throughput. For the three-state case, although the larger number of users in the virtual Active user pool can improve the scheduling efficiency, it does not provide additional performance benefit over the two-state case. This is because with the bursty nature of the HTTP traffic, the maximum allowable number of users in the virtual Active user pool, i.e. 50, is more than sufficient to support the number of simultaneous users in a sector that have data in their buffer. Overall, the priority based state transition provides 18% user capacity gain over the best-case (i.e. two-state case) timer-based state transition.

8.2.2.3 WAP Traffic

Figure 53 and Figure 54 show the user capacity performance and sector throughput performance respectively for the different MAC state transition scenarios with WAP traffic. Overall, we observe that the three-state case is better than the two-state case for a particular state transition algorithm. This is because the WAP traffic is much burstier than FTP and HTTP traffic, with small packet sizes and mean inter-object arrival time, gateway response time and reading time range from 1.6 seconds to 5.5 seconds. In this case, the Control-hold state serves as an intermediate state that allows user to reside when its buffer becomes temporarily empty. For the timer-based state transition, since the packet size of each WAP object is small, a user does not stay in the Active state for a long duration thus allowing more dynamic transitioning of users between the Active state and the Control-hold state. This is contrary to the situation in FTP and HTTP traffic.

Therefore the three-state timer-based scenario has a 40% capacity gain over the two-state timer-based scenario.

For the priority based state transition, the three-state case outperforms the two-state case by 50% in terms of user capacity and sector throughput. This is because of the larger number of users in the virtual Active user pool that enhances scheduling efficiency. For the same MAC state configuration, the priority based state transition provides 33% and 46% capacity gain over the timer-based state transition, for the two-state case and the three-state case respectively.

8.2.2.4 Mixed FTP/HTTP/WAP Traffic

The user capacity performance and sector throughput performance for mixed traffic are shown in Figure 55 and Figure 56 respectively. The performance of each MAC state scenario is skewed by the percentage of each type of traffic. Overall, the best case (i.e. three-state case) priority based state transition provides 41% user capacity gain over the best case (i.e. three-state case) timer-based state transition. For both the timer-based state transition algorithm and the priority based state transition algorithm, the three-state MAC state machine outperforms the two-state MAC state machine by 29% and 39% respectively in terms of user capacity.

8.2.3 Comparison of State Transition Frequencies

As described in Section 8.2.2, the proposed priority based state transition algorithm significantly outperforms the conventional timer-based state transition algorithm in terms of system capacity. However, it is expected that the priority based state transition will incur higher state transition frequency thus higher signalling overhead. In this section, we examine the state transition frequency of the priority based state transition algorithm and

compare with that of the timer-based state transition algorithm. We perform the study using the three-state scenario. The performance is evaluated at the capacity point of Scenarios 1B and 2B, i.e., 66 users and 93 users respectively as previously shown in Section 8.2.2.

Figure 57 shows the rate for the different MAC state transition, i.e., Dormant to Active, Control-hold to Dormant, Active to Control-hold and Control-hold to Active, for the three-state timer-based state transition case and the three-state priority based state transition case. We can see that the state transition rate of the priority based state transition algorithm is generally higher than that of the timer-based state transition algorithm. There is only a small difference in the transition rate of Dormant to Active and Control-hold to Dormant between the two algorithms. On the other hand, the priority based algorithm has about 55% higher state transition rate between the Control-hold state and the Active state compared to the timer-based algorithm. This is because of the much more dynamic transition of users between the Active and Control-hold states based on users' priority.

The higher the state transition rate, the higher the signalling overhead associated with state transition. We therefore reduce the state transition priority update rate with the objective of reducing the state transition rate. Figure 58 shows the state transition rate for the case of state transition priority update rate of 8Hz and compares with that of the baseline 800Hz case. We can see the state transition rate between the Active state and the Control-hold state is greatly reduced with the reduced state transition priority update rate. The state transition rate between the Active state and the Control-hold state is now comparable to that of the timer-based state transition algorithm. However, the state

transition rates for Dormant to Active and Control-hold to Dormant have slightly increased with the reduced state transition priority update rate. This is because a user in the Control-hold state has a higher probability of transitioning to the Dormant state due to the reduced probability of transitioning to the Active state.

We further evaluate the impact of reducing state transition priority update rate on the system performance. As shown in Figure 59, reducing the state transition priority update rate from 800Hz to 8Hz does not result in a significant system performance degradation. Overall, the reduced state transition priority update rate of 8Hz provides a good trade off between signalling overhead and system performance.

8.3 Summary

In this chapter, we proposed a novel priority based state transition algorithm for both the existing 3G cellular system as well as the proposed B3G multi-mode wireless access systems. We also conducted detailed performance studies of the proposed priority based state transition algorithm and the conventional timer-based state transition algorithm under different MAC states scenarios. Our overall findings are as follow:

- The proposed priority based state transition algorithm significantly outperforms the conventional timer-based state transition algorithm for all the traffic types considered, i.e., FTP, HTTP, WAP and mixed traffic.
- For the conventional timer-based state transition algorithm, a two-state MAC state machine consisting of the Dormant state and the Active state is best suited for FTP and HTTP traffic. A three-state MAC state machine consisting of the Dormant state, the Control-hold state and the Active state is best suited for the WAP traffic.

- For the proposed priority based state transition algorithm, a three-state MAC state machine is best suited for all the traffic types considered.
- For the proposed priority based state transition algorithm, a state transition priority update rate of 8Hz is adequate to provide a good trade-off between system performance and signalling overhead associated with state transition.

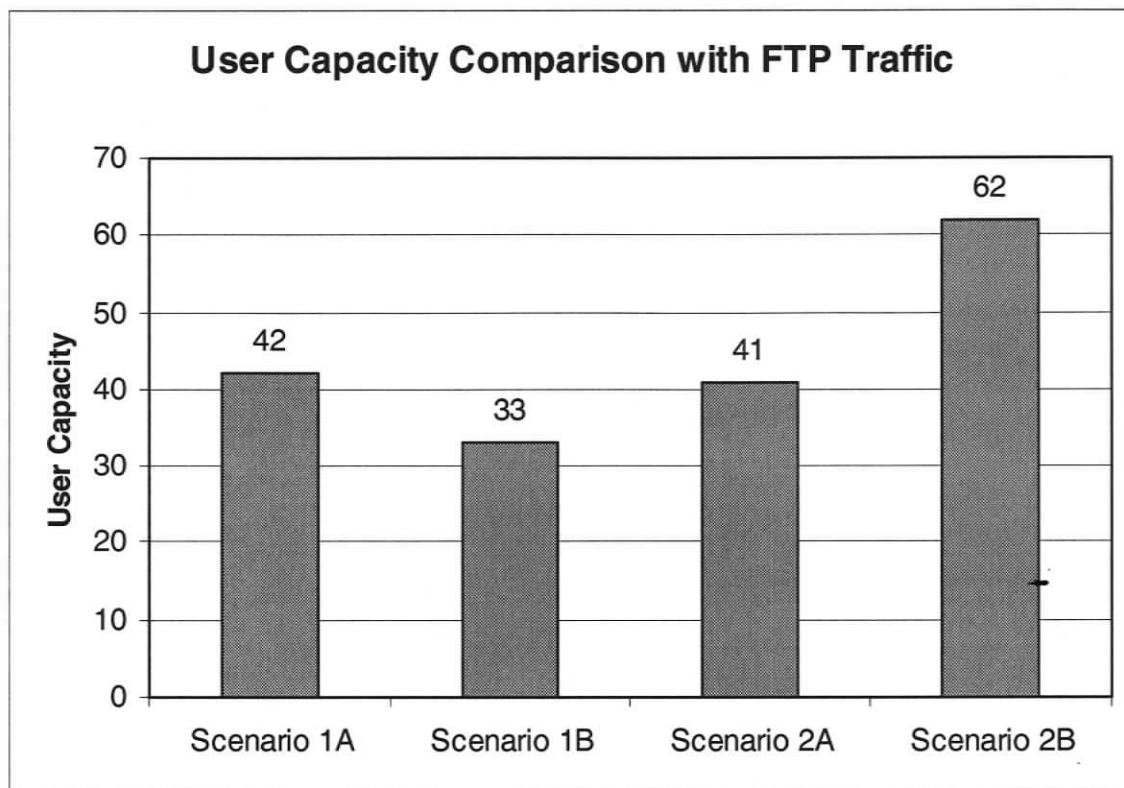


Figure 49 Comparison of User Capacity for Different MAC States Transition Algorithm Scenarios with FTP Traffic

