Design and performance analysis of medium access control protocols for QOS in Wimax

Lu, Jin Chang

2010

Lu, J. C. (2010). Design and performance analysis of medium access control protocols for QOS in Wimax. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/41502>

<https://doi.org/10.32657/10356/41502>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**DESIGN AND PERFORMANCE ANALYSIS OF MEDIUM
ACCESS CONTROL PROTOCOLS FOR QOS IN WIMAX**

LU JINCHANG

**SCHOOL OF ELECTRICAL AND ELECTRONIC
ENGINEERING**

2010

DESIGN AND PERFORMANCE ANALYSIS OF MEDIUM
ACCESS CONTROL PROTOCOLS FOR QOS IN WIMAX

LU JINCHANG

2010

DESIGN AND PERFORMANCE ANALYSIS OF MEDIUM ACCESS CONTROL PROTOCOLS FOR QOS IN WIMAX

LU JINCHANG

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University in
partial fulfillment of the requirement for the degree of
Doctor of Philosophy in Communications Engineering

2010

Acknowledgements

Primarily, I wish to thank my supervisor, A/P Maode Ma. Without his support, guidance, and constant encouragement, this dissertation would not have been possible. I truly appreciate all the suggestions, professional advices and numerous great ideas given by him.

It is also a great pleasure to thank my fellow graduate students and friends. I have enjoyed many technical discussions with Zheng Jialing. I also wish to thank Frankie Chan for his technical help.

Last, but not least, my boundless gratitude is devoted to my family for their love and support during my graduate school career. The path to complete this dissertation was far from clear during some difficult times, and I would like to especially thank my wife for her understanding and supportive encouragement.

Table of Contents

ACKNOWLEDGEMENTS	I
TABLE OF CONTENTS	III
LIST OF ABBREVIATIONS	IX
LIST OF FIGURES	XIII
LIST OF TABLES	XVII
SUMMARY	XIX
CHAPTER 1 INTRODUCTION	1
1.1 Motivation.....	1
1.2 Objectives	4
1.3 Major Contributions of the Thesis	6
1.4 Organization of the Thesis	10
CHAPTER 2 FUNDAMENTAL CONCEPTS OF QOS AND QOS SUPPORT MECHANISMS IN WiMAX SYSTEMS	13
2.1 Fundamental Concepts of QoS	14
2.2 WiMAX System.....	15
2.3 Overview of QoS Support Mechanisms in WiMAX Systems.....	23

2.3.1 QoS Support Architecture in WiMAX Systems	23
2.3.2 Bandwidth Management QoS Mechanisms in WiMAX Systems	25
2.3.2.1 Resource Reservation Mechanisms.....	25
2.3.2.2 Connection Admission Control Mechanism	26
2.3.2.3 Cross-Layer Resource Management	28
2.3.3 Traffic Handling Mechanisms in WiMAX Systems	29
2.3.3.1 Traffic Classification.....	29
2.3.3.2 Channel Access Mechanism.....	30
2.3.3.3 Traffic Policing	31
2.3.3.4 Buffer Management	32
2.3.3.5 Congestion Control	33
2.3.3.6 Packet Scheduling Algorithm.....	35
2.3.3.6.1 Holonomic Packet Scheduling Scheme	37
2.3.3.6.2 Hierarchical Scheduling Scheme	39
2.3.3.6.3 Cross-Layer Packet Scheduling Scheme	42
2.3.3.7 Adaptive Power Control.....	46
2.4 Two-Dimensional Mapping Schemes in WiMAX OFDMA System	48
2.5 Discussion	50
2.6 Summary	53
 CHAPTER 3 HIERARCHICAL QOS SUPPORT ARCHITECTURE AND THREE-TIER SCHEDULING SCHEME IN WIMAX 57	
3.1 Hierarchical QoS Support Architecture	58
3.2 Hierarchical Scheduling Scheme	59
3.2.1 DL/UL Sub-channel Bandwidth Allocation	60
3.2.2 Tier 1 Scheduling Scheme	63
3.2.2.1 Bandwidth Allocation among Different Service Classes	63
3.2.2.2 Allocated PS and BE Bandwidth Allocation among SSs	65
3.2.3 Tier 2 Scheduling Scheme	65
3.2.4 Tier 3 Scheduling Scheme	67
3.3 System Queueing Model and Packet Delay Analysis	68
3.4 Performance Evaluation	71
3.4.1 Simulation Design.....	71
3.4.2 Numerical Calculation of Average Packet Delay	72
3.4.3 Experimental Results	73
3.4.3.1 Performance Evaluation by Simulation.....	74
3.4.3.2 Performance Comparison between Simulation and Theoretical Analysis.....	78
3.5 Summary	82

CHAPTER 4 CROSS-LAYER MAC PROTOCOL AND HOLISTIC OPPORTUNISTIC SCHEDULING WITH APC SCHEME	83
4.1 AMC Schemes vs. SINR in WiMAX	86
4.2 Proposed MAC-PHY Cross-Layer Framework and MAC Protocol.....	91
4.3 Holistic Opportunistic Scheduling Algorithm	96
4.4 Adaptive Power Control Scheme.....	102
4.4.1 The First Function of APC: Requested Transmission Power Optimization	103
4.4.2 The Second Function of APC: APC Algorithm for rtPS Traffic	107
4.5 Performance Evaluation.....	110
4.5.1 Simulation Design.....	110
4.5.2 Experimental Results	112
4.6 Summary	120
CHAPTER 5 PROPOSED CAC SOLUTIONS TO SUPPORT QOS	123
5.1 Cross-Layer Elastic CAC.....	124
5.1.1 QoS Framework with CAC in tandem with Scheduling Scheme	126
5.1.2 Cross-Layer Elastic CAC Algorithm	130
5.1.3 The Cooperation between Opportunistic Scheduling and ECAC	137
5.1.4 Queueing Analytical Model	138
5.1.4.1 Effective Uplink Channel Bandwidth Analysis for Wireless Channel with AMC	138
5.1.4.2 Analysis of Average Packet Delay for Each Service Class.....	140
5.2 Cross-Layer Cognitive Radio-Based Framework and CAC Scheme	142
5.2.1 Intelligent Spectrum Management.....	144
5.2.2 Cross-Layer Cognitive Radio-Based QoS Support Framework	146
5.2.3 Cross-Layer Cognitive Radio-Based CAC Scheme.....	149
5.2.4 Queueing Analytical Model for WiMAX OFDM System.....	154
5.2.4.1 WiMAX OFDM System Queueing Model	154
5.2.4.2 Effective UL Channel Bandwidth Analysis for WiMAX OFDM Wireless Channel with AMC	158
5.2.4.3 Average Packet Delay Analysis for Each Service Class in WiMAX OFDM System.....	159
5.3 Performance Analysis	164
5.3.1 Performance Analysis for the Proposed ECAC Admission Control Scheme	165
5.3.1.1 Simulation Design	165
5.3.1.2 Simulation Results	167
5.3.1.3 Case Study for the ECAC Admission Control Scheme.....	174
5.3.2 Performance Analysis for the CRCAC Admission Control Scheme	177

5.3.2.1 Simulation Design	177
5.3.2.2 Simulation results for the CRCAC Scheme	181
5.3.2.3 Case Study for the CRCAC Admission Control Scheme	188
5.3.2.3.1 Experiment 1	189
5.3.2.3.2 Experiment 2	191
5.4 Summary	192
CHAPTER 6 COGNITIVE RADIO-BASED FRAMEWORK AND SELF- OPTIMIZING TEMPORAL-SPECTRUM BLOCK SCHEDULING	195
6.1 Research Background	198
6.1.1 Autonomic Computing	198
6.1.2 Intelligent Spectrum Management	201
6.1.3 WiMAX WirelessHUMAN™ Physic Layer	202
6.2 Cross-Layer Cognitive Radio-Based QoS Support Framework and Self- Optimizing Temporal-Spectrum Block Scheduling	205
6.2.1 Cross-Layer Cognitive Radio-Based QoS Support Framework	205
6.2.2 Cross-Layer Cognitive Radio-Based Self-Optimizing Scheduling cum Mapping Scheme	210
6.2.2.1 BS UL Channel Inter-SS Scheduling	211
6.2.2.2 CR-Based Self-Optimizing Sub-Carrier and Symbol-Duration Scheduling cum Mapping Scheme	213
6.3 Explanation of SOS ² M Scheme	219
6.3.1 SOS ² M Scheduling UGS Connections	220
6.3.2 SOS ² M Scheduling rTPS Connections	220
6.3.3 SOS ² M Scheduling nrtPS Connections	223
6.3.4 SOS ² M Scheduling BE Connections	224
6.3.5 Self_Optimizing_Mapping (CID, TSB) Algorithm	224
6.3.6 Discussion on SOS ² M Algorithm	229
6.4 Queueing Analysis for WiMAX OFDMA System	233
6.5 Performance Analysis	236
6.5.1 Simulation Design	237
6.5.2 Performance Comparison between Simulation Results and Theoretical Results	238
6.6 Summary	247
CHAPTER 7 CONCLUSION AND RECOMMENDATIONS	249
7.1 Conclusion	249

7.2 Recommendations for Future Research	255
7.2.1 Further Improvement on Current Results	255
7.2.2 Potential Research Directions	258
Author's Publications.....	265
Bibliography	269

List of Abbreviations

AC	autonomic computing
ACL	active connection list
AMC	adaptive modulation and coding
APC	adaptive power control
BE	best effort
BER	bit error rate
BPSK	binary phase shift keying
BS	base station
BW	bandwidth
BWA	broadband wireless access
CAC	connection admission control
CCE	channel condition estimator
CID	connection identifier
CR	cognitive radio
CSI	channel specification index
Cspec	channel specification
DL	downlink
DPI	dynamical priority index

DSA	dynamic service addition
DSC	dynamic service change
DSD	dynamic service deletion
FDD	frequency division duplex or duplexing
FEC	forward error correction
$G(m, BS)$	link gain from source SS_m to the BS
HOS	holistic opportunistic scheduling
IE	information element
LOS	line-of-sight
MAC	medium access control
$Max\_R(t)$	maximum transmission rate of the system
MIMO	multiple-input multiple-output
ML	maximum latency
MSTR	maximum sustained traffic rate
MSTR	maximum sustained traffic rate
NLOS	non-line-of-sight
NPSI	normalized predictive starvation index
nrtPS	non-real-time polling service
NTDSI	normalized time-delay satisfaction index
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
OMAC	opportunistic MAC
OWA	open wireless architecture
$P(BS, m)$	transmission power of the BS to SS_m
PHY	physical layer

PMP	point-to-multipoint
PS	polling service
PSI	predictive starvation index
QAM	quadrature amplitude modulation
QoS	quality of service
QPSK	quadrature phase-shift keying
RSSI	receive signal strength indicator
rtPS	real-time polling service
SAP	service access point
SC	single carrier
SFID	service flow identifier
SINR	signal-to-interference and noise ratio
SP	scheduling priority
SS	subscriber station
STBC	space-time block code
TCP	transmission control protocol
TDD	time division duplex or duplexing
TDM	time division multiplexing
TDMA	time division multiple access
TDSI	time-delay satisfaction index
Tspec	traffic specification
UGS	unsolicited grant service
UL	uplink
VC	virtual clock
WiAMX	worldwide interoperability for microwave access

WirelessHUMAN	wireless high-speed unlicensed metropolitan area network
WirelessMAN	wireless metropolitan area network
WRAN	wireless region area network

List of Figures

Figure 2.1 IEEE 802.16-2004 protocol layering.....	16
Figure 2.2 Existing QoS architecture of IEEE 802.16-2004	22
Figure 2.3 Hierarchical taxonomy of QoS support mechanisms in WiMAX systems	23
Figure 2.4 Hierarchical structure of bandwidth allocation	41
Figure 3.1 Hierarchical QoS support architecture with three-tier scheduling scheme.....	60
Figure 3.2 WiMAX UL channel queueing model: Multi-class priority TDMA queueing system.....	69
Figure 3.3 PS packet loss rate vs. traffic load for PS traffic	75
Figure 3.4 PS throughput vs. traffic load	76
Figure 3.5 BE throughput vs. traffic load	76
Figure 3.6 UL throughput vs. traffic load	77
Figure 3.7 UGS packet delay vs. traffic load.....	78
Figure 3.8 PS packet delay vs. traffic load (model without packet dropping)...	79