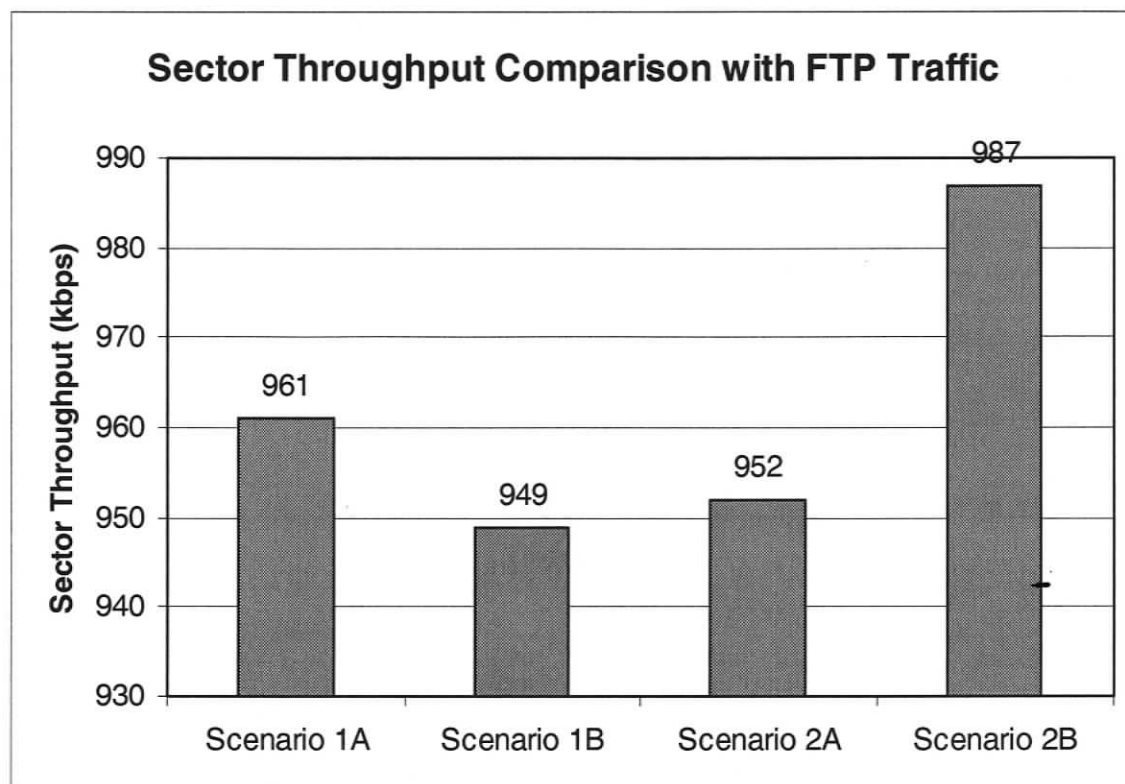


Figure 50 Comparison of Sector Throughput for Different MAC States Transition Algorithm Scenarios with FTP Traffic

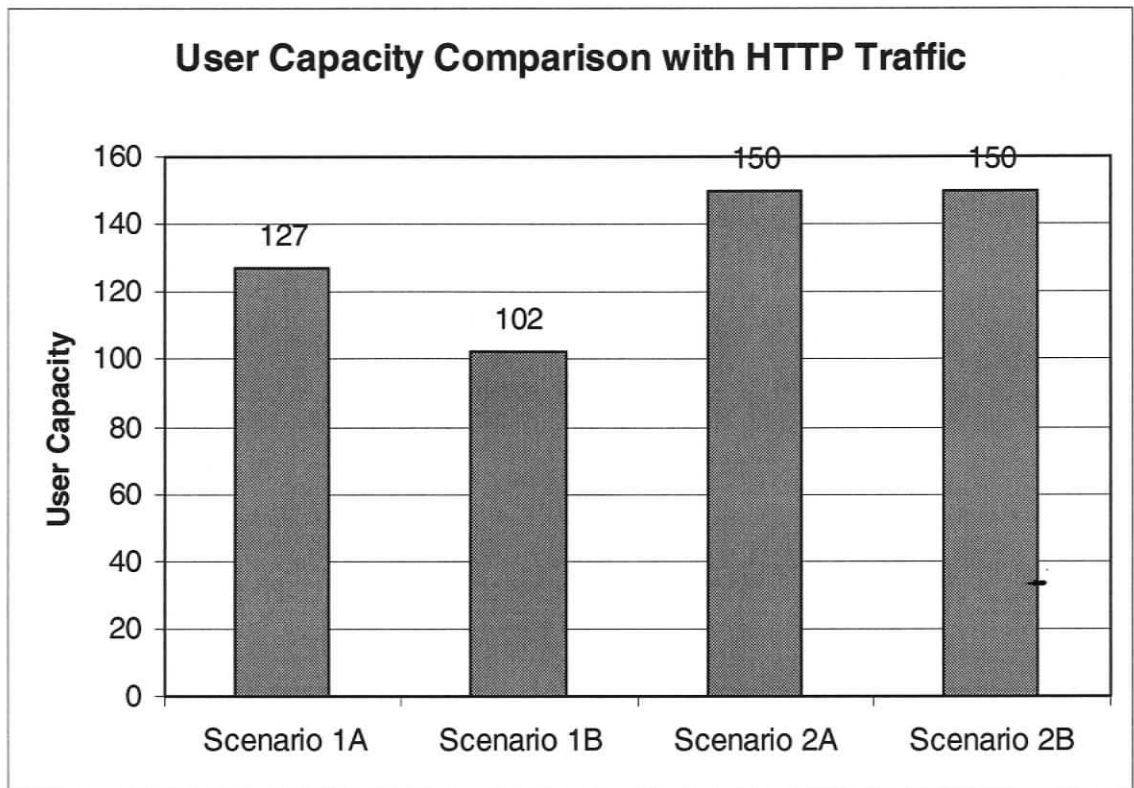


Figure 51 Comparison of User Capacity for Different MAC States Transition Algorithm Scenarios with HTTP Traffic