Figure 3.9 BE packet delay vs. traffic load.....	81
Figure 4.1 Theoretical BER vs. SINR relationship for various AMC schemes	88
Figure 4.2 Proposed cross-layer MAC protocol with QoS support framework	92
Figure 4.3 Loss rate vs. traffic load for rtPS traffic	113
Figure 4.4 rtPS delay vs. traffic load	114
Figure 4.5 rtPS UL throughput vs. traffic load	116
Figure 4.6 nrtPS UL throughput vs. traffic load	117
Figure 4.7 System UL throughput vs. traffic load	118
Figure 4.8 UGS delay vs. traffic load	120
Figure 5.1 Cross-layer QoS support framework with CAC in tandem with the HOS scheduling scheme	127
Figure 5.2 ECAC flowchart	131
Figure 5.3 Intelligent Spectrum Access White Spaces in Temporal-Spectrum domain.....	145
Figure 5.4 Cross-layer CR-based QoS support framework for WiMAX	147
Figure 5.5 CRCAC flowchart	150
Figure 5.6 Data transmission in WiMAX OFDM with TDD duplex mode	155
Figure 5.7 rtPS CBP vs. connection arrival rate	169
Figure 5.8 nrtPS CBP vs. connection arrival rate	170
Figure 5.9 Average number of rtPS on-going connections vs. connection arrival rate.....	171
Figure 5.10 Average number of nrtPS on-going connections vs. connection	

arrival rate	172
Figure 5.11 Average number of rtPS& nrtPS on-going connections vs. connection arrival rate.....	173
Figure 5.12 rtPS CBP vs. connection arrival rate	182
Figure 5.13 nrtPS CBP vs. connection arrival rate	184
Figure 5.14 Average number of rtPS on-going connections vs. connection arrival rate.....	185
Figure 5.15 Average number of nrtPS on-going connections vs. connection arrival rate.....	186
Figure 5.16 Average number of rtPS& nrtPS on-going connections vs. connection arrival rate	187
Figure 6.1 OFDMA frame structure (TDD, PUSC)	204
Figure 6.2 UL burst mapping structure.....	204
Figure 6.3 Proposed cross-layer Cognitive Radio-based QoS support framework	207
Figure 6.4 UML collaboration diagram of CR-based self-optimizing scheduling cum mapping scheme.....	211
Figure 6.5 WiMAX OFDMA system UL channel queueing model: Four M/G/1 queueing system.....	234
Figure 6.6 UGS packet delay vs. traffic load.....	240
Figure 6.7 rtPS packet delay vs. traffic load.....	241
Figure 6.8 Loss rate vs. traffic load for rtPS traffic	242

Figure 6.9 rtPS UL throughput vs. traffic load	243
Figure 6.10 nrtPS UL throughput vs. traffic load	244
Figure 6.11 System UL throughput vs. traffic load	244
Figure 6.12 BE packet delay vs. traffic load.....	245

List of Tables

Table 3.1 Numerical calculation of mean delay	73
Table 4.1 BER equation vs. E_b/N_0 for various modulation schemes	86
Table 4.2 Channel bandwidth vs. bit rate with various modulation schemes....	88
Table 4.3 Receiver minimum input level sensitivity (dBm)	90
Table 4.4 AMC vs. receiver SINR & min sensitivity requirement.....	90
Table 5.1 Average delay of UGS, rtPS and nrtPS	168
Table 5.2 Initial input traffic for the ECAC scheme.....	175
Table 5.3 New arrival input traffic for the ECAC scheme	176
Table 5.4 Simulation parameters based on IEEE 802.16d wirelessHUMAN ^{TM} OFDM system	179
Table 5.5 Numerical results of average packet delay of UGS and rtPS in one spectrum	181
Table 5.6 Initial input traffic flows for the CRCAC scheme.....	189
Table 5.7 New arrival input traffic flows for the CRCAC experiment 1 & 2 .	190
Table 6.1 Simulation parameters based on IEEE 802.16d WirelessHUMAN ^{TM} -	

OFDMA system 238

Table 6.2 Numerical results of average packet delay 239

Summary

This dissertation investigates QoS MAC protocol, architecture and QoS mechanisms. Scheduling algorithms, Adaptive Power Control (APC), Connection Admission Control (CAC) and mapping scheme are proposed to provide QoS provisioning to real-time and non-real-time applications.

The main part of the thesis presents four main contributions.

- Firstly, the thesis presents a three-tier framework and scheduling scheme to provide QoS in the WiMAX WirelessMAN-SC (Single Carrier) point-to-multipoint (PMP) system. The hierarchical scheduling algorithms perform inter-SS (Subscriber Station), inter-connections and intra-connection packet scheduling to improve QoS for PS (Polling Service) service class while achieving a high Best Effort (BE) throughput.
- Secondly, a cross-layer MAC protocol and QoS support framework associated with a two-stage opportunistic scheduling scheme and an adaptive power control scheme are proposed in WiMAX WirelessMAN-SC PMP systems.

The two-stage opportunistic scheduling algorithm is termed as Holistic Opportunistic Scheduling (HOS) with features of channel-awareness, queue-awareness and traffic QoS-awareness. It determines the dynamic

Scheduling Priority of each packet by its four key scheduling parameters, namely the dynamical priority index, the channel specification index, the normalized time-delay satisfaction index and the normalized predictive starvation index. The proposed HOS is a type of opportunistic scheduling algorithm in view of communication over a spatiotemporally varying wireless link whereby the multi-user diversity is exploited to maximize bandwidth efficiency and system throughput.

The Adaptive Power Control scheme that has been proposed is associated with the HOS scheduling algorithm. It is to optimize wireless channel utilization and to increase the transmission power of a Subscriber Station (SS) to transmit packets from real-time Polling Service (rtPS) class to the Base Station (BS).

- Thirdly, the thesis focuses on a CAC scheme that works in tandem with the scheduling algorithm. Two novel QoS support frameworks with CAC schemes are proposed.

The first proposal is a novel two-stage cross-layer QoS support framework with the cross-layer CAC scheme. The proposed CAC is termed as Elastic CAC (ECAC). It works in tandem with the Holistic Opportunistic Scheduling (HOS) in a single carrier WiMAX PMP system.

The second proposal is a novel cross-layer Cognitive Radio-based QoS support framework with the cross-layer Cognitive Radio-based CAC (CRCAC). This proposal aims at providing QoS provisioning to the heterogeneous traffic in WiMAX WirelessHUMAN^{TM}- OFDM PMP systems. The CRCAC has advanced features of cross-layer approach and

intelligent spectrum spreading ability. It is able to recognize and use one of the unused spectrum portions, hence multiplying the system capacity without a fixed spectrum bandwidth constraint.

- Finally, a novel cross-layer Cognitive Radio (CR)-based QoS support framework and a self-optimizing scheduling scheme are proposed in WiMAX WirelessHUMAN^{TM} - OFDMA PMP systems. The proposed scheduling scheme is a joint self-optimizing sub-carrier allocation and symbol-duration scheduling cum mapping scheme (SOS²M). It is an intelligent self-optimizing scheduling scheme which considers the mapping constraint. It has an intelligent spectrum spreading ability to recognize and use one of the unused spectrum portions. The solution can improve the system capacity without the constraint of a fixed spectrum bandwidth.

Chapter 1 Introduction

1.1 Motivation

Broadband wireless communication has seen a huge development since the last decade. With advanced wireless communication technology, it can accommodate many applications supported by its capability to access Internet anywhere at any time. Certainly, each type of application needs certain level of service commitment from wireless networks in order to operate successfully. This leads to a requirement of Quality of Service (QoS). There are a number of factors and components, which affect the performance of real-time and non-real time applications. Delivering such a heterogonous traffic with QoS over a time-varying wireless channel is one of the major challenges of research in the area.

Broadband Wireless Access (BWA) systems, e.g. IEEE 802.16-2004 standard [1], provide fixed-wireless access between the Subscriber Station (SS) and the Internet service provider (ISP) through the Base Station (BS). BWA systems have been deployed not only to be a complement and expansion of existing last mile wired networks such as cable modem and xDSL but also to be a competitor to wired broadband access networks. Due to the upcoming air interface technologies, which promise to deliver high transmission data rates, BWA

systems will become an attractive alternative [2][3].

The Medium Access Control (MAC) layer of the IEEE 802.16-2004 was designed for point-to-multipoint (PMP) broadband wireless access applications. It was designed to meet the requirements of very-high-data-rate applications with a variety of QoS requirements. The signaling and bandwidth allocation algorithms have been designed to accommodate hundreds of terminals per channel. The standard allows each terminal to be shared by multiple end users. The services required by the end users can be varied in their bandwidth and latency requirements, which demand the MAC layer protocol to be flexible and efficient over a vast range of different data traffic models. The system has been designed to include legacy Time-Division Multiplex (TDM) voice and data, Internet Protocol (IP) connectivity, and Voice over IP (VoIP).

The bandwidth request and grant mechanism has been designed to be scalable, efficient, and self-correcting. The IEEE 802.16-2004 does not lose efficiency when presented with multiple connections per terminal, multiple QoS levels per terminal, and a large number of statistically multiplexed users. It takes advantage of a wide variety of request mechanisms, balancing the stability of contentionless access like Polling Service (PS) with the efficiency of contention-oriented access.

The IEEE 802.16-2004 MAC provides QoS differentiation for different types of applications that might operate over IEEE 802.16-2004 networks. The IEEE 802.16-2004 standard defines the following types of services:

- **Unsolicited Grant Services (UGS):**

UGS is designed to support Constant Bit Rate (CBR) services, such as T1/E1 emulation, and Voice over IP (VoIP) without silence suppression.

- **Real-Time Polling Services (rtPS):**

rtPS is designed to support real-time services that generate variable size data packets on a periodic basis, such as MPEG video or VoIP with silence suppression.

- **Non-Real-Time Polling Services (nrtPS):**

nrtPS is designed to support non-real-time services that require variable size data grant burst types on a regular basis.

- **Best Effort (BE) Services:**

BE services are typically provided by the Internet today for Web surfing.

Since IEEE 802.16-2004 MAC protocol is connection oriented, the application first establishes the connection with the BS as well as the associated service flow (UGS, rtPS, nrtPS or BE). BS will assign the connection with a 16-bit connection identifier (CID). The connection can represent either an individual application or a group of applications such as multiple tenants in a building (all in one SS) sending data with the same CID.

All packets from the application layer in the SS are classified by the connection classifier based on CID and are forwarded to an appropriate queue. At the SS, the scheduler will retrieve the packets from the queues and transmit them to the network in the appropriate time slots or sub-carriers as defined by the UL-Map sent by the BS. The UL-Map is determined by the Uplink Bandwidth Allocation Scheduling scheme based on the BW-request messages that report the current queue size of each connection in SS.

IEEE 802.16-2004 medium access control specifies QoS signaling mechanisms such as bandwidth requests and bandwidth allocation. However, IEEE 802.16-2004 standard left the details of the QoS based packet scheduling algorithms and reservation management that determine the uplink and downlink

bandwidth allocation for rtPS, nrtPS and BE service classes, undefined.

IEEE 802.16-2004 defines the connection signaling (connection request, response and connection deletion) between SS and BS, but it does not define the details of the admission control algorithm.

IEEE 802.16-2004 PHY gives Adaptive Modulation and Coding scheme (AMC) and the conceptually power control scheme. It also left the details of the AMC and Adaptive Power Control (APC) algorithm, undefined.

This dissertation focuses on the QoS solutions that are to complete the missing parts in the QoS architecture specified in the IEEE 802.16-2004 standard.

1.2 Objectives

The objective of this research is to design and to undertake performance analysis of Medium Access Control protocols for QoS in broadband wireless access networks (WiMAX).

The first objective of this work is to provide QoS support framework alongside with proper scheduling algorithms for efficient bandwidth sharing, and QoS provisioning to rtPS traffic. To deal with a sophisticated system like WiMAX, the following factors should be considered when proposing new algorithms:

- The existing WiMAX QoS signaling mechanisms and DL/UL_Map mechanisms should be fully applied;
- Per-connection scheduling overhead should be minimized;
- The parameters of QoS for each connection should be guaranteed.

Therefore, the proposed QoS framework and the associated scheduling

algorithm should be in a hierarchical structure.

The second objective of this work is to exploit the time-varying nature of the radio environment to improve the spectral efficiency while maintaining a certain level of QoS satisfaction for each connection or user in the WiMAX systems. Considering the impact of evolving traffic characteristics on scheduling and the impact of air interface on scheduling, a MAC-PHY cross-layer design with opportunistic scheduling algorithm and adaptive power control scheme could be designed and implemented with the aims not only to exploit the multi-user diversity over wireless channel, but also to maximize bandwidth efficiency and system throughput.

The third objective of this work is to design Connection Admission Control (CAC) schemes that work in tandem with the scheduling algorithm in ensuring the number of connections in the network so as to control the congestion, connection-level QoS satisfaction and packet-level QoS satisfaction while achieving system efficiency through optimizing the network resources. Both CAC and scheduling algorithm have an important role since they can manage and guarantee the QoS requirements of the connections. A single scheduling algorithm cannot guarantee all the QoS requirements of traffic without the support of a suitable CAC scheme and vice versa. The choice of a CAC scheme is critical for the performance of a scheduling algorithm. Some suitable CAC schemes should be proposed in tandem with some versions of the opportunistic scheduling scheme.

The fourth objective of this work is to try to set up a self-optimization framework and resource allocation and scheduling scheme based on Autonomic Computing technique with a MAC-PHY cross-layer approach. It has two goals to

achieve.

Firstly, it should be able to expand the system throughput beyond the specifications in the IEEE 802.16-2004 standard. As broadband wireless access becomes ubiquitous, this makes the already heavily crowded radio spectrum much scarcer. Cognitive Radio (CR) is a promising technology to alleviate the increasing stress on the fixed radio spectrum.

Secondly, current scheduling decisions are made based solely on bandwidth requirements, system capacity, QoS of connections or PHY time-varying channel without considering the mapping constraint. Further, considering the impact of mapping constraint on scheduling decisions and the complexity of the mapping solution, the ubiquitous wireless communication can be seen to demand for a higher level of self-organization and self-optimizing ability. Autonomous scheduling mechanisms, in micro-view to design an optimally MAC protocol, an adaptive sub-carrier allocation, and a packet-level scheduling scheme with the mapping constraint in WiMAX OFDMA systems, should be proposed, designed and implemented.