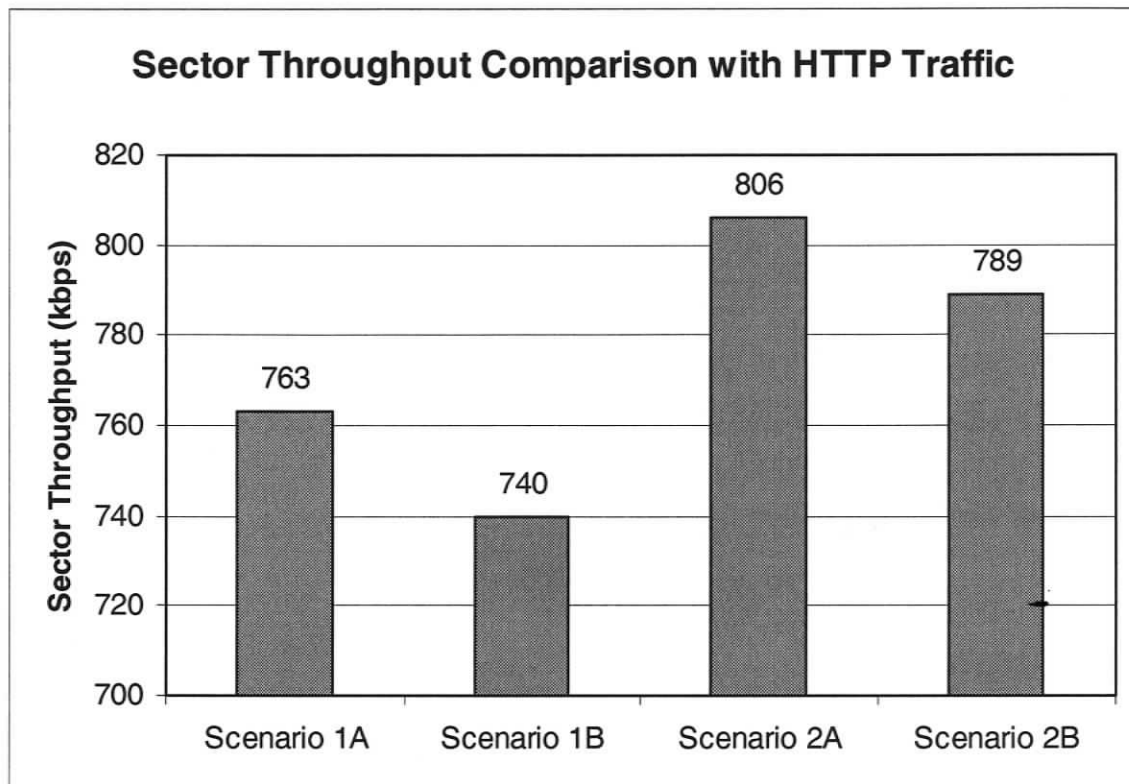


Figure 52 Comparison of Sector Throughput for Different MAC States Transition Algorithm Scenarios with HTTP Traffic

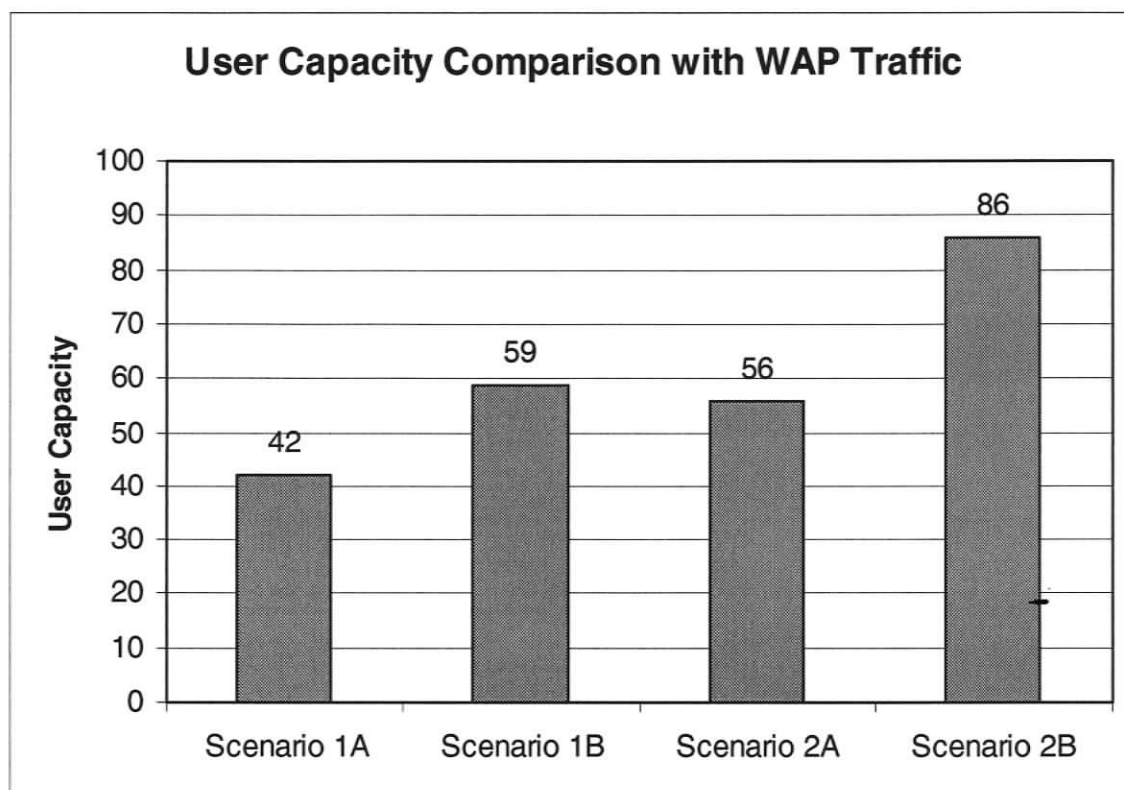


Figure 53 Comparison of User Capacity for Different MAC States Transition Algorithm Scenarios with WAP Traffic

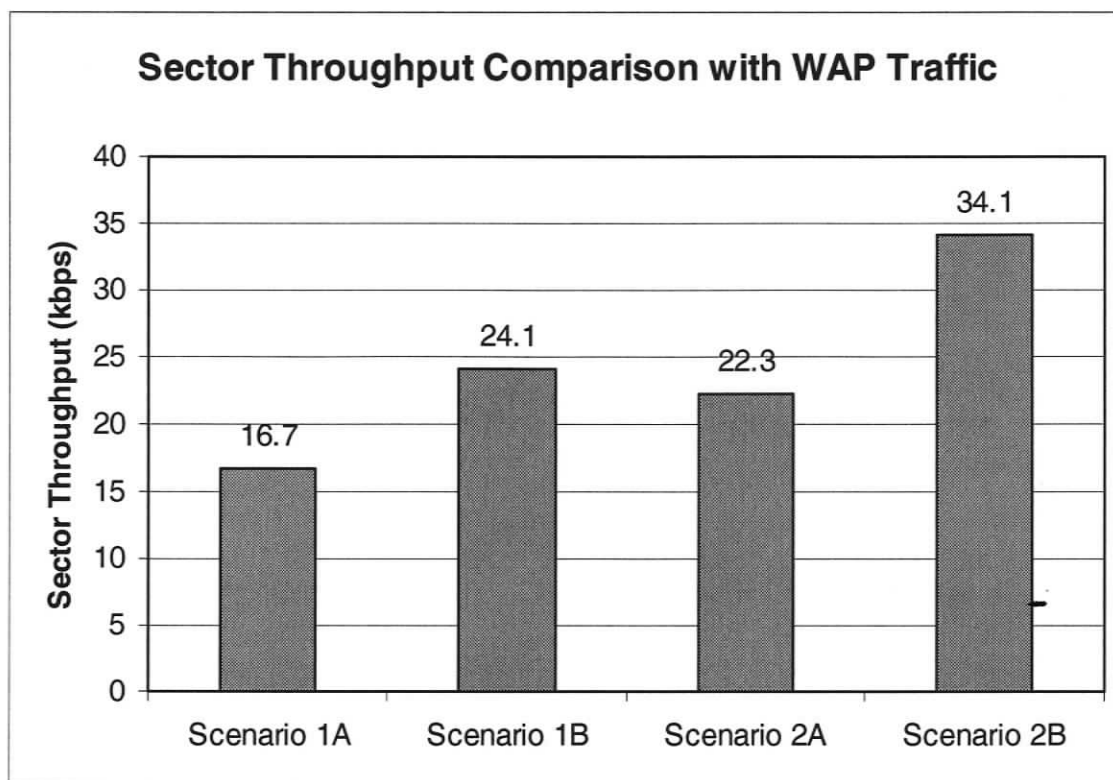


Figure 54 Comparison of Sector Throughput for Different MAC States Transition Algorithm Scenarios with WAP Traffic