1.3 Major Contributions of the Thesis

The goal of the research is to design QoS solutions to provide differentiated QoS to the heterogeneous traffic while achieving the maximum system throughput and spectral efficiency in WiMAX systems. To achieve this goal, the following QoS support solutions that have been proposed include:

- QoS support architectures
- Hierarchical scheduling scheme

- Opportunistic scheduling scheme with APC
- Elastic CAC scheme and Cognitive Radio-based CAC scheme
- Cognitive Radio-based self-optimizing temporal-spectrum block scheduling scheme with mapping scheme.

The main contributions in this thesis are as follows:

1. A novel QoS support framework alongside with a hierarchical scheduling scheme have been proposed to provide QoS support in WiMAX WirelessMAN-SC/TDD PMP systems. The proposed 3-tier architecture and scheduling schemes can provide QoS support for a wide range of real-time and non-real-time applications. The detailed 3-tier scheduling scheme is as follows:
  - Dynamic Resource Reservation (DRR) algorithm has been proposed to allocate the total bandwidth to the UL sub-channel and the DL sub-channel dynamically. The detailed theoretical derivation of the optimal solution of the DRR algorithm based on the M/M/1 queueing theorem has been carried out.
  - Priority-based Queue Length Weighted (PQLW) scheduling algorithm has been proposed for inter-class scheduling and Max-Min Fair Sharing (MMFS) scheduling scheme has been applied to inter-SS scheduling within each service class at the BS as Tier 1 scheduling.
  - Self-Clocked Fair Queueing (SCFQ) [4] and Weighted Round Robin (WRR) [5] scheduling schemes have been applied to inter-connection scheduling within each service class at each SS as Tier 2 scheduling.
  - Earliest Deadline First (EDF) [6] and Shortest Packet Length First (SPLF) [7] scheduling schemes have been applied to intra-connection

packet scheduling within each of the Polling Service (PS) and BE connections carrying burst traffic respectively as Tier 3 scheduling.

2. Taking the channel characteristics, queue characteristics and traffic QoS characteristics of traffic flows into account, a cross-layer MAC protocol and QoS support framework associated with a two-stage opportunistic scheduling scheme cum APC scheme has been proposed to provide QoS support to the heterogeneous traffic in single carrier WiMAX PMP systems. The two-stage opportunistic scheduling algorithm has been termed as Holistic Opportunistic Scheduling (HOS) with features of channel-awareness, queue-awareness and traffic QoS-awareness. Associated with the proposed MAC protocol and Holistic Opportunistic Scheduling scheme, the APC scheme has been designed for a SS or the BS to perform two functions. The first function of the APC is to optimize the requested transmission power for UL or DL transmission. The second function is to perform an adaptive power control function to increase the transmission power of a SS in order to improve its received signal $SINR$ at the BS receiver when the SS transmits packets from rtPS service class to the BS. With the APC, the SS can transmit rtPS packets at a higher data transmission rate to improve QoS for real-time traffic in the UL transmission.
3. Two novel QoS support frameworks with CAC schemes are proposed.
  - The first proposal is a novel two-stage cross-layer QoS support framework with a cross-layer CAC scheme. Taking the complexity of the WiMAX system and the impact of AMC on CAC schemes into account, a novel cross-layer QoS support framework and a cross-layer CAC algorithm have been proposed. The proposed CAC is termed as

Elastic CAC (ECAC). It works in tandem with the Holistic Opportunistic Scheduling (HOS) in a single carrier WiMAX PMP system. The ECAC consists of three modules, namely Bandwidth Allocation Estimation Module (BAEM), Bandwidth Elastic Module (BEM) and QoS Control Module (QoS_CM). The proposed ECAC can provide the connection-level QoS to the heterogeneous traffic in WiMAX WirelessMAN-SC PMP systems.

- The second proposal is a novel cross-layer Cognitive Radio-based QoS support framework with the cross-layer Cognitive Radio-based CAC (CRCAC). The proposal aims at providing QoS provisioning to the heterogeneous traffic in WiMAX WirelessHUMAN^{TM}- OFDM PMP systems. The proposed CRCAC consists of three modules, namely, Bandwidth Allocation Estimation Module (BAEM), QoS Control Module (QoS_CM) and Bandwidth Spreading Module (BSM). With the MAC-PHY cross-layer approach, it can intelligently explore unused spectrums and spread over them. Thus, the WiMAX system can operate under more than one spectrum to improve the system capacity significantly and provide guaranteed QoS to real-time traffic in the networks below 11 GHz unlicensed frequency band.
4. A novel cross-layer Cognitive Radio-based QoS support architecture and a joint self-optimizing sub-carrier allocation and symbol-duration scheduling cum mapping scheme (SOS²M) has been proposed to provide QoS to the heterogeneous traffic in WiMAX WirelessHUMAN^{TM} - OFDMA PMP systems. The proposed framework is equipped with a CR-based Intelligent Spectrum Management Module at the PHY layer. It can explore and

recognize unused spectrum portions. The system can operate under more than one spectrum. As a result, the system throughput can be improved significantly. The proposed scheduling algorithm has taken the mapping constraint into account. It has a unique feature, which is the cooperation of scheduling and mapping scheme. The efficiency of the mapping scheme can fine-tune the scheduling decision by the proposed function of self-optimizing autonomic computing. The self-optimizing autonomic computing can reduce the complexity of scheduling and mapping significantly. The QoS requirements of real-time traffic can be satisfied while a higher system efficiency and throughput can be achieved.

Queueing modeling analysis on each proposal has been carried out. Extensive simulations to evaluate the performance of the proposed solutions have been conducted. Some of them are evaluated by comparing them with the theoretical queueing analysis results or with other solutions that have some similar functions.

1.4 Organization of the Thesis

This thesis has been organized into seven chapters. Besides this introduction (Chapter 1), there are six more chapters. In Chapter 2, an overview of the fundamental concept of QoS, a basic introduction to WiMAX systems and a comprehensive overview of state-of-the-art QoS support mechanisms are given. Existing proposals for QoS provisioning have been classified into three major categories. They are the QoS support architecture, the bandwidth management mechanism and the traffic handling mechanism. Representative schemes from each of the categories have been evaluated with respect to major distinguishing

characteristics of the WiMAX MAC layer and PHY layer as specified in the IEEE 802.16-2004 standard. The proposed existing two-dimensional mapping schemes for WiMAX OFDMA systems are also summarized briefly.

In Chapter 3, a hierarchical QoS support architecture and a 3-tier scheduling scheme are presented. The queueing system model for the UL channel of WiMAX WirelessMAN-SC PMP systems is derived. The performance analysis of the system model is presented. The results of analysis are compared with that of the proposal.

In Chapter 4, the relationship among AMC, SINR and receiver sensitivity requirement in WiMAX WirelessMAN-SC PMP systems is derived. The graph of BER against SINR is plotted. A lookup table that reflects AMC vs. SINR as well as receiver sensitivity (RSS) requirement is tabulated. A MAC-PHY cross-layer QoS support architecture and the Holistic Opportunistic Scheduling (HOS) algorithm with an adaptive power control scheme are presented.

In Chapter 5, two novel QoS support frameworks with CAC schemes are presented. A novel cross-layer CAC scheme termed as Elastic CAC (ECAC) is presented in the first part of Chapter 5. The proposed ECAC is in tandem with the HOS scheduling scheme in WiMAX WirelessMAN-SC PMP systems. A novel cross-layer Cognitive Radio-based QoS support framework and a cross-layer Cognitive Radio-based CAC (CRCAC) are presented in the second part of Chapter 5. The proposed CRCAC aims at providing QoS to the heterogeneous traffic in WiMAX WirelessHUMAN^{TM}-OFDM PMP systems.

In Chapter 6, a novel cross-layer CR-based QoS support framework and a joint self-optimizing sub-carrier allocation and symbol-duration scheduling cum mapping scheme (SOS²M) are presented to provide QoS to the heterogeneous

traffic in WiMAXHUMAN^{TM}-OFDMA PMP systems.

Finally, in Chapter 7, the thesis is concluded with a summary. A brief discussion of possible future research work is presented.

Chapter 2 Fundamental Concepts of QoS and QoS Support Mechanisms in WiMAX Systems

This chapter briefly introduces the fundamental concepts of QoS and its quantitative parameters that influence the design of the QoS support mechanisms and solutions. The WiMAX system is briefly introduced. This chapter also presents various existing QoS support mechanisms that enable QoS in WiMAX systems. Existing proposals based on MAC layer with state-of-the-art technology have been classified into three main categories: QoS support architecture, bandwidth management mechanism and traffic handling mechanism. Representative schemes from each of the categories have been evaluated with respect to major distinguishing characteristics of the WiMAX MAC layer and PHY layer as specified in the IEEE 802.16-2004 standard. The existing two-dimensional mapping schemes proposed for WiMAX OFDMA PHY layer also have been summarized briefly.

2.1 Fundamental Concepts of QoS

The QoS term can be interpreted in different ways. In general, QoS can be described from two perspectives: user perspective and network perspective. In the user perspective, QoS refers to the application quality from the perspective of the user. In the network perspective, QoS refers to the service quality that the network offers to applications or users in terms of network QoS parameters that include latency or delay of packets traveling across the network, as well as packet loss rate and throughput.

From the network perspective, the network's goal is to provide the QoS to meet the users' needs adequately while maximizing the network resources' utilization. To achieve this goal, the network can analyze the application requirements, manage the network resources and deploy various network QoS provisioning mechanisms.

QoS parameters quantitatively represent the applications' QoS requirements. They are:

- Throughput
- Delay
- Delay jitter
- Error rate
- Packet loss rate

Network may use a combination of QoS support mechanisms, i.e., per-flow and quantitative, per-class and quantitative. Some networks may include multiple types of QoS support mechanisms in order to support a wide range of

applications.

2.2 WiMAX System

The IEEE 802.16-2004 and its amendment IEEE 802.16e (2005) standard have specified the Worldwide Interoperability for Microwave Access (WiMAX) system. The IEEE 802.16-2004 and IEEE 802.16e standard defines the Physical (PHY) and Medium Access Control (MAC) layer of the fixed and mobile broadband wireless access systems, respectively. This research project mainly focuses on IEEE 802.16-2004 standard.

Broadband Wireless Access (BWA) systems, e.g. WiMAX systems, provide fixed-wireless access between the SS and the Internet service provider (ISP) through the BS. BWA systems have been deployed not only to be a complement and expansion of existing last mile wired networks such as cable modem and xDSL but also to be a competitor to wired broadband access networks. Due to the upcoming air interface technologies, which promise to deliver high transmission data rates, BWA systems will become an attractive alternative.

The MAC layer of IEEE 802.16-2004 was designed for PMP broadband wireless access applications. It is designed to meet the requirements of very-high-data-rate applications with a variety of QoS requirements. The signaling and bandwidth allocation algorithms have been designed to accommodate hundreds of terminals per channel. The standard allows each terminal to be shared by multiple end users. The services required by the end users can be varied in their bandwidth and latency requirements, which demand the MAC layer protocol to be flexible and efficient over a vast range of different data traffic models. The

system has been designed to include legacy Time-Division Multiplex (TDM) voice and data, Internet Protocol (IP) connectivity, and Voice over IP (VoIP).

The MAC layer of IEEE 802.16-2004 is composed of three sub-layers as shown in Figure 2.1.

Figure 2.1 IEEE 802.16-2004 protocol layering

From bottom to top: the Security Sub-layer, the MAC Common Part Sub-layer (CPS), and the Service Specific Convergence Sub-layer (CS). The former deals with security and network access authentication procedures. CPS carries out the key MAC functions. It is connection-oriented. The CS sub-layer provides the interface to the upper layers, decides the MAC service class for the specific connection and initializes the resource allocation requests of the CPS.

The bandwidth request and grant mechanism has been designed to be scalable, efficient, and self-correcting. The IEEE 802.16-2004 does not lose efficiency

when presented with multiple connections per terminal, multiple QoS levels per terminal, and a large number of statistically multiplexed users. It takes advantage of a wide variety of request mechanisms, balancing the stability of contentionless access with the efficiency of contention-oriented access.

WiMAX system in PMP topology includes one BS and many SSs. It can operate both in Time-Division Duplex (TDD) and Frequency-Division Duplex (FDD) modes. The TDD MAC frame is divided into uplink (UL) and downlink (DL) sub-frames. Service flows, which are uniquely identified by a 32-bit service flow identifier (SFID), may be transmitted in either the uplink or the downlink sub-frame. Service flows may be created, changed, or deleted. This is accomplished through a series of MAC management messages referred to as Dynamic Service Addition (DSA), Dynamic Service Change (DSC) and Dynamic Service Deletion (DSD). Service flows can be three types, namely, provisioned service flows, admitted service flows, and active service flows. They are associated with QoS requirement parameter, namely, Provisioned QoS Parameter Set, Admitted QoS Parameter Set and Active QoS Parameter Set respectively. Admitted and active service flows are mapped onto a 16-bit connection identifier (CID). They are controlled and maintained by a Connection Admission Control (CAC) scheme in an Active Connection List (ACL) at the BS.

The traffic flows at each SS are classified according to four types of scheduling services, namely, Unsolicited Grant Service (UGS), Real-time Polling Service (rtPS), Non-real-time Polling Service (nrtPS) and Best Effort (BE) service. The mandatory QoS service flow parameters for UGS service are Maximum Sustained Traffic Rate (MSTR), Maximum Latency (ML), Tolerated Jitter and Request/Transmission Policy. The mandatory QoS service flow parameters for

rtPS are Minimum Reserved Traffic Rate (MRTR), MSTR, ML, and Request/Transmission Policy. The mandatory QoS service flow parameters for nrtPS are MRTR, MSTR, Traffic Priority and Request/Transmission Policy. The mandatory QoS service flow parameters for BE are MSTR, Traffic Priority, and Request/Transmission Policy.

Bandwidth is always requested on a CID basis and bandwidth is allocated on a SS basis. BS should have an admission control policy to decide whether the QoS of a connection can be satisfied. On the uplink (from SS to BS), the BS determines the number of time slots or sub-carriers that each SS will be allowed to transmit in an uplink sub-frame. This information is broadcasted by the BS through the uplink map message (UL-Map) at the beginning of each frame. UL-Map contains Information Element (IE), which includes the transmission opportunities, i.e. the time slots or sub-carriers in which the SS can transmit during the uplink sub-frame. After receiving the UL-Map, each SS will transmit data in the predefined time slots or sub-carriers as indicated in IE. The BS uplink-scheduling module determines the IEs using bandwidth request Protocol Data Unit (PDU) (BW-request) sent from SSs to the BS.

The IEEE 802.16-2004 PHY specifies five different PHY specifications, namely, WirelessMAN-SC (Single Carrier) PHY specification, WirelessMAN-SCa PHY specification, WirelessMAN-OFDM PHY specification, WirelessMAN-OFDMA PHY specification and WirelessHUMAN^{TM} PHY specification.

1. WirelessMAN-SC PHY specification targets systems operating in the range of 10 to 66 GHz. Line-of-sight (LOS) propagation paths are a practical necessity in such systems. The standard specifies single-carrier modulation

schemes to be with Forward Error Correction (FEC). The air interface supports both FDD and TDD operation modes. The size of frame and the sub-frames can be adjusted, enabling the PHY layer to adopt adaptive modulation and coding schemes. Each sub-frame is divided into numbers of time slots. In the BE service, the bandwidth requests from all SSs will contend for the transmission to the BS. By the polling schemes, every SS is implicitly polled periodically by the BS to send its bandwidth request. The bandwidth requests will be comprehensively scheduled at the BS for each SS transmission. The BS informs each SS to share the UL transmission by the scheduling results through the UL-Map.

2. WirelessMAN-SCa PHY specification targets non-line-of-sight (NLOS) systems operating in frequency bands below 11 GHz.
3. WirelessMAN-OFDM PHY is designed for NLOS operation in the frequency bands below 11 GHz. It uses Orthogonal Frequency-Division Multiplexing (OFDM) with a 256-point transform. Multiple-access is made available by Time Division Multiple Access (TDMA). This air interface is mandatory for license-exempt bands.
4. WirelessMAN-OFDMA PHY is based on OFDM modulation and designed for NLOS operation in the frequency bands below 11 GHz. For licensed bands, channel bandwidth is limited to the regulatory provisioned bandwidth divided by any power of 2 no less than 1.0 MHz. It uses Orthogonal Frequency-Division Multiple Access (OFDMA) with a 2048-point transform. In this system, multiple-access is provided by addressing a subset of the multiple carriers to individual receivers. IEEE 802.16-2004 allows for different mappings of sub-carriers to sub-channels, which are logical

groupings of sub-carriers. In Adjacent Sub-carrier Permutation (ASP) mode, adjacent groups of sub-carriers (number of sub-carriers can vary) are taken together to form a sub-channel. Thus, a sub-channel is formed of sub-carriers in adjacent frequency bands. In Distributed Sub-carrier Permutation (DSP), the sub-carriers in a sub-channel are widely apart in the frequency spectrum.

OFDMA may be considered a hybrid of time- and frequency-division multiple access. In the time domain, data is transmitted in the form of frames. The minimum time-frequency unit that can be allocated to a user is a slot. A slot consists of a sub-channel over one, two, or three OFDM symbols, depending on the sub-channelization scheme that is used.

5. WirelessHUMAN^{TM} PHY specification not only complies with the WirelessMAN-SCa PHY, the WirelessMAN-OFDM PHY and the WirelessMAN-OFDMA PHY, but also further complies with the Dynamical Frequency Selection (DFS) protocols. WirelessHUMAN^{TM} PHY specification is implemented for license-exempt frequencies below 11 GHz. The license-exempt nature introduces new mechanisms such as DFS to detect and avoid interference. Hence, it provides the platform whereby Cognitive Radio (CR) can perform.

The standard defines three different modulation schemes. On the uplink, support for QPSK is mandatory, while 16-QAM and 64-QAM are optional. The downlink supports QPSK and 16-QAM, while 64-QAM is optional.

In addition to these modulation schemes, the IEEE 802.16-2004 PHY also defines various FEC schemes on the uplink as well as the downlink. These include Reed-Solomon (RS) codes, RS concatenated with inner Block Convolution Codes (BCC), and turbo codes. Support for such a wide variety of

modulation and coding schemes permits vendors to tradeoff efficiency for robustness depending on the channel conditions.

The advanced technology of the IEEE 802.16-2004 PHY requires equally advanced radio link control (RLC), particularly the capability of the PHY to transit from one burst profile to another. The RLC must control this capability as well as the traditional RLC functions of power control and ranging. The system uses the Receiver Sensitivity (RSS) as a parameter together with the Signal-to-Interference and Noise Ratio (SINR) thresholds of receivers used by the AMC scheme to select different burst profiles in order to maximize the network throughput and maintain the Bit Error Rate (BER) under a preset level e.g. $BER \leq 1 \times 10^{-5}$.

The IEEE 802.16-2004 standard targets the large amount of spectrum potentially available for point-to-multipoint (PMP) systems in the 10 to 66 GHz frequency range. Due to wide variations in the local regulations in different regions, no frequency plan is specified in this standard. However, sufficient commonality exists to specify a default Radio Frequency (RF) channel bandwidth for each major region.

Figure 2.2 shows the existing QoS architecture of IEEE 802.16-2004.

Uplink bandwidth allocation scheduling resides in the BS to control all the uplink packet transmissions. The communication path between SS and BS has two directions: uplink channel (from SS to BS) and downlink channel (from BS to SS). On the downlink (from BS to SS), the transmission is relatively simple because the BS is the only one that transmits packets during the downlink sub-frame.

IEEE 802.16-2004 medium access control, which is based on the concepts of

connections and service flows, specifies QoS signaling mechanisms such as bandwidth requests and bandwidth allocation.

Figure 2.2 Existing QoS architecture of IEEE 802.16-2004

However, as shown in Figure 2.2, IEEE 802.16-2004 standard left the details of the connection admission control algorithm undefined. IEEE 802.16-2004 standard also left the details of the QoS based packet-scheduling algorithms and reservation management that determine the uplink and downlink bandwidth allocation, undefined. IEEE 802.16-2004 PHY frames out AMC and the conceptual power control scheme. It also left the details of the AMC and adaptive power control algorithm undefined.

2.3 Overview of QoS Support Mechanisms in WiMAX Systems

In recent years, QoS support architecture and QoS support mechanisms have been proposed [8-34, 37-44, 47-57] in WiMAX systems. They can be classified according to the following taxonomy as shown in Figure 2.3

Figure 2.3 Hierarchical taxonomy of QoS support mechanisms in WiMAX systems

2.3.1 QoS Support Architecture in WiMAX Systems

An inclusive architecture was proposed to support QoS mechanisms in IEEE 802.16 standard [8]. The authors developed some compatible methods for specific modules such as Scheduler, Traffic Shaper, and Request and Grant

Manager to optimize delay, throughput and bandwidth utilization metrics.

In paper [9], the authors proposed MAC layer cross to network upper layer QoS framework in the downlink mode and uplink mode to provide QoS support in WiMAX systems. The proposed cross-layer QoS framework integrated Layer 3 (L3) and Layer 2 (L2) QoS in the IEEE 802.16 network. Main functional blocks in the framework include QoS mapping from L3 to L2, Admission Control, Fragment Control, and Remapping. Fragment Control handles the data frames from the same IP datagram as a group in L2 operations to reduce useless transmission. Remapping is designed for more flexible use of L2 buffers by changing the mapping rules from IP QoS to L2 service type under congested situation of the rtPS queue.

In paper [10], a cross-layer design framework was proposed for IEEE 802.16e OFDMA systems that is compatible with WiBro based on various kinds of cross-layer protocols for performance improvement: a cross-layer adaptation framework and a design example of primitives for cross-layer operation between its MAC and PHY layers. In the proposed model, the MAC layer contains a User Grouper, Scheduler, and Resource Controller. Each functional entity exploits physical layer information to increase system throughput. The physical layer consists of a diversity channel Physical-layer Protocol Data Unit (PPDU) controller, AMC channel PPDU controller, control information controller, and hybrid-automatic repeat request (HARQ) functional blocks. AMC sub-channel users and diversity sub-channel users are classified by the user grouper. Since the properties of AMC sub-channels and diversity sub-channels are quite different, the grouping of users into two channel types is essential if system throughput is to be increased. The scheduler determines the scheduling of users and the

quantity of packets that should be scheduled in the current frame. For cross-layer optimization, the scheduler should be designed to exploit not only the information of PHY but also the information of application layer.

2.3.2 Bandwidth Management QoS Mechanisms in WiMAX Systems

Bandwidth management mechanisms are mechanisms that manage the network resources by coordinating and configuring network devices' traffic handling mechanisms. The main mechanisms are:

- Resource reservation
- Connection admission control
- Cross-layer resource management

2.3.2.1 Resource Reservation Mechanisms

Resource reservation mechanisms inform the network entities on the QoS requirements of the applications using the network resources. The network devices will use this information to manage the network resources in order to meet such requirements. The resource reservation mechanisms include the following functions:

- Provision of resource reservation signaling that notifies all devices along the communication path on the multimedia application's QoS requirements.
- Delivery of QoS requirements to the connection admission control mechanism that decides if there are available resources to meet the QoS

requirements of the new connection.

- Notification of the application regarding the admission result.

The representative proposal tailored for WiMAX is Dynamic Resource Reservation (DRR) scheme [11]. The basic principle of DRR is that the reserved bandwidth will vary between a minimum and a maximum value as per the bandwidth utilized by the clients. The proposal is able to optimize reservation and utilization of bandwidth for Committed Bandwidth (CB) type traffic. However, it is very difficult to select parameters C_m and T . Moreover, the fluctuations of the reserved bandwidth from C_M to C_m will increase signaling costs.

2.3.2.2 Connection Admission Control Mechanism

Admission control is a network QoS procedure. Admission control determines how bandwidth is allocated; therefore, it needs to be implemented between network edges and core to control the traffic entering the network. The role of CAC is to control the number of connection flows into the network. A new connection request is progressed only when sufficient resources are available at each successive network element to establish the connection through the whole network based on its service category, traffic contract, and QoS, while the agreed QoS of all existing connections are still maintained. Admission control is useful in situations where a certain number of connections may all share a link, while an even greater number of connections cause significant degradation in all connections to the point of making them all useless, such as in congestive collapse.

In paper [12], a reservation-based CAC scheme (R-CAC) was proposed. It is

characterized by defining thresholds of allocated bandwidth at BS for each class of service based on the priority of the service. The proposal introduced two parameters, denoted as C_u and C_r . C_u is the bandwidth exclusively reserved for UGS service, which is allocated with the highest priority in the four service classes supported in IEEE 802.16d systems. C_r is the bandwidth exclusively reserved for UGS and rtPS to mitigate the bandwidth competition coming from the other two types of service. The residual bandwidth (i.e., $C - C_u - C_r$) is the only part that a BS can assign to the nrtPS connection requests to meet their minimum bandwidth requirements, and it can also be assigned to UGS and rtPS services. D-CAC [13] is a Priority Support-based CAC. The proposed scheme can give the highest priority for UGS flows and maximize the bandwidth utilization by bandwidth borrowing and degradation. Token bucket-based CAC [14] is to control each connection. The connection is controlled by two token bucket parameters: token rate r_i (bps) and bucket size b_i (bits). When a traffic flow wants to establish a connection with a BS, it sends these two parameters to the BS and waits for a response from the BS. An extra parameter, delay requirement d_i , will be sent by rtPS flow. In order to avoid starvation of some classes, the proposal suggested that a threshold is set for each class. Combining with a proper scheduler, token bucket-based CAC can reserve bandwidth needed by real-time flows and thus delay requirements of rtPS flows can be promised. Binary search approach fairly allocating bandwidth CAC [15] used Gaussian model for aggregated traffic in large network and Chernoff bound method to obtain upper bound blocking probability. Based on the analysis result, binary search approach was applied to solve the problem that given a total bandwidth, fairly allocating bandwidth to each class of multimedia traffic in IEEE 802.16d

networks. The total bandwidth (C) is completely portioned for four classes of traffic, and the partition value C_i ($i = 1, 2, 3, 4$) is calculated by the above binary search algorithm using Chernoff bound. If a new connection of UGS and BE arrives at SS, it will send a request to BS for bandwidth. It is granted bandwidth if the new aggregated bandwidth including this connection is less than C_i ($i = 1, 2, 3, 4$). Else, it is blocked. If a new connection of rtPS and nrtPS arrives at SS, the connection is always admitted, but the burst within the connection will be blocked when the total used bandwidth is larger than C_i ($i = 2, 3$). This mechanism can guarantee the pre-required upper bound blocking probability for UGS and BE connections, and the burst blocking probability for rtPS and nrtPS.