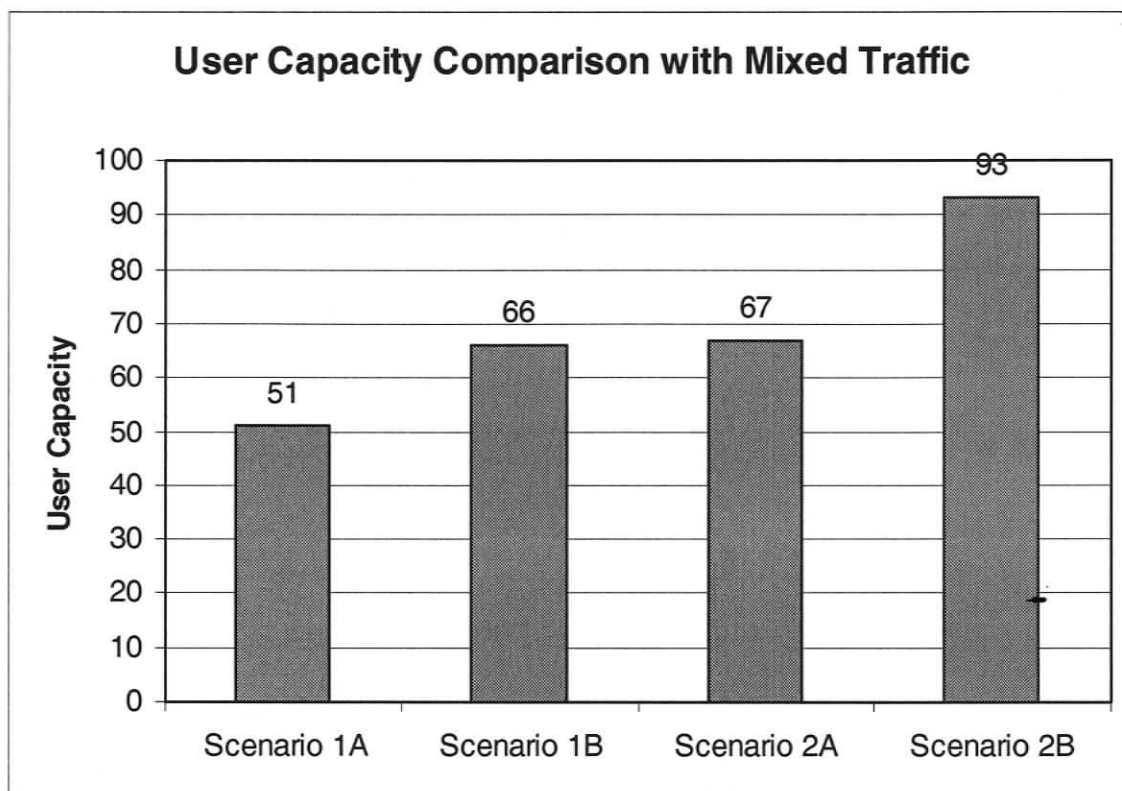


Figure 55 Comparison of User Capacity for Different MAC States Transition Algorithm Scenarios with Mixed FTP/HTTP/WAP Traffic

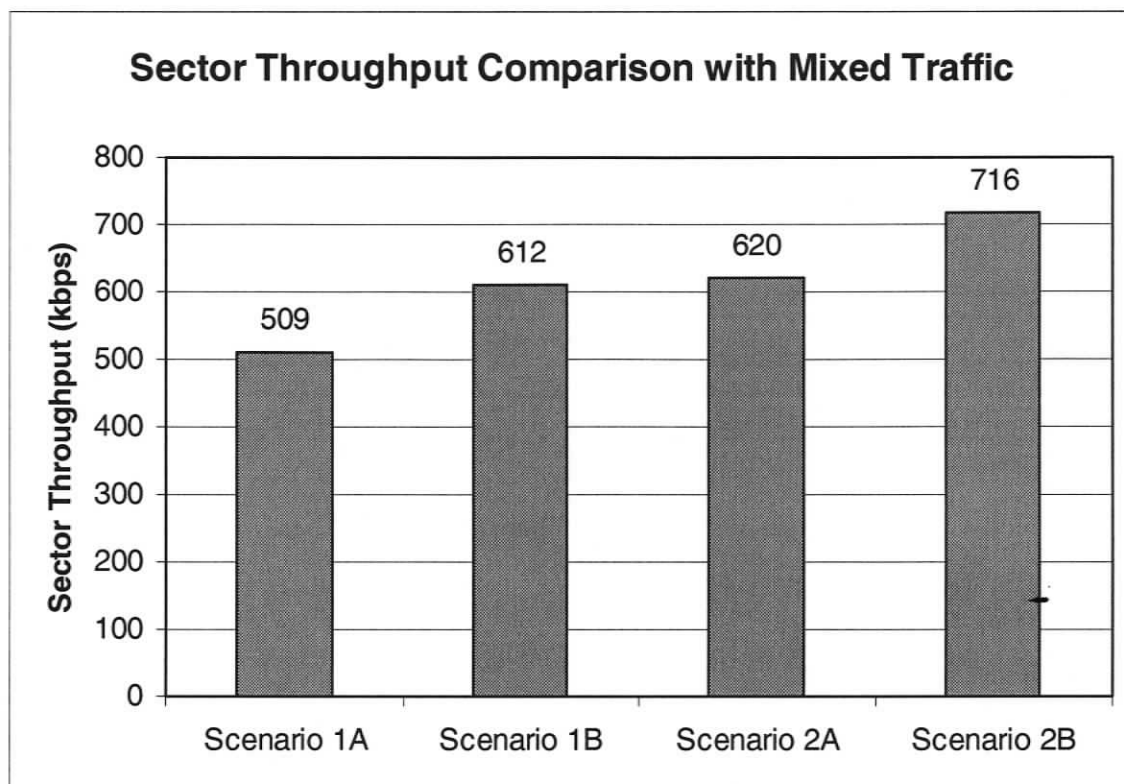


Figure 56 Comparison of Sector Throughput for Different MAC States Transition Algorithm Scenarios with Mixed FTP/HTTP/WAP Traffic

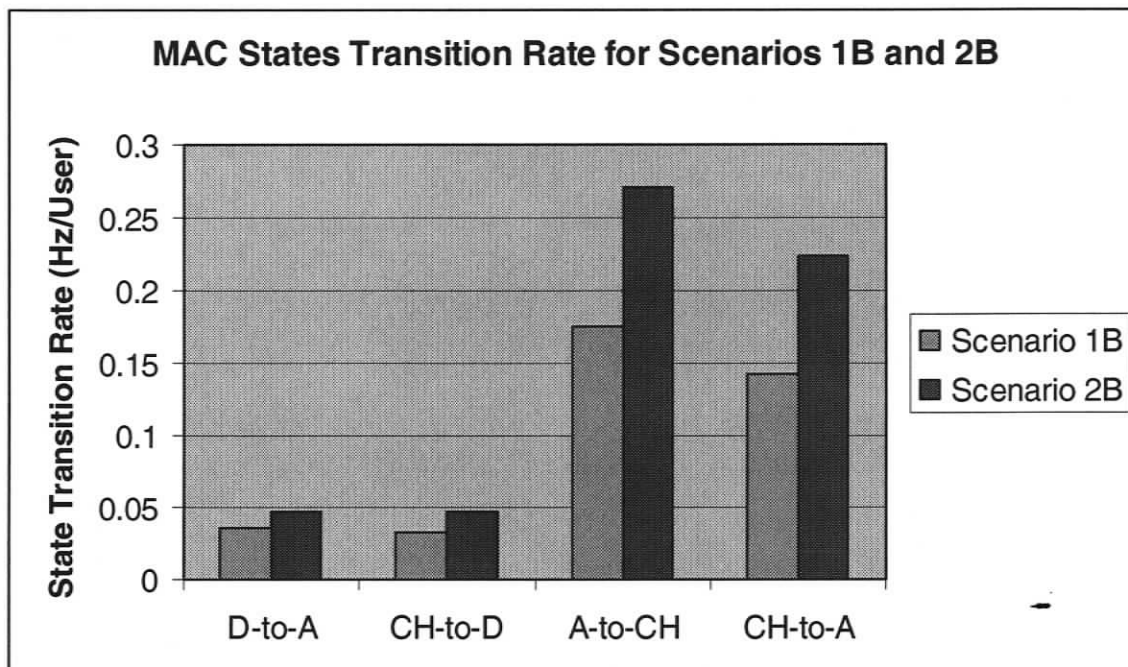


Figure 57 Comparison of MAC States Transition Rate for Scenario 1B (Three-State Timer-Based State Transition) and Scenario 2B (Three-State Priority Based State Transition), where D-to-A is Dormant to Active Transition, CH-to-D is Control-hold to Active Transition, A-to-CH is Active to Control Hold Transition, CH-to-A is Control-hold to Active Transition.