An optimization-based policy has been suggested in [16, 17]. A degradation strategy has been used in [18-20]. More conservative algorithm would maintain the QoS provisioning for on-going connections and simply reject a new service flow [21] if the remaining bandwidth is less than the bandwidth requirement of the new service flow. The AMC-induced CAC [22] and TCP-Aware CAC [23] have adopted a cross-layer approach. Although a CAC scheme for an opportunistic scheduling based on Minimum Rate Outage and Delay Outage which regulates the number of connections or users in the system to obtain multi-user diversity has been proposed [24], it has not been combined with the opportunistic scheduling algorithm substantively.

2.3.2.3 Cross-Layer Resource Management

A few prior works deal with MAC cross to the higher layers resource management, i.e. the MAC to application and transport layers. Paper [25] was aimed at providing end-to-end QoS guarantee service using IntServ and DiffServ

in connection oriented WiMAX PMP and mesh networks. The work maps RSVP in the IP layer and DSA/DSC/DSD in the MAC layer. The authors proposed that message exchange for DSA and DSC could be deployed to carry QoS parameters of IntServ services for end-to-end resource (bandwidth/buffer) reservation. For DiffServ services, on the other hand, a number of per-hop behaviors (PHBs) for different classes of aggregated traffic could be mapped into different connections directly.

2.3.3 Traffic Handling Mechanisms in WiMAX Systems

Traffic handling mechanisms are the mechanisms that classify, handle, police and monitor the traffic across the network. The main mechanisms are:

- Classification
- Channel access
- Traffic policing
- Buffer management
- Congestion avoidance
- Packet scheduling
- Adaptive power control

2.3.3.1 Traffic Classification

The classification mechanism identifies and separates different traffic into flows or group of flows. Therefore, each flow or group of flows can be handled differently. Application traffic is identified by the classification mechanism and is forwarded to the appropriate queue awaiting service from other mechanism such

as traffic shaping and packet scheduling. The granularity level of the classification mechanism can be per-user, per-flow or per-class depending on the type of QoS provided. To identify and classify the traffic, the traffic classification mechanism requires some form of tagging or marking of packets.

In paper [26], a two-stage packet classification algorithm is proposed. The authors suggested that prefix-based fields and range-based fields should be processed separately in two stages. In the first stage, packets are classified by a scheme that matches with their prefix-based fields, while other range-based fields are processed by a range-based scheme in the second stage. The Prefix-Matching-Tree (PMT) is used in the first stage to handle prefix-based fields. The PMT is constructed by a prefix-based matching scheme that can speedup the searching process. Each tree node of the first stage is connected to the second stage according to protocol types (TCP or UDP). The Range-Matching-Tree (RMT) is employed to deal with range-based fields in the second stage.

2.3.3.2 Channel Access Mechanism

In wireless networks, all hosts communicate through a shared wireless medium. When multiple hosts try to transmit packets on the shared communication channel, collisions can occur. Therefore, wireless networks need a channel access mechanism that controls the access to the shared channel. Collision-based channel such as Random access and collision-free channel access such as TDMA or Polling channel access mechanism can provide different QoS support.

In IEEE 802.16d, there are two access protocols based on multi-channel slotted Aloha and periodic polling. The former is a contention based access protocol, while the latter provides periodic polling services without contention, such as

Unsolicited Grant Service (UGS), e.g., Voice over IP and real-time or non-real-time services. In addition, IEEE 802.16d allows unicast polling for a single SS or multicast polling for groups of SSs as periodic polling service.

In paper [27], the performance of IEEE 802.16d Random Access Protocol was evaluated by using Transient Queueing Analysis. The authors derived the random access success probability from the system equilibrium. Retransmission probability is also derived by including a binary exponential back-off algorithm. In paper [28], the authors considered a capacity allocation scheme of periodic polling services for a multimedia traffic in an IEEE 802.16d network. Considering a BS that assigns a Subscriber Station contiguous M uplink sub-frames for uplink traffic transmission and ν vacation frames for saving power and opportunities for other Subscriber Stations, they proposed a capacity allocation scheme for a multimedia traffic in WiMAX systems. The bandwidth allocated to a SS will be returned to the BS, when the SS's queue is empty during M contiguous uplink sub-frames. The returned bandwidth will be allocated to other types of services requested by the multichannel slotted Aloha.

2.3.3.3 Traffic Policing

Traffic policing is the mechanism that monitors the admitted sessions' traffic so that the sessions do not violate their QoS contract. The traffic policing mechanism makes sure that all traffic that passes through it will conform to the agreed traffic parameters. In the case of violation, a traffic policing mechanism will be enforced by shaping the traffic and dropping the traffic to enforce compliance with that contract. Traffic sources that are aware of a traffic contract sometimes apply traffic shaping in order to ensure their output stay within the contract. Due

to the fact that traffic policing shapes the traffic based on some known quantitative traffic parameters, multimedia (real-time) application are naturally compatible to traffic policing. Most multimedia application traffic (voice, video) is generated by a standard codec, which generally provides certain knowledge of the quantitative traffic parameters. Traffic policing can be applied to individual multimedia flows. Non-real-time traffic does not provide quantitative traffic parameters and usually demands bandwidth as much as possible. Therefore, traffic policing enforces non-real-time traffic based on network policy. Such policing is usually enforced on aggregated non-real-time traffic flows. Traffic policing, in cooperation with other QoS mechanisms, usually can provide QoS support.

In paper [29], a new traffic shaping based on the concept of “Fair Marker” (FM) was proposed to enforce fairness among distinct flows. FM controls token distribution from the token bucket to the flows originating from the same subscriber network. The FM explores the duality between packet queueing and token bucket utilization. Fairness in token distribution is a function of the fair allocation algorithm used by FM. In order to reach this purpose, it records information regarding the consumption of tokens by the monitored flows.

2.3.3.4 Buffer Management

Buffer management refers to any particular discipline used to regulate the occupancy of a particular queue where packets may be held (or dropped). Buffer is set to improve link utilization and system performance, but it also increases packet’s queue delay. With the increase of user demands for service quality, providing stable and low delay has been the primary requirement of real-time

services. The most important and easy to control part of total delay is queue delay. Therefore, how to set the capacity of the buffer, how to control buffer length efficiently and how to achieve the tradeoff between throughput and queue delay are the important problems to be solved in buffer management and QoS control of whole networks. At present, support [30] for wired and wireless network is included for drop-tail (FIFO) queueing, Random Early Detection (RED) buffer management, Class-Based Queueing (CBQ) (including a priority and round-robin scheduler), and variants of Fair Queueing including Fair Queueing (FQ), Stochastic Fair Queueing (SFQ), and Deficit Round-Robin. RED is the most intensive researched class of AQM (Active Queue Management); it selectively discards packets of some flows based on predetermined probability.

In WiMAX systems, the BS is a likely bottleneck for downlink (DL) TCP connections due to the difference in available bandwidth between the fixed network and the wireless link. This may result in buffer overflows or excessive delays at the BS, as these buffers are connection-specific. In order to avoid buffer overflows, different Active Queue Management (AQM) methods may be applied at the BS. In paper [31], the authors analyzed RED, Packet Discard Prevention Counter (PDPC) and time-to-live based RED AQM mechanisms and proved that they are indeed very useful: AQM reduces DL delays at the BS in WiMAX system considerably without sacrificing TCP throughput.

2.3.3.5 Congestion Control

Congestion control concerns controlling traffic entry into a network, to avoid congestive collapse by attempting to avoid oversubscription of any of the processing or link capabilities of the intermediate nodes and networks and taking

resource reducing steps, such as reducing the rate of sending packets. In congestion control, the packet loss information can serve as an index of network congestion for effective rate adjustment, but in a wireless network environment, common channel errors due to multipath fading, shadowing, and attenuation may cause bit errors and packet loss, which are quite different from the packet loss caused by network congestion. Therefore wireless packet loss can mistakenly lead to dramatic performance degradation.

If the link capacity in one system is temporarily degraded due to a high traffic load in the WLAN system (congestion) or due to interference, two effects, leading to inefficiency on the IEEE 802.16d link, could occur:

- Loss of data due to buffer overflow and there will be unnecessary retransmissions;
- Waste of bandwidth due to unused reserved transmission opportunities.

To avoid the above-mentioned effects, a congestion control mechanism needs to be worked out to dynamically adapt the QoS demands of a connection during runtime for a specifically defined period. In paper [32], the authors proposed Dynamic Service Change (DSC) congestion control mechanism to support the Explicit Congestion Notification mechanism in future deployment of TCP. The authors contributed to IEEE 802.16-REVd_D1(2003) standard by adding MAC_DSC_TEMP.request and MAC_DSC_TEMP.indication message in MAC entity to provide the temporary traffic reduction mechanism to overcome congestion effect.

Since network congestion is directly related to the congestion packet loss, packet loss can be caused by either congestion loss or wireless channel errors, resulting from multipath fading, shadowing, or attenuation. New approach of

congestion control over wireless network is to perform packet loss classification so that congestion control algorithms can more effectively adapt the sending rate based on congestion loss instead of from wireless loss. In paper [33], the authors considered two packet loss classes, congestion loss and wireless loss and proposed Packet Loss Classification (PLC) method based on the trend of ROTT(relative one-way trip time) to assist packet loss classification in the ambiguous area of ROTT distribution.

By taking advantage of the QoS features offered by one of the four proposed WiMAX service flow arrangement, paper [34] was aimed at a more flexible layer constructing and subscription while reliable in diverse channel conditions and fitting users' demand. Through effective integration of Packet Loss Classification (PLC) [33], end-to-end available bandwidth probing, congestion control via layered structure and packet level FEC, for layered multicast applications over WiMAX for disseminating scalable extension of H.264/AVC [35] compressed video is proposed. The optimality comes from the best tradeoff of a number of video layers subscription with a number of additional FEC packets insertion simultaneously to satisfy the estimated available bandwidth and wireless channel error condition.

2.3.3.6 Packet Scheduling Algorithm

Packet scheduling refers to the decision process used to choose which packets should be serviced or dropped. Packet scheduling is the process of resolving contention for bandwidth. A scheduling algorithm has to determine the allocation of bandwidth among the users and their transmission order. One of the most important tasks of a scheduling scheme is to satisfy the QoS requirements of its

users while efficiently utilizing the available bandwidth.

Many legacy-scheduling algorithms, capable of providing certain guaranteed QoS, have been developed for wire-line networks. However, these existing service disciplines, such as Fair Queueing scheduling, Virtual Clock, and EDD, are not directly applicable in wireless networks because they do not consider the varying wireless link capacity and the location-dependent channel state. The characteristics of wireless communication pose special problems that do not exist in wire-line networks. These include:

- high error rate and bursty errors
- location-dependent and time-varying wireless link capacity
- scarce bandwidth
- user mobility
- power constraint of the mobile hosts

All of the above characteristics make developing efficient and effective scheduling algorithms for wireless networks very challenging.

WiMAX systems provide services for heterogeneous classes of traffic with different QoS requirements. Currently, there is an urgent need to develop new technologies for providing QoS differentiation and guarantees in WiMAX systems. Among all the technical issues that need to be resolved, packet scheduling in WiMAX systems is one of the most important.

In this sub-section, the existing proposed scheduling algorithms for QoS support in WiMAX systems are assessed thoroughly with respect to the characteristics of the IEEE 802.16d MAC layer and PHY layer. Those scheduling algorithms are classified into three categories: Holonomic Approach [37-40], Hierarchical Approach [21][41-44] and Cross-layer Approach [46-57] with

respect to the nature of scheduling algorithm mechanism. Holonomic Approach uses single layer scheduling scheme, in contrast, Hierarchical Approach uses several layers or stages scheduling schemes. Cross-layer Approach uses information from several layers in Open System Interconnection (OSI) Reference Model. It is a design by the violation of reference layered communication architecture with respect to the particular layered architecture [36]. Each of three categories can further be classified as per-flow, per-class, per-packet and hybrid scheduling algorithms. Representative schemes in each of these categories will be discussed next.

2.3.3.6.1 Holonomic Packet Scheduling Scheme

In paper [37], the authors applied Holonomic Approach and proposed one layer hybrid-scheduling scheme which combines per-flow and per-class scheduler termed as “Frame Registry Tree Scheduler” (FRTS). The proposal aims at providing differentiated treatment to data connections, based on their QoS characteristics. This approach focuses on properly preparing future transmitted frames by using a tree-based approach. The tree consists of six levels; root, time-frame, modulation, subscriber, QoS and connection level. The first level is taken to be the root. The second level represents time-frame immediately after the current time frame. The third level represents the available modulation types. The fourth level organizes all the connections according to the SS. Each SS has one uplink node and one downlink node at this level. The fifth level organizes the connections according to their QoS. The last level consists of leaves for each active connection queue. The data structure presented achieves time-frame creation and reduces the processing needs at the beginning of each frame. The

algorithm schedules each packet at the last time-frame before its deadline. This allows more packets to be transmitted and hence an increased throughput. This method also avoids fragmentation of transmissions to/from the same SS or same modulation. Another good feature is its ability to handle changes in the connection characteristics like modulation type or service type of the channel.