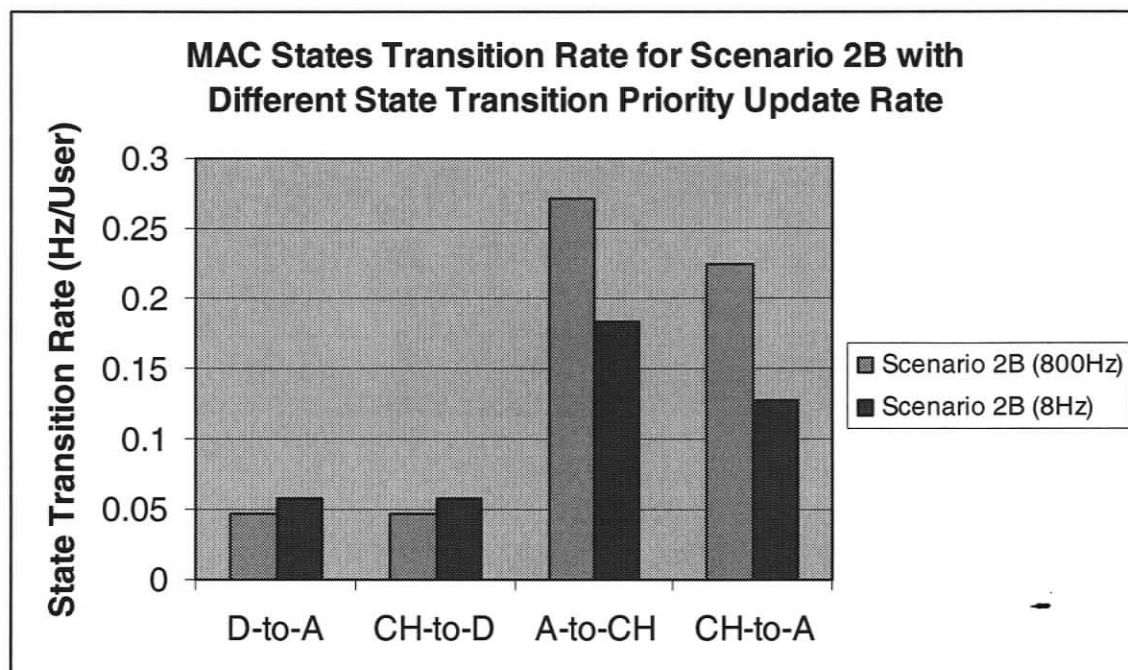


Figure 58 Comparison of MAC States Transition Rate for Scenario 2B (Three-State Priority Based State Transition) with State Transition Priority Update Rate of 800Hz and 8Hz.

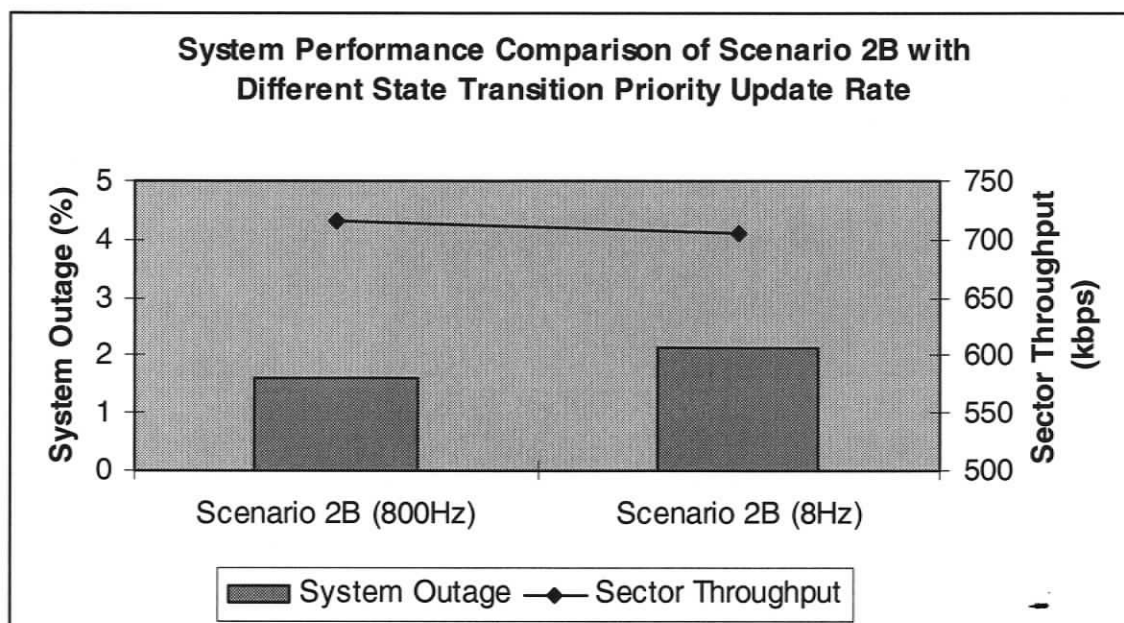


Figure 59 Comparison of System Outage and Sector Throughput for Scenario 2B (Three-State Priority Based State Transition) with State Transition Priority Update Rate of 800Hz and 8Hz.

Chapter 9

Conclusion

In this thesis, we investigated and proposed MAC solutions for B3G wireless access systems aimed to support different deployment scenarios; and encompass different access technologies while maintaining seamless access and mobility from a user's perspective.

We first proposed a novel B3G multi-mode access system framework that comprised of heterogeneous PL modes targeted for different deployment scenarios and performance targets, while anchored by a common layer 2/3 protocol stack. Each PL mode is optimized for specific mobility conditions, network topology, service types, and backward compatibility requirements to existing 3G systems. The proposed framework is superior to existing approaches of mere integration of heterogeneous access systems through IP layer or network layer convergence, since the common layer 2/3 protocols enable seamless mobility and handoff for users moving from one deployment environment to another. The common layer 2/3 also facilitates dynamic radio resource/load/spectrum management across the different PL modes to achieve optimum spectrum efficiency and QoS support. The proposed B3G multi-mode access system concept is also complimentary to the IP layer convergence approach which is required when the B3G access system inter-works with existing 2G/3G heterogeneous access systems. We provided detailed design of the common layer 2/3 protocol structure and operations; mobility management across PL modes and across different cell hierarchies; spectrum overlay options and system access. Also, we provided a specific multi-mode system configuration that overlays CDMA PL mode with OFDMA PL mode.

We further addressed the issues of system migration from 3G to B3G. We introduced the MC DS-CDMA system concept as the first evolution step of the current 3G cellular systems towards the proposed B3G multi-mode wireless access system framework. We provided a detailed description of the proposed MC DS-CDMA system including the system configuration, operation and the service-driven common layer 2/3 protocol stack that anchors the multiple DS-CDMA carriers. A cross-carrier layer 2 scheduler was proposed to optimize system capacity while ensuring the QoS of different services are met. The proposed cross-carrier scheduler can be generalized to cross PL modes scheduler as shown in Chapter 6. Detailed performance studies were conducted to evaluate the performance of a specific MC DS-CDMA system, i.e. MC-DV, and compared with that of the corresponding single carrier system, i.e. 1xEV-DV. Our studies showed that MC-DV provides significant performance and capacity gain over 1xEV-DV, in particular for bursty traffic and delay sensitive traffic which are key traffic characteristics in 3G and B3G wireless systems. The capacity gain was shown to be as high as 2.7 times for real-time service.