In [38], the authors proposed Token Bank Fair Queueing (TBFQ) scheduling algorithm at packet level for BWA systems. TBFQ uses the priority index E_i/r_i to keep track of the normalized service received by backlogged flows. E_i is the number of tokens exchanged between the ‘bank’ and flow i . E_i is negative if the flow continues to borrow tokens from the bank to serve the traffic that exceeds its average rate. E_i is positive if the traffic is below its average rate. The flows are first served based on their token generation rate to guarantee the throughput and latency, and then the remaining bandwidth is distributed based on their priority index. The parameters (*debt limit*, *credit burst*, and *creditable threshold*) determine the dynamic behavior of the algorithm. TBFQ can be adapted to operate under varying channel error conditions. It has demonstrated its effectiveness in achieving fairness, maximizing channel utilization, and fast convergence to guaranteed throughput. Its ability to serve and isolate real-time traffic and data traffic that is under severe error conditions makes it a suitable candidate as the wireless packet-scheduling algorithm for BWA systems.

In order to maximize throughput of non-real-time traffic with satisfying QoS requirements of real-time traffic, the authors [39] proposed urgency and efficiency based packet scheduling (UEPS) algorithm that was designed not only to support multiple users simultaneously but also to offer real-time and non-real-time services to a user at the same time. The UEPS algorithm uses the time-utility

function as a scheduling urgency factor and the relative status of the current channel to the average one as an efficiency indicator of radio resource usage. The proposed packet scheduler assigns priorities to the packets to be transmitted, based on the channel status reported by the user equipments as well as the QoS statistics maintained by the BS. Since the scheduler works in a global timeline, a time utility function (TUF) is used for the scheduling. Two scheduling factors, the *urgency* of scheduling and the *efficiency* of radio resource usage, are used to schedule real-time and non-real-time traffic packets at the same time. The TUF is used to represent the urgency of scheduling while the channel state is used to indicate the efficiency of radio resource usage.

A new per-flow based scheduling algorithm was proposed [40], referred as Service Criticality (SC) based scheduling scheme. SC is based on a dual of buffer occupancies at nodes and allowable latencies for a particular service. In the proposed scheme, a flow would receive the service through bandwidth allocation depending on degree of service criticality (SC_{index}) computed at SS and conveyed to BS during bandwidth request burst of UL. The core of the proposed SC scheme is the way SS computes the SC_{index} for active flows existing at node. The computation of SC_{index} depends on both, the buffer occupancy and latency experienced by flow. Proposed scheme employs linear dependence on buffer occupancy and tunable sigmoid like relation with experienced latency to compute SC_{index} .

2.3.3.6.2 Hierarchical Scheduling Scheme

In paper [41], the issue of differentiated service provisioning was addressed with the non-real-time polling service in WiMAX systems. The proposed

solution has been designed to have an ability to accommodate integrated traffic in the networks with effective scheduling schemes. A hierarchical scheduling algorithm to provide service differentiation to enhance the nrtPS service in WiMAX systems was proposed. In order to meet the time constraints of real-time messages as much as possible and avoid scheduling starvation of the non-real-time messages, the data transmission scheduling has been divided into two levels, inter-class scheduling and intra-class scheduling. With the objective to provide service differentiation between the real-time and non-real-time classes of traffic, the Proportional Delay Differentiation (PDD) model was proposed as inter-class scheduling algorithm. Intra-class scheduling scheme was designed as a priority assignment scheme based on tardy rate, the message with less transmission time has higher priority to be served. The message, which has been delayed a longer time in the queue also has higher priority to be transmitted. With this scheme, the total time for messages to stay in the network could be reduced and the starvation of the messages with longer transmission time could be avoided.

In paper [21], the authors proposed a hybrid scheduling algorithm that combines Earliest Deadline First EDF [6], WFQ and FIFO scheduling algorithms. It has two-tier hierarchical scheduling structure. The overall allocation of bandwidth is done in a strict priority manner. EDF scheduling algorithm is used for SSs of the rtPS class, WFQ is used for SSs of the nrtPS class and FIFO for SSs of the BE class. In paper [42], the authors enhanced the proposal scheduling architecture in [21], the scheduling architecture is similarly divided into two layers as shown in Figure 2.4 The first layer is for bandwidth requests. The authors suggested Deficit Fair Priority Queue (DFPQ) for scheduling at this layer. The second layer scheduling is for the data traffic. UGS is not scheduled because

it already has a reserved bandwidth. For the other three traffic classes, a hybrid of scheduling algorithm is proposed. The authors suggested EDF for rtPS traffic, where the packets with earliest deadline are scheduled first. For nrtPS, WFQ is proposed. The bandwidth left is allocated to each BE traffic in a round robin (RR) manner.

Figure 2.4 Hierarchical structure of bandwidth allocation

Compared with fixed bandwidth allocation, the proposed solution [42] improves the performance of throughput under unbalanced uplink and downlink traffic. Furthermore, better performance in fairness can be achieved by the proposed DFPQ algorithm than that by strict PQ scheduling.

Paper [43] extended the scheduling architecture in paper [42]. It proposed a Preemptive DFPQ scheduling algorithm that enhances the DFPQ algorithm proposed in [42], and improves the performance of the rtPS service class. The Preemptive DFPQ defines for each non-preemptive queue a *Quantum Critical*

(Q_{crit}) to give the queue another chance to serve critical packets. Q_{crit} value is a percentage of the original value of the queue's quantum. Queues are allowed to use Q_{crit} to serve critical packets only. The processing of critical packets continues until the Q_{crit} of the non-preemptive queue becomes less than or equals to zero. Q_{crit} is initialized only once per frame and not at every round like the quantum Q . This algorithm gives more chances to rtPS packets to be serviced before the expiration of their deadlines.

The authors in [44] proposed an architecture consisting of three schedulers. The first scheduler concerns UGS and rtPS flow, as well as rtPS and nrtPS polling flow. EDF scheduling is applied in this scheduler. The second scheduler concerns flows requiring a minimum bandwidth, mainly for nrtPS. WFQ scheduling is used here where the weight is the size of the requested bandwidth. The third scheduler is used for BE traffic and here, WFQ scheduling is employed too, where the weight is the traffic priority. Among three schedulers, the first level has the highest priority, and only after all the packets have been served, the second scheduler is considered. The third scheduler comes when the first two have become free. The delay and delay-jitter character for UGS, rtPS, nrtPS and BE can be improved simultaneously using the proposed architecture.

2.3.3.6.3 Cross-Layer Packet Scheduling Scheme

Wireless communication systems have unique characteristics – namely, time-varying channel conditions and multi-user diversity. A new MAC design and new scheduling solutions need to be developed that are specifically tailored for this wireless communication environment [45]. Opportunistic MAC (OMAC) is the modern view of communicating over spatiotemporally varying wireless link. The

cross-layer nature embeds OMAC with the potential to revolutionize the design of wireless data networks from physical to data link layers.

The wireless resources (bandwidth and power) are more scarce and expensive than their wired counterparts, because the overall system performance degrades dramatically due to multi-path fading, Doppler, and time-dispersive effects caused by the wireless air interface. Unlike wired networks, even if large bandwidth/power is allocated to a certain wireless connection, the loss and delay requirements may not be satisfied when the channel experiences deep fades. In this scenario, the scheduler plays an extremely important role. An OMAC seeks to pick among competing users the one who is currently experiencing the relatively best channel conditions in each scheduling instant. Judicious schemes should be developed to support prioritization and resource reservation in wireless networks, in order to enable guaranteed QoS with efficient resource utilization.

Recently, various opportunistic scheduling schemes have been proposed for wireless communication systems. They can exploit the time-varying nature of the radio environment to improve the spectral efficiency while maintaining a certain level of QoS satisfaction for each connection or user in various wireless networks. The proposed opportunistic scheduling schemes can be further classified into channel-aware only and channel-aware & queue-aware algorithm based on their functionality of scheduling algorithms [46].

Max Carrier-to-Noise Ratio Scheduling (MCS) [47] is a typical channel-aware opportunistic scheduling scheme. The MCS scheme is to allocate resources to users with the best channel conditions to achieve high system throughput.

On top of channel characteristics, traffic characteristics also play an important role in the design of opportunistic scheduling algorithms, channel-aware &

queue-aware opportunistic scheduling schemes were proposed [48-53]. The Proportional Fair Scheduling (PFS) schemes [48-50] attempt to trade-off among the throughput, efficiency and fairness among users by taking packet length into account [49], or estimating future channel quality [50]. Modified Largest Weighted Delay First (M-LWDF) [51] is a modified version of PFS scheduler that tries to meet QoS requirement in terms of head-of-line packet delay. Traffic-Aided Opportunistic Scheduling (TAOS-1) [52] is a heuristic opportunistic scheduling scheme that unifies file size information and wireless channel variations in order to reduce the completion time of file transmissions. Exponential Rule Scheduler (EXPRule) [53] attempts to equalize the weighted delays of all buffers when their differences become larger in a wireless system.

Cross-layer opportunistic scheduling designs tailored for WiMAX systems have been proposed [54-57]. Paper [54] addressed a channel-aware scheduling algorithm with a per-flow channel-error-compensation technique based on the typical features of WiMAX systems: class-based QoS guarantees and per-flow resource assignment. The Worst-Case Fair Weighted Fair Queuing + (WF2Q+) algorithm was suggested to manage the flow-level and class-level granularity per-class queues to provide the target QoS to each WiMAX service class while achieving fairness of traffic flows belonging to the same class. In paper [55], the authors proposed a cross-layer design of packet scheduling and resource allocation in OFDMA wireless networks, which concentrates on downlink scheduling in the BS. In OFDMA systems, each channel is subdivided into a number of sub-carriers that can be controlled or allocated separately, making the PHY very robust. In the proposed system, the sub-carriers are grouped into Allocation Units (AU) for allocation so that overheads are minimized. The BS

estimates the sub-channel condition of each user and allocates resources to the users on a frame-by-frame basis. The authors considered a system where some SS have real-time traffic and other systems have non-real-time traffic, i.e. real-time and non-real-time traffic that do not co-exist in the same SS. The BS can estimate the channel gain of each user on a sub-channel and an AU can be independently allocated to a particular user. The authors in [55] formulated a linear programming method and subsequently obtained a sub-optimal solution that reduces the computational time. The scheme also proposed a packet scheduler at the BS MAC layer that provides equal priority to both real-time and non-real-time traffic, except if the ratio of the waiting time of a real-time packet in the MAC queue and the delay constraint of the packet exceeds a certain threshold. In paper [55], the authors proposed a 3-stage resource allocation algorithm. The first stage schedules the urgent packets. The second stage schedules the non-real-time packets and the real-time packets that are not urgent; higher priority is given to users whose channel quality is better regardless of the traffic type. The third stage allocates the unallocated AUs, if any. After the packet scheduler decides the rate requirements of each user, the actual resources are allocated in the PHY layer. This is carried out by using a sub-optimal heuristic algorithm that first allocates the sub-channels (or sub-carriers) constraints in order to maximize the utilization (or minimize the transmission power). After that, the algorithm reallocates (or swaps) sub-channels, so as to satisfy the rate requirements of each user. Though this kind of cross-layer scheme can be followed in the downlink at the BS, it cannot be adopted for uplink scheduling since individual nodes cannot decide the channel condition and place request for those particular channels.

In paper [56], the authors proposed a priority-based scheduler at the medium access control layer where each connection employs AMC scheme at the PHY layer. The authors defined a Priority Function (PRF) for each connection admitted in the system and updated it dynamically depending on the wireless channel quality, QoS satisfaction and service priority through MAC-PHY cross-layer manner. The strategy proposed in [57] further enhanced the architecture in [10] by proposing a joint packet scheduling and sub-channel allocation scheme (JPCAS) in WiMAX OFDMA system. The basic principle of the proposal is firstly to assign a priority to each packet based on its integrated QoS requirements, PHY Channel State Information (CSI) and service priority in each sub-channel; then according to such priority queues in each sub-channel, the packet is transmitted on a sub-channel with good CSI.

2.3.3.7 Adaptive Power Control

Appropriate power control can not only reduce power consumption, but also improve system capacity by adjusting coverage ranges and improve spatial reuse efficiency. Higher transmission rates require higher SINR to maintain the same bit error rate. Due to this close relationship between rate and channel condition, incorporating transmission rate selection into MAC design is another promising way to increase the system performance.

Paper [58], which considered Adaptive Power Allocation (APA), emphasizes on how to share the limited power resource of BS among different WiMAX subscribers and further influence the access bandwidth of each subscriber. The authors suggested that APA and CAC have to work cooperatively to provide cross-layer resource management to build a WiMAX access network. A power

allocation scheme that produces optimal revenue, known as the optimal revenue criterion, was studied. In order to investigate the APA revenue of a certain scheme, the revenue rate of each type of service as the revenue generated by a bandwidth unit was defined.

Let rer^{UGS} , rer^{rtPS} , rer^{nrtPS} and rer^{BE} be the revenue rates of the following, respectively:

- Unsolicited Grant Service
- Real-time Polling Service
- Non-real-time Polling Service
- Best-Effort Service

Obviously, different services have different prices due to their specific QoS requirements. As a result, rer^{UGS} , rer^{rtPS} , rer^{nrtPS} and rer^{BE} always take distinct values. To maximize the APA revenue, the potential revenue of each subscriber, which is defined as the revenue that could be achieved if all arriving traffic is served, must be investigated. The potential revenue of a given subscriber k is determined by the amount of UGS, rtPS, nrtPS, and BE traffic load in its local network and the price of service (i.e., rer^{UGS} , rer^{rtPS} , rer^{nrtPS} and rer^{BE}). Let TL_k^D denote the arriving downlink traffic load in subscriber k 's local network, and suppose this traffic load can generate potential revenue RD_k . Then, the revenue-to-bandwidth ratio of the k^{th} subscriber is defined as $RBR_k^D = RD_k / TL_k^D$. Since different WiMAX subscribers can have different amounts of UGS, rtPS, nrtPS and BE traffic in their local networks, they hold distinct revenue-to-bandwidth ratios. The optimal revenue-criterion-based APA optimization has inherent preference to allocate more power resource to the sub-carriers that belong to the subscriber with high revenue-to-bandwidth ratio.