We closely examined the different load balancing schemes across the heterogeneous PL modes in the proposed B3G multi-mode access system. The PL modes share the spectrum resource in TDM, FDM or CDM fashion. We presented specific system configurations for all three cases of spectrum multiplexing. For cases where PL modes are either TDM or FDM, we proposed an integrated load balancing and scheduling or ILBS scheme to maximize the overall system capacity while ensuring users' QoS and SLA are met. For cases where PL modes are CDM, we proposed a dynamic Walsh codes and BS transmit power sharing scheme between power-controlled dedicated traffic

channels and rate-controlled packet data channels across multiple CDMA carriers. Performance studies showed that our proposed schemes significantly outperformed the conventional load balancing schemes. For the TDM and FDM cases, the proposed dynamic load balancing using ILBS provides a system outage reduction of a few times to tens of times compared to the conventional semi-static load balancing. The corresponding capacity gain for VoIP at 2% system outage criterion ranges from 12.5% to 2 times for the cases studied. For the case of CDM, the sector throughput gain of the proposed dynamic load balancing of voice and data in the same carrier versus semi-static segregation of voice and data into separate carriers is 12% to 24%. With multi-carrier scheduling coupled with the dynamic load balancing, a further sector throughput gain of 11% and corresponding FTP user capacity gain of 43% to 47% were shown.

We proposed a universal MAC states concept for the B3G multi-mode access system. The universal MAC state machine anchors the heterogeneous PL modes such that when a user switches from one PL mode to another, the MAC state and the associated context information of the MS can be retained in order to minimize packet loss and PL mode switching/handoff latency. We demonstrated the importance of retaining the current MAC state during PL mode switching by showing the impact of state transition latency and MAC states on overall system performance. Our performance studies showed that with retention of intermediate MAC states, the system performance can be improved by 43% and 34% respectively in terms of user capacity and sector throughput.

We investigated in detail the decision criteria used to transition a user from one MAC state to another. It is an important aspect of MAC states management for both the existing 3G systems and the B3G systems. The decision criteria should aim to maximize system

capacity while meeting users' QoS and SLA requirements, and at the same time achieve power-saving. We proposed a novel priority based state transition algorithm that achieves these objectives. We compared the performance of our proposed scheme with that of the conventional timer-based state transition algorithm. We showed that the priority based state transition algorithm outperformed the conventional algorithm by 18% to 48% in term of user capacity for the different traffic types considered. We also provided insight into the optimum MAC state machine configuration to support different traffic types. For the conventional timer-based state transition algorithm, a two-state MAC state machine consisting of the Dormant state with fast call setup (or the Suspended state) and the Active state is recommended for FTP and HTTP traffic. A three-state MAC state machine consisting of the Dormant state with fast call setup (or the Suspended state), the Control-hold state and the Active state is recommended for WAP traffic or other types of low rate, highly bursty traffic. For the proposed priority based state transition algorithm, a three-state MAC state machine consisting of the Dormant state with fast call setup (or the Suspended state), the Control-hold state and the Active state is recommended for all traffic types.

Overall, in this thesis, we provided key MAC solutions and system framework proposal for B3G systems envisioned in ITU-R M.1645 recommendation. We aim to contribute to and influence the research community on the direction of B3G system design. Our work has been published in [46] and [47], and has been submitted for review in [48].

9.1 Suggestions for Future Work

We suggest continuing to build on the proposed B3G multi-mode access system framework as follows:

1. Design of detailed signalling messages and signalling flow for system access and handoff between PL modes.
2. Design of physical layer configurations of the different PL modes. Physical layer configurations of existing system such as 1xEV-DO, HSDPA, 802.16e etc, can be used with further enhancements to improve spectral efficiency.
3. Detailed investigation of multi-hop relay as part of the multi-mode access system. Application of some of the proposed MAC solutions on resource/load/spectrum management to relay.
4. Incorporation of other types of embedded networks into the multi-mode access system framework such as sensor network, adhoc network and peer-to-peer communication.

Bibliography

- [1] TIA/EIA-856-A, cdma2000 High Rate Packet Data Air Interface Specification, Revision A, March 2004.
- [2] 3GPP TR 25.855 V5.0.0, Technical Specification Group Radio Access Network; High Speed Downlink Packet Access; Overall UTRAN Description (Release 5), September 2001.
- [3] ITU-R M.1645, "Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000," June, 2003.
- [4] M. Inoue, K. Mahmud, H. Murakami, M. Hasegawa, "MIRAI: A solution to seamless access in heterogeneous wireless networks," Proceedings of IEEE International Conference on Communications, vol. 2, May 2003, pp. 1033-1037.
- [5] C. Politis, T. Oda, S. Dixit, A. Schieder, H.-Y. Lach, M. I. Smirnov, S. Uskela, R. Tafazolli, "Cooperative networks for the future wireless world," IEEE Communication Magazine, vol. 42, no. 9, September 2004, pp.70-79.
- [6] J. McNair, F. Zhu, "Vertical handoffs in fourth-generation multinet network environments," IEEE Wireless Communication Magazine, vol. 11, issue 3, June 2004, pp. 8 – 15.
- [7] 3GPP TS 05.01 V8.9.0, Technical Specification Group GSM/EDGE Radio Access Network; Physical layer on the radio path; General description (Release 1999), November, 2004.
- [8] TIA/EIA/IS-136, TDMA Air Interface - Mobile Station - Base Station Compatibility, 1995.
- [9] TIA/EIA-2000-0, CDMA 2000® Series of Standard for Spread Spectrum Systems, Release 0 Addendum 2, June 2001.
- [10] TIA/EIA-2000-A, CDMA 2000® Series of Standard for Spread Spectrum Systems, Release A Addendum 2, February 2002.
- [11] TIA/EIA/IS-2000-B, CDMA 2000® Series of Standard for Spread Spectrum Systems, Release B, April 2002.
- [12] TIA-2000-C, CDMA 2000® Series of Standard for Spread Spectrum Systems, Release C, May 2002.
- [13] TIA-2000-D, CDMA 2000® Series of Standard for Spread Spectrum Systems, Release D, March 2004.

- [14] 3GPP TS 25.201 V3.4.0, Technical Specification Group Radio Access Network; Physical layer - General description (Release 1999), June 2006.
- [15] IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005, IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1, 2005.
- [16] TIA/EIA-95, Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System, 1995.
- [17] TIA/EIA-95 Rev A, Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System, 1995.
- [18] TIA/EIA-95 Rev B, Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System, 1997.
- [19] L. Litwin, "The Medium Access Control Sublayer," IEEE Potentials, vol. 20, no. 4, pp. 30 – 34, October – November 2001.
- [20] A. S. Tanenbaum, *Computer Networks*, Prentice-Hall International, Inc., Second Edition, 1989.
- [21] IEEE 802.11A-1999, IEEE Standard for IT-Telecommunications and information exchange between systems LAN/MAN - Part II: Wireless LAN Medium Access control (MAC) and Physical Layer (PHY) Specification: High-speed Physical Layer in the 5GHz Band, 1999.
- [22] IEEE 802.11B/COR 1-2001, IEEE Standard for IT-Telecommunications and information exchange between systems LAN/MAN - Part II: Wireless LAN Medium Access control (MAC) and Physical Layer (PHY) Specification: High-speed Physical Layer Extension in the 2.4GHz Band, 2001.
- [23] IEEE 802.11G-2003, IEEE Standard for IT-Telecommunications and information exchange between systems LAN/MAN - Part II: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band, 2003.
- [24] IEEE 802.11E-2005, IEEE Standard for IT-Telecommunications and information exchange between systems LAN/MAN – Part II: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements, 2005.
- [25] Y. Xiao, "IEEE 802.11E: QoS provisioning at the MAC layer," IEEE Wireless Communication Magazine, vol. 11, no. 3, pp. 72-79, June 2004.

- [26] J. W. Robinson, T. S. Randhawa, "Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function," *IEEE Journal of Selected Areas in Communications*, vol. 22, no. 5, pp. 917 – 928, June 2004.
- [27] 3GPP2 C.R1002-0, *cdma2000 Evaluation Methodology*, Revision 0, December 2004.
- [28] 3GPP TR 25.892 V6.0.0, Technical Specification Group Radio Access Network; Feasibility Study for Orthogonal Frequency Division Multiplexing (OFDM) for UTRAN enhancement (Release 6), June 2004.
- [29] IEEE P 802.20 PD-09 Version 1.0, "802.20 Evaluation Criteria – Ver. 1.0," September, 2005.
- [30] 3GPP2 C.S0025-0, Markov Service Option (MSO) for cdma2000 Spread Spectrum Systems, Ballot Resolution Version, March, 2000.
- [31] 3GPP2 C.S0014-0, *Enhanced Variable Rate Codec (EVRC)*, Version 1.0, December 1999.
- [32] ITU-T G.114, "One Way Transmission Time," May 2003.
- [33] E. Yanmaz, O. K. Tonguz, "Dynamic load balancing and sharing performance of integrated wireless networks," *IEEE Journal of Selected Areas in Communications*, vol. 22, no. 5, pp. 862 – 872, June 2004.
- [34] Y. Matsunaga, R. H. Katz, "Inter-domain radio resource management for wireless LANs," *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2183 – 2188, March 2004.
- [35] G. Bianchi, I. Tinnirello, "Improving load balancing mechanisms in wireless packet networks communications," *Proceedings of IEEE International Conference on Communications (ICC)*, pp. 891 – 895, April/May 2002.
- [36] X. Liu, E. K. P. Chong, N. B. Shroff, "Optimal opportunistic scheduling in wireless networks," *Proceedings of IEEE Vehicular Technology Conference 2003 - Fall (VTC2003-Fall)*, pp. 1417-1421, September 2003.
- [37] A. Jalali, R. Padovani, R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," *Proceedings of IEEE Vehicular Technology Conference 2000 - Spring (VTC2000-Spring)*, vol. 3, pp. 1854-1858, May 2000.
- [38] S. Vrzić, M. –H. Fong, J. Li, "Fairness and delay bound scheduler," *Nortel Networks Internal Report*, August 2001.
- [39] R. Steele, *Mobile Radio Communications*, Pentech Press Publishers – London, IEEE Press New York, 1992.

- [40] ITU-R M.1225, "Guidelines for evaluation of radio transmission technologies for IMT-2000," 1997.
- [41] M. Chatterjee, S. K. Das, "Optimal MAC State Switching for cdma2000 Networks," Proceedings of 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002), pp. 400 - 406, June 2002.
- [42] F. A. D. Ferrer, C. V. Rosales, "Performance Analysis of cdma2000 MAC Layer with Voice/Data Integration," Proceedings of the 14th International Conference on Electronics, Communications and Computers (CONIELECOMP '04), pp. 35 – 40, February 2004.
- [43] Nortel Networks, "RL voice only results," 3GPP2 TSG-C WG3 contribution C30-20021106-003R1, November 2002.
- [44] Qualcomm, "Motivation of the differential C/I feedback," 3GPP2 TSG-C WG5 contribution C50-20011120-011, November 2001.
- [45] S. Lee; S.-Y. Yoon; S.-J. Maeng, "Air-link performance of the cdma2000 based mobile wireless network," Proceedings of IEEE Global Telecommunications Conference 2001 (GLOBECOM '01), pp. 3405 - 3428, November 2001.
- [46] M.-H. Fong, G. Wu, W. Tong, J. Li, A. Gulliver, V. K. Bhargava, "Dynamic and scalable bandwidth allocation for beyond 3G CDMA systems," Proceedings of 5th International Workshop on Multi-Carrier Spread Spectrum, September 2005.
- [47] M. -H. Fong, T.A. Gulliver, V. K. Bhargava, "Multi-mode access system with anchor layer 2/3 protocol for beyond 3G wireless networks," Proceedings of IEEE Vehicular Technology Conference 2006 - Spring (VTC2006-Spring), May 2006.
- [48] M. -H. Fong, T. A. Gulliver, V. K. Bhargava, "Load balancing for multi-mode access systems in beyond 3G wireless networks," submitted to IEEE Communications Letters.

Appendix A

Calculation of Number of Active State and Control-hold State Users

In this appendix, we provide the detailed calculation conducted to obtain the number of Active state and Control-hold state users assumed in Chapter 7 and Chapter 8.

We use the number of voice users supported in the RL of 1xEV-DV as the basis for the calculation. In [43], the maximum number of simultaneous voice users supported per sector in the RL is shown to be 42, based on a full RL system-level simulation, the per-user outage criterion and the ROT criterion defined in [27]. The aggregate RL capacity can be expressed as

$$(28) \quad C_{RL} = 42 \times (1 + (T/P)_{eff_voice}),$$

where $(T/P)_{eff_voice}$ is the effective traffic-to-pilot ratio (TPR) for a voice user.

$(T/P)_{eff_voice}$ is calculated as

$$(29) \quad (T/P)_{eff_voice} = P_{full} \times (T/P)_{full} + P_{1/2} \times (T/P)_{1/2} + P_{1/4} \times (T/P)_{1/4} + P_{1/8} \times (T/P)_{1/8},$$

where P_{full} , $P_{1/2}$, $P_{1/4}$ and $P_{1/8}$ are the probability of occurrence of full rate, half rate, quarter rate and eighth rate voice frames respectively, as given in Section 3.2.6.

$(T/P)_{full}$, $(T/P)_{1/2}$, $(T/P)_{1/4}$ and $(T/P)_{1/8}$ are the TPRs of Reverse Fundamental Channel (R-FCH) for full rate, half rate, quarter rate and eighth rate frames respectively.

These TPRs are defined in [12] and are summarized in Table 12 below.

Table 12 TPR for R-FCH

Frame type	T/P (dB)
Full rate	3.75
Half rate	-0.25
Quarter rate	-2.25
Eighth rate	-5.875

Therefore, we obtain that

$$(30) \quad C_{RL} = 81.8 .$$

The maximum number of Active state supported per sector, N_{\max_active} , is limited by the RL since the RL pilot channel (R-PICH) and the RL channel quality indicator channel (R-CQICH) are always on regardless of whether an Active state user has data to transmit on the FL or RL. We next calculate the maximum number of Active state data users per sector by assuming an Active state user has a 50% activity factor on the R-FCH, and the loading of R-CQICH can be represented by a CQICH-to-pilot ratio (C/P) of -4.3dB as given in [27]. We have

$$(31) \quad C_{RL} = N_{\max_active} \times (1 + 0.5 \cdot (T/P)_{full} + (C/P)_{active}) .$$

Using the values of $(T/P)_{full}$ and $(C/P)_{active}$ previously given, we obtain

$$(32) \quad N_{\max_active} \approx 32 .$$

With the presence of Control-hold state users, the RL capacity is shared between the Active state users and the Control-hold state users. A Control-hold state user transmits gated R-PICH and gated R-CQICH. We assume a gating rate of $\frac{1}{4}$, i.e. the R-PICH and the R-CQICH are transmitted once every four slots. Ideally with R-PICH transmitted at $\frac{1}{4}$

gating rate, the RL interference generated by the R-PICH is $\frac{1}{4}$ or -6dB. However, as shown in [45], due to the reduced power control rate, the required R-PICH received E_c/N_t increases, which leads to a RL interference reduction of -2.3dB or 0.589 compared to the case of non-gated R-PICH. For the gated R-CQICH transmission, we assume full CQI feedback for the gated-on slot, rather than a mixed of full and differential CQI feedback assumed for Active state users in [27]. This is to ensure the $\frac{1}{4}$ rate CQI feedback can reliably track the fast fading of high mobility users. The C/P of a full CQI transmission is 3dB as shown in [44]. Therefore, the average C/P for a user in Control-hold state is (3-6)dB = -3dB or 0.5. The RL capacity shared by the Active state users and the Control-hold state users can be expressed as

$$(33) \quad C_{RL} = N_{active} \times (1 + 0.5 \cdot (T/P)_{full} + (C/P)_{active}) + N_{ch} \times ((P_{ch}/P) + (C/P)_{ch}),$$

where N_{active} is the number of Active state users, N_{ch} is the number of Control-hold state users, (P_{ch}/P) is the ratio of R-PICH transmission power in Control-hold state to the R-PICH transmission power in Active state and equals to 0.589, $(C/P)_{ch}$ is the C/P in Control-hold state and equals to 0.5. Therefore, for a given N_{active} , we can obtain the corresponding N_{ch} . The different combinations of N_{active} and N_{ch} are summarized in Table 13.