2.4 Two-Dimensional Mapping Schemes in WiMAX OFDMA System

In WiMAX OFDMA PHY specification, a data region is a two-dimensional allocation of a group of contiguous sub-channels in a group of contiguous OFDMA symbols. This allocation may be visualized as a rectangle, such as a 4×3 rectangle [1].

One of the most basic approaches for packet mapping in OFDMA downlink or uplink subframes is Fixed Mapping Algorithm. Fixed bursts are predefined in advance and then packets are assigned to them. In this approach, a subframe is pre-segmented into a set of M fixed rectangles of different heights and widths. Pre-segmentation has to be done usually by using an expected distribution of packets and available rates, creating various bursts to achieve efficient use of wireless resources.

The fixed allocation algorithm is simple. The criterion for selecting a pre-segmented burst for a transmission packet is that the size of the pre-segmented burst must be larger than that of the packet. However, the main disadvantage of the algorithm is the fact that the prior knowledge of the distribution of packets is required in advance [59].

Another approach for burst construction in OFDMA downlink or uplink subframes is Raster Scanning Approach, in which the subframe is filled row by row, from left to right and from top to bottom while allocating requests according to specific criteria [59-62]. The algorithm in [60] is similar to the algorithm in [59] but it allows a burst compaction if there are more than one burst that belongs

to the same physical mode. It allocates bursts in the time domain first and then the frequency domain (left to right and top to bottom). However, this algorithm does not consider unused space, which results in a reduced throughput. A “biggest first” mapping algorithm [61] discussed the issue of optimizing resource allocation for rectangular mapping. The algorithm [62] basically allocates the burst from right to left and bottom to top.

The Raster Scanning Approach algorithm is more efficient than the Fixed Mapping Algorithm, because the size of a burst is not predefined and is changed according to the packet size. The main disadvantage of this approach is that the number of bursts is not considered. As the number of bursts increases, the number of control data items, such as the MAC header and the cyclic redundancy check (CRC), in a frame increases, which causes degradation of frequency efficiency. Moreover, UL/DL-Map IEs have to be prepared for each burst. Therefore if the number of bursts is large, the size of the UL/DL-Map becomes large, which also degrades frequency efficiency [59].

Another approach for burst construction in OFDMA downlink or uplink subframes is Optimizing Placement Algorithm [63]. Prior to any optimization, the proposal first fits each burst within the global time-frequency rectangle of the downlink frame, and minimizes the total duration of the frame (number of OFDM symbols), in order to minimize bandwidth waste. The main disadvantage of the algorithm is computational complexity, which increases quickly when the number of burst increases. The algorithm is only suitable for a limited number of bursts.

The two-dimensional rectangular mapping problem is a variation of the bin-packing problem. This packing problem is known to be NP-complete. The

complexity of the solution grows exponentially with the number of bursts [64]. There have been many proposals [65, 66, 67] to tackle this problem. However, there is no easy way to achieve optimality with simple computations.

2.5 Discussion

IEEE 802.16-2004 medium access control, which is based on the concepts of connections and service flows, specifies QoS signaling mechanisms such as bandwidth requests and bandwidth allocation. However, as shown in Figure 2.2, there are deficiencies in IEEE 802.16-2004 standard. Those important components for QoS provisioning, are left unspecified in the standard. Hence, they remain as open research problems.

- The details of the QoS based packet-scheduling algorithms and reservation management that determine the uplink and downlink bandwidth allocation are left undefined.
- The details of the connection admission control algorithm are left undefined.
- The details of the AMC and adaptive power control algorithm are left undefined.

Currently, many scheduling solutions have been proposed in WiMAX systems. However, the current existing scheduling schemes proposed to provide QoS in WiMAX WirelessMAN-SC PMP networks have the following shortcomings:

- There is no DL/UL sub-channel bandwidth allocation scheme to allocate the total bandwidth to DL/UL sub-channels.
- The information carried in the UL/DL_Map, specified in the standard, to

determine the time slots for the packet transmission of different service types has not been fully utilized.

- The problems of the starvation of low priority service class exist.
- Burst nature of the real-time or non-real-time traffic have been ignored.

Cross-layer MAC and opportunistic scheduling designs tailored for WiMAX systems have been proposed [54-57]. However, the existing opportunistic scheduling schemes proposed for WiMAX systems only take channel characteristics and one of the traffic characteristics like delay or queue length into account without evaluation of some common performance metrics such as packet loss rate. The scheduling schemes have not been applied to packet scheduling in each traffic flow. In addition, the priority coefficient setting for rtPS and nrtPS traffic connections is rigid while in fact, traffic load and average long-term channel conditions vary dynamically.

Current scheduling decisions are made solely based on bandwidth requirements, system capacity, QoS of connections and PHY time-varying channel without considering the mapping constraint. Actually, Downlink (DL) and Uplink (UL) resource scheduling are done in two steps in WiMAX OFDMA systems. In the first step, those proposed schedulers decide the number of sub-carriers and the number of slots in units of symbol-duration to be allocated to each Subscriber Station (SS) and those allocated sub-carriers and symbol-durations are further distributed to each connection at the SS. In the second step, the mapping algorithms allocate the scheduled UL data bursts from each SS into two-dimensional rectangular “areas” [1] whereby the length of the area is the number of sub-carriers and the width is the number of OFDM symbols.

The two-dimensional rectangular mapping problem is a variation of the bin-

packing problem. There have been many proposals to tackle this problem. However, there is no easy way to achieve optimality with simple computations.

Both CAC and the scheduling algorithm have an important role since they can manage and guarantee the QoS requirements of the connections. A single scheduling algorithm cannot guarantee all the QoS requirements of traffic without the support of a suitable CAC and vice versa. CAC schemes tailored for WiMAX networks have been proposed in [12-25]. The combination of a scheduling policy with an efficient CAC scheme to guarantee QoS in WiMAX networks has become an important issue. Although some CAC policies have been proposed to combine with an uplink scheduling algorithm [14, 21, 25] to provide QoS support in WiMAX, very few of them were catered for WiMAX networks with the tight combination of an opportunistic scheduling algorithm.

Currently, the proposed CAC schemes operate under the Fixed Spectrum Allocation (FSA) scheme with a pre-defined fixed spectrum with a fixed channel bandwidth or fluctuating channel bandwidth if the AMC scheme is used. No matter how efficient the CAC is, the number of connections or users that can be admitted in the system would be limited under the limited bandwidth constraint in the pre-defined fixed spectrum.

The optimal revenue-criterion-based APA optimization has inherent preference to allocate more power resource to the sub-carriers that belong to the subscriber with a high revenue-to-bandwidth ratio. However, it is hard to guarantee QoS to real-time applications in terms of packet delay and loss rate.

2.6 Summary

This chapter briefly introduces the fundamental concepts of QoS and the features of MAC and PHY specifications in WiMAX systems.

This chapter also has summarized various QoS support mechanisms in WiMAX systems. Existing proposals have been classified into three main categories: QoS support architecture, bandwidth management mechanism and traffic handling mechanism. Representative schemes from each of the categories have been evaluated with respect to major distinguishing characteristics of the MAC layer and PHY layer of WiMAX systems as specified in the IEEE 802.16-2004 standard. The proposed two-dimensional mapping schemes for WiMAX OFDMA PHY layer have also been introduced briefly.

This chapter can serve as a good starting point for researchers working on this area since it gives them an overview of the research efforts conducted for the past decade.

Based on the intensive review of the state-of-the-art proposals contribution to WiMAX systems in the past few years, some new research issues or trends emerged. They are highlighted as follows.

The computational complexity of a proposed algorithm strongly influences its scalability. The QoS support framework in WiMAX systems could be a hierarchical structure. A system model and scheduling solution with a hierarchical structure will be presented in the next chapter.

Opportunistic MAC (OMAC) is the modern view of communicating over spatiotemporally varying wireless channel whereby the multi-user diversity is exploited rather than combated to maximize bandwidth efficiency or system

throughput. The cross-layer nature embedded OMAC has the potential to revolutionize the design of wireless data networks from physical to data link layers.

Cross-layer opportunistic scheduling and adaptive power control scheme will be presented to provide QoS to the heterogeneous traffic in the WiMAX WirelessMAN-SC system in Chapter 4.

Scheduling algorithms provide mechanisms for bandwidth allocation and multiplexing at the packet level. Admission control and congestion control policies are all dependent on the specific scheduling disciplines used in WiMAX systems. For the uplink traffic, the scheduling algorithm has to work in tandem with CAC to satisfy the QoS requirements. The CAC algorithm ensures that a connection is accepted into the network only if its QoS requirements can be satisfied and the performance of existing connections in the network has not deteriorated. A novel CAC solution in tandem with the opportunistic scheduling scheme to provide QoS provisioning in the WiMAX WirelessMAN-SC system will be presented in the first part of Chapter 5.

There is a large amount of research results on CAC scheme to provide QoS to the traffic with Maximum Latency constraint in WiMAX systems. However, all those proposed CAC schemes operate under the Fixed Spectrum Allocation (FSA) scheme with pre-defined fixed spectrum and fixed channel bandwidth or fluctuation channel bandwidth if AMC scheme is used. No matter how efficient the CAC is, the number of connections or users, which can be admitted in the system, would be limited under the limited bandwidth constraint in the pre-defined fixed spectrum. A novel cross-layer CR-based QoS support framework and a cross-layer CR-based CAC (CRCAC) scheme will be presented in the

second part of Chapter 5. The proposed solution has advanced features of cross-layer approach and intelligent spectrum spreading ability. It is able to recognize and use one of the unused spectrum portions, hence multiplying the system capacity without a fixed spectrum bandwidth constraint in WiMAX WirelessHUMAN^{TM}-OFDM PMP systems.

Many schemes have been proposed to improve the network performance in terms of packet delay, packet loss rate or system throughput. However, no scheme so far has considered the issue of the sub-carrier allocation, symbol-duration scheduling and temporal-spectrum mapping by Autonomic Computing technique for WiMAX OFDMA systems.

Hence, in Chapter 6, a novel cross-layer Cognitive Radio-based QoS support framework and a joint autonomic self-optimizing sub-carrier allocation and symbol-duration scheduling cum mapping scheme (SOS²M) will be presented. With new emerging technologies, like Cognitive Radio and Autonomic Computing technique, the proposed solution is able to provide QoS to the heterogeneous traffic in WiMAX OFDMA PMP systems.

Chapter 3 Hierarchical QoS Support Architecture and Three-Tier Scheduling Scheme in WiMAX

Scheduling algorithm is important in the provisioning of guaranteed Quality of Service parameters such as packet delay, packet loss rate and system throughput. To date, many scheduling solutions have been proposed for WiMAX systems. Comprehensive surveys on those scheduling proposals are provided in Chapter 2. However, the current existing scheduling schemes proposed to provide QoS in WiMAX WirelessMAN-SC PMP networks have the shortcomings as follows:

- There is no DL/UL sub-channel bandwidth allocation scheme to allocate total bandwidth to DL/UL sub-channels.
- The information carried in the UL/DL_Map, specified in the standard, to determine the time slots for the packet transmission of different service types has not been fully utilized.
- The problems of the starvation of low priority service class exist.
- Burst nature of the real-time or non-real-time traffic have been ignored.

In this chapter, a hierarchical QoS support architecture and a three-tier

scheduling scheme are proposed to provide QoS support for a wide range of real-time and non-real-time applications in WiMAX WirelessMAN-SC PMP systems. The proposed solution can improve the WiMAX system performance through enhancements of the QoS support framework with dynamic bandwidth allocation and efficient scheduling schemes.

3.1 Hierarchical QoS Support Architecture

The following factors have been taken into consideration to design the QoS support architecture and corresponding scheduling schemes for QoS provisioning in WiMAX PMP systems:

- The existing WiMAX QoS signaling mechanisms and DL/UL_Map mechanism should be fully applied;
- Per-connection scheduling overhead should be minimized;
- The parameters of QoS for each connection should be guaranteed.

As shown in Figure 2.2, those important components for providing QoS, are left unspecified in the standard. This chapter focuses on scheduling architecture and scheduling scheme aspects in WiMAX WirelessMAN-SC PMP systems.

The proposed QoS provisioning architecture aims at providing scheduling modules to complete two of the missing parts in the QoS architecture specified in the IEEE 802.16-2004 standard as shown in Figure 2.2. As stated in Section 2.5, there is no DL/UL sub-channel bandwidth allocation scheme to allocate the total bandwidth to DL/UL sub-channels. The DRR algorithm is proposed at the BS to allocate the overall bandwidth to the UL and DL sub-channel dynamically. As the scheduling scheme for rtPS, nrtPS and BE service classes remains as an open

research problem, the hierarchical scheduling module is proposed at the BS as well as at the SS. Due to the complex of the WiMAX system, a divide and conquer methodology is adopted. The scheduling scheme for data transmission is further divided into three hierarchical tiers. Tier 1 scheduling exists at the BS only. It performs two functions. The first function is to allocate the overall bandwidth of UL or DL to four service classes. The second function is to distribute the allocated bandwidth of each service class to every SS. Tier 2 scheduler further allocates the bandwidth, which is granted by Tier 1 scheduler, to different connections within each service class at the SS. Tier 3 scheduling is to determine the transmission priority of a packet in a connection with burst traffic at each SS.