In our performance studies in Chapter 7 and Chapter 8, we have chosen $N_{active} = 19$ and $N_{ch} = 31$ to provide a good balance between the number of Active state users and the total number of virtual Active users.

Table 13 **Different Combinations of Number of Active State Users and Number of Control-hold State Users**

N_{active}	N_{ch}	Total number of Virtual Active Users
0	75	75
5	63	68
10	52	62
15	40	55
19	31	50
20	28	48
30	5	35
32	0	32

VITA

Surname: FONG

Given Names: Mo-Han

Place of Birth: Kuala Lumpur, Malaysia

Date of Birth: September 2, 1971

Educational Institutions Attended:

University of Victoria, Canada 1993 to 2006

Nanyang Technological University, Singapore 1989 to 1993

Degrees Awarded:

M. A. Sc. University of Victoria, 1995

B. Eng. Nanyang Technological University, 1993

Honours and Awards:

Nortel Networks Second Quarter 2005 Technology Award of Excellence (2005)

University of Victoria Fellowship (1993 – 1995)

University of Victoria Research Assistantship (1993 – 1995)

Canadian Advanced Technology Association (CATA) Scholarship for excellent research in the discipline of telecommunications (1995)

The Martlet Chapter IODE Graduate Scholarship for Women (1994/95)

The British Columbia Advanced Systems Institute (ASI) Graduate Assistance Scholarship (1993/94)

Association of Consulting Engineers Singapore Gold Medal for top academic standing (1993)

Hewlett-Packard Prize for distinguished academic performance (1993)

3 Motorola Book Prizes for top standing in 3 major subjects: Communication Systems, Microwave Engineering and Digital Communications (1993)

Gold Medal Award of the 4th Varsity Micro-mouse Contest from Nanyang Technological University, Singapore (1991)

Granted Patents:

1. H. Zhang, M. –H. Fong, S. Periyalwar, “Dynamic automatic retransmission request in wireless access networks,” United States Patent Number 7,032,153, April 2006.
2. M. –H. Fong, L. Strawczynski, G. Wu, W. Tong, “Dual protocol layer automatic retransmission request scheme for wireless air interface,” United States Patent number 6,931,569, August 2005.
3. M. –H. Fong, G. Wu, “Dynamic, dual-mode wireless network architecture with split layer 2 protocol,” China P.R. Patent number ZL00189036.4, May 2005.
4. M. –H. Fong, G. Wu, “Network resource sharing during handover of a mobile station between cellular wireless networks,” United States Patent number 6,799,039, September 2004.
5. M. –H. Fong, H. Zhang, G. Wu, D. Yu, A. Schmidt, “Automatic retransmission request layer interaction in a wireless network,” United States Patent number 6,760,860, July 2004.
6. N. Guo, M. –H. Fong, A. Gutierrez, W. Tong, “Sub-frames for cdma2000,” China P.R. Patent number 99804286.2, May 2004.
7. F. Khaleghi, N. Senarath, G. Wu, R. Matyas, P. Larijani, M. –H. Fong, “Signalling scheme and messaging structure to support the smoothing of large bit rate transmissions,” United States Patent number 6,704,328, March 2004.
8. W. Huang, H. Zhang, M. –H. Fong, “Network driven cell switching and handoff with load balancing for wireless systems,” United States Patent number 6,701,150, March 2004.
9. M. –H. Fong, C. –C. Wang, G. Wu, “Multiplexing/demultiplexing schemes between wireless physical layer and link layer,” United States Patent number 6,590,882, July 2003.
10. N. Ahmadvand, G. Wu, M. –H. Fong, “Data link control protocol for 3G wireless system,” United States Patent number 6,542,490, April 2003.
11. B. Massie, C. Leretaille-Gauthier, A. Lamani, W. Tong, M. –H. Fong, “Procédé et dispositifs de transmission de données avec mécanisme d’acquiescement harq restart –

- method and devices for transmitting data with acknowledgement,” France Patent number 2819661, March 2003.
12. M. –H. Fong, G. Wu, “Active set management in cellular wireless network that supports high data rate forward link transmissions,” United States Patent number 6,493,328, December 2002.
 13. N. Secord, M. –H. Fong, G. Wu, L. Strawczynski, A. Jalali, R. Matyas, N. Guo, W. Lee, A. Bigloo, “Systems and methods of channel coding and inverse-multiplexing for multi-carrier CDMA systems,” United States Patent number 6,373,831, April 2002.
 14. J. Li, M. –H. Fong, D. Yu, “Method and system for optimally allocating orthogonal codes,” United States Patent number 7,072,319, December 2001.
 15. M. –H. Fong, G. Wu, D. C. Bees, R. W. Dijkerman, “Circuit-mode data service support over the A-interface,” United States Patent number 6,259,684, July 2001.

Publications:

Journal

1. R. Hu, D. Paranchych, M. –H. Fong, G. Wu, “Handoff management with network architecture evolution in IEEE802 broadband wireless mobility networks,” submitted to IEEE Communication Magazine.
2. M. –H. Fong, T. A. Gulliver, V. K. Bhargava, “Load balancing for multi-mode access systems in beyond 3G wireless networks,” submitted to IEEE Communications Letters.
3. M. –H. Fong, V. K. Bhargava, Q. Wang, “Concatenated orthogonal/PN spreading sequences and their application to cellular DS-CDMA systems with integrated traffic,” IEEE Journal on Selected Areas in Communications, Special Issue on Wireless Local Communications, April/May, 1996.

Conference

4. M.-H. Fong, G. Wu, W. Tong, J. Li, A. Gulliver, V. K. Bhargava, "Dynamic and scalable bandwidth allocation for beyond 3G CDMA systems," Proceedings of 5th International Workshop on Multi-Carrier Spread Spectrum, September 2005.
5. M. -H. Fong, T.A. Gulliver, V. K. Bhargava, "Multi-mode access system with anchor layer 2/3 protocol for beyond 3G wireless networks," Proceedings of IEEE Vehicular Technology Conference 2006 - Spring (VTC2006-Spring), May 2006.
6. M. -H. Fong, V. K. Bhargava, Q. Wang, "Line rate(s) selection for cellular DS-CDMA systems with integrated traffic," Proceedings of IEEE Global Telecommunications Conference 1995 (Globecom '95), November 1995.
7. M. -H. Fong, V. K. Bhargava, Q. Wang, "Optimum signalling for cellular DS-CDMA systems with integrated traffic," Book of Abstracts of the Canadian Institute for Telecommunications Research (CITR) Annual Conference '95, August 1995.
8. M. -H. Fong, V. K. Bhargava, Q. Wang, "Concatenated orthogonal/PN spreading scheme for cellular DS-CDMA systems with integrated traffic," Proceedings of IEEE International Conference on Communications 1995 (ICC '95), pp. 905-909, June 1995.
9. M. -H. Fong, "Signalling for DS-CDMA systems with integrated traffic," presented at the British Columbia Advanced Systems Institute (ASI) Graduate Students Presentation Day '95, Vancouver, BC, March 1995.
10. M. -H. Fong, Q. Wang, V. K. Bhargava, "Concatenated orthogonal/PN cods for DS-CDMA systems in a multi-user and multipath fading environment," Proceedings of IEEE Global Telecommunications Conference 1994 (Globecom '94), pp. 1642-1646, Nov. 1994.
11. M. -H. Fong, V. K. Bhargava, Q. Wang, "Application of the concatenated orthogonal/PN spreading scheme for high rate transmission," Book of Abstract of the Canadian Institute for Telecommunications Research (CITR) Annual Conference '94, August 1994.
12. C. W. Chan, K. K. Low, C. G. Ee, M. -H. Fong, "Distributed staff directory system for a campus," Proceedings of IEEE SICON/ICIE '93 Singapore, September 1993.