Based on the abovementioned factors, a three-tier with Dynamic Resource Reservation (DRR) hierarchical scheduling architecture and corresponding scheduling schemes are proposed as shown in Figure 3.1.

3.2 Hierarchical Scheduling Scheme

The proposed hierarchical scheduling scheme follows the framework specified in the IEEE 802.16d standard with the following unique features.

- The Dynamic Resource Reservation (DRR) scheme was proposed to allocate bandwidth to UL sub-frame and DL sub-frame dynamically.
- The Priority-based Queue Length Weighted (PQLW) scheduling algorithm was proposed for inter-class scheduling and the MMFS scheduling was introduced for inter-SS scheduling within each class of service at the BS as Tier 1 scheduling.
- The Self-Clocked Fair Queueing (SCFQ) [4] and Weighted Round Robin

(WRR) [5] scheduling schemes were applied to inter-connection scheduling within each class of service at each SS as Tier 2 scheduling. Earliest Deadline First (EDF) [6] and Shortest Packet Length First (SPLF) [7] scheduling were applied to packet scheduling within each of the connections carrying burst traffic as Tier 3 scheduling.

Figure 3.1 Hierarchical QoS support architecture with three-tier scheduling scheme

3.2.1 DL/UL Sub-channel Bandwidth Allocation

Since the BS is linked to a backbone network, it has traffic flows from Internet service providers (ISP) to SSs. The UL channel is shared by multiple SSs to access the Internet. Thus, the traffic load of UL and DL could be unbalanced. The proposed DRR scheduler is to allocate overall bandwidth to the DL and UL sub-channel dynamically according to the change of traffic load from different SSs.

The bandwidth of UL or DL sub-channel can be calculated based on the DRR

scheduler as follows:

$$DL\_Bandwidth = R \times Total\_Bandwidth$$

$$UL\_Bandwidth = (1-R) \times Total\_Bandwidth$$

where R is the ratio of DL sub-channel's bandwidth to the entire network bandwidth.

Lemma 3.1: In case of using dynamic DL and UL sub-channel allocation algorithm to achieve minimum end-to-end average packet delay, the optimal value of R is:

$$R = \frac{DL\_QL + 1}{DL\_QL + UL\_QL + 2} \quad (3.1)$$

where DL_QL and UL_QL represent the queue length of DL sub-channel and UL sub-channel respectively.

Proof:

From the system's point of view, SSs in WiMAX system send packets via a UL sub-channel to the BS. The BS then forwards the packets to its destination via a DL sub-channel. Assume packet arrival at each SS is a Poisson arrival process and the packet length is exponentially distributed. Let λ_n is the average arrival rate of the n^{th} SS for UL transmission, where $n=1, 2, \dots, N$. Let λ_{UL} denote the average arrival rate of UL traffic from all SSs where $\lambda_{UL} = \sum_{n=1}^N \lambda_n$. It is also a

Poisson arrival process based on the Poisson superposition property that is merging of N Poisson streams with mean λ_n , resulting in a Poisson with mean rate $(\lambda_1 + \lambda_2 + \dots + \lambda_N)$. Similarly, the average arrival rate of DL traffic to all SSs, λ_{DL} , is also a Poisson process. Based on Kleinrock independence approximation [68], the DL and UL can be modeled as two individual $M/M/1$ queueing system in

series with service rate μ_{DL} and μ_{UL} respectively to provide service to DL traffic and UL traffic with a constraint as:

$$\mu_{total} = \mu_{DL} + \mu_{UL} \quad (3.2)$$

where μ_{total} is the total service rate of the system. Let T_{DL} be the mean time spent by a packet to get through the DL channel and T_{UL} be the mean time spent by a packet to get through the UL channel. Let T_{total} be the average end-to-end packet delay, which is the average time spent by a packet to get through UL from SSs to the BS then to get through the DL from the BS to its destination. T_{total} is expressed as:

$$T_{total} = T_{DL} + T_{UL} \quad (3.3)$$

According to Little's Theorem, the average number of packets (N) in a queueing system is equal to the average arrival rate (λ) of packets to that system multiplied by the average time (T) a packet spent in that system. T_{DL} and T_{UL} are expressed as:

$$T_{DL} = N_{DL}/\lambda_{DL} = \rho_{DL}/[\lambda_{DL}(1 - \rho_{DL})] = 1/(\mu_{DL} - \lambda_{DL}) \quad (3.4)$$

$$T_{UL} = N_{UL}/\lambda_{UL} = \rho_{UL}/[\lambda_{UL}(1 - \rho_{UL})] = 1/(\mu_{UL} - \lambda_{UL}) \quad (3.5)$$

From (3.2), μ_{UL} is given by:

$$\mu_{UL} = \mu_{total} - \mu_{DL} \quad (3.6)$$

Substituting (3.4) and (3.5) into (3.3), T_{total} is calculated as:

$$\begin{aligned} T_{total} &= 1/(\mu_{DL} - \lambda_{DL}) + 1/(\mu_{total} - \mu_{DL} - \lambda_{UL}) \\ &= (\mu_{total} - \lambda_{DL} - \lambda_{UL}) / [(\mu_{DL} - \lambda_{DL}) \times (\mu_{total} - \mu_{DL} - \lambda_{UL})] \end{aligned} \quad (3.7)$$

The main objective of the DRR scheduler is to allocate overall bandwidth to UL sub-channel and DL sub-channel with the aim of minimizing the average end-to-end packet delay. The optimization problem can be formulated as:

$$\text{Minimize } T_{total} = (\mu_{total} - \lambda_{DL} - \lambda_{UL}) / [(\mu_{DL} - \lambda_{DL}) \times (\mu_{total} - \mu_{DL} - \lambda_{UL})] \quad (3.8)$$

Subject to $\mu_{DL} + \mu_{UL} = \mu_{total}$;

$$(\lambda_{DL} + \lambda_{UL}) / \mu_{total} < 1;$$

Let $y = 1 / T_{total}$,

$$y = [-\mu_{DL}^2 + \mu_{DL}(\mu_{total} + \lambda_{DL} - \lambda_{UL}) - \lambda_{DL}\mu_{total} + \lambda_{DL}\lambda_{UL}] / (\mu_{total} - \lambda_{DL} - \lambda_{UL}) \quad (3.9)$$

This is a quadratic function. The vertex is $(-b/2a, (4ac-b^2)/4a)$. The optimal solution can be obtained as:

$$\mu_{DL} = (\mu_{total} + \lambda_{DL} - \lambda_{UL}) / 2 \quad (3.10)$$

R is the ratio of DL sub-channel bandwidth to the entire network bandwidth. R can be expressed as μ_{DL} / μ_{total} .

$$R = 1/2 + (\lambda_{DL} - \lambda_{UL}) / 2\mu_{total}$$

$$R = 1/2 + R\lambda_{DL} / 2\mu_{DL} - (1-R)\lambda_{UL} / 2\mu_{UL} \quad (3.11)$$

According to $M/M/1$ queueing theorem, λ_{DL} / μ_{DL} and λ_{UL} / μ_{UL} are expressed as:

$$\lambda_{DL} / \mu_{DL} = DL\_QL / (DL\_QL + 1) \quad (3.12)$$

$$\lambda_{UL} / \mu_{UL} = UL\_QL / (UL\_QL + 1) \quad (3.13)$$

Substituting (3.12) and (3.13) into (3.11), R is calculated as:

$$R = (DL\_QL + 1) / (DL\_QL + UL\_QL + 2) \quad (3.14)$$

Thus, Lemma 3.1 is proved.

3.2.2 Tier 1 Scheduling Scheme

Tier 1 scheduler works at the BS only. It has two functions, which are to allocate the overall bandwidth to the four service classes and to allocate the bandwidth of every service class to each SS.

3.2.2.1 Bandwidth Allocation among Different Service Classes

At the BS, Tier 1 UL scheduling algorithm collects bandwidth requests from

all SSs and allocates bandwidth to rtPS, nrtPS and BE service classes. UGS service class is allocated with fixed bandwidth based on its bandwidth requirement specified in the standard.

The rtPS and nrtPS service classes can be combined under a general Polling Service (PS) umbrella as suggested in [41]. In order to overcome the starvation problem in the strict priority queue algorithm [21], PQLW scheduling algorithm is proposed to allocate overall UL bandwidth excluding the bandwidth for UGS to PS and BE service classes.

Let

$$PI_i = P_i + QR_i, \quad i=1, 2 \quad (3.15)$$

$$QR_i = \sum_{j=1}^n Q_{ij} / \sum_{i=1}^2 \sum_{j=1}^n Q_{ij}, \quad (3.16)$$

$$f_i = PI_i / \sum_{i=1}^2 PI_i, \quad (3.17)$$

where P_i is the priority of the i^{th} service class. It is set to 1 for PS service class and 0 for BE service class. PI_i is the Priority Index of the i^{th} service class. QR_i is the ratio of the bandwidth request of the i^{th} service class to the entire PS and BE bandwidth requests. Q_{ij} represents the bandwidth request for the i^{th} service class sent by the j^{th} SS and f_i is the factor of the i^{th} service class bandwidth proportional to the entire PS and BE bandwidth. The bandwidth for PS and BE service classes can be calculated as follows:

$$C_i = C_{PB} \times f_i \quad i=1, 2 \quad (3.18)$$

where C_i is the bandwidth allocated to the i^{th} service class. C_{PB} is the bandwidth for PS and BE service classes which is UL channel bandwidth excluding the time slots for UGS, the time slots for PS_response and the time

slots for BE bandwidth request contention period. As a result, BE service class will get at least a small portion of C_{PB} based on its bandwidth request. The starvation problem can be resolved.

3.2.2.2 Allocated PS and BE Bandwidth Allocation among SSs

The BS Tier 1 scheduler performs the second function to distribute the PS and BE bandwidth allocated in the UL sub-frame to every SS by MMFS [69] scheduling algorithm to fully utilize the allocated bandwidth.

Consider a set of SSs $1, 2, \dots, N$, which has PS Bandwidth_request $Sp_{s1}, Sp_{s2}, \dots, Sp_{sN}$ where $Sp_{s1} \leq Sp_{s2} \leq \dots \leq Sp_{sN}$. C_{PS} is the bandwidth of PS allocated in the UL sub-frame. Initially allocate C_{PS}/N of the bandwidth to every SS. The remaining bandwidth ($C_{PS}/N - Sp_{s1}$) is still available as total bandwidth exceeds from the bandwidth request of SS₁. Distribute the remaining bandwidth to remaining $(N-1)$ SSs. Each of them gets $[C_{PS}/N + (C_{PS}/N - Sp_{s1})/(N-1)]$. This process ends when C_{PS} is fully distributed to all SSs. The bandwidth allocation for the BE service class to each SS follows the same way.

The outcome of the scheduling is the total bandwidth for UGS service class, the total bandwidth for PS and BE service class at each SS. Then it is loaded into the DL/UL _Map and broadcasted by the BS at the beginning of each DL sub-frame.

3.2.3 Tier 2 Scheduling Scheme

Tier 2 scheduling is essentially a type of connection scheduling. The operation of the scheduler working at every SS is that, when a SS receives the UL_Map, it extracts the bandwidth allocation information for the SS, which is the total bandwidth for the UGS service class, the total bandwidth for the PS service class

and the total bandwidth for the BE service class at that SS. Then it applies different scheduling algorithms to allocate bandwidth among different traffic connections within PS and BE service classes based on their traffic characteristics.

The detailed scheduling scheme is as follows:

1. Bandwidth allocation within PS class connections: SCFQ scheduling is applied to allocate bandwidth among different connections in the PS service class because this algorithm can provide performance guarantees in terms of bandwidth bound and end-to-end delay bound to each connection. The algorithm transmits packets in the order of their finishing numbers. Let $P(i, k, t)$ be the length of a newly arrived packet, the k^{th} packet, to connection i at time t . Let $F(i, k-1, t)$ be the finishing number for the $(k-1)^{th}$ packet. The finishing number of the k^{th} packet is:

$$F(i, k, t) = \max[F(i, k-1, t), CF] + P(i, k, t)/\Phi_i \quad (3.19)$$

where CF is the current finishing number of the packet currently in service. Φ_i is the weight of the i^{th} connection. It is calculated as:

$$\Phi_i = C \times \frac{1/d_i}{\sum_{i=1}^n 1/d_i} \quad (3.20)$$

where C is the channel transmission rate, d_i is average deadline of packets in the queue of the i^{th} connection. As connections in PS service class have Maximum Latency requirement, average deadline of packets in a connection of PS service class is used to set the weight of the connection. The packet in a PS connection with short average deadline has high weight thus it has a small finishing number, it will be transmitted first to reduce its dropping probability due to its deadline constraint.

2. Bandwidth allocation within BE connections: In order to protect well-

behaved traffic flows against ill-behaved ones and to achieve fairness among BE connections, the remaining bandwidth will be allocated based on WRR algorithm to those BE connections which have successfully sent out bandwidth requests during the contention period.

For BE connections $j=1, 2, \dots, J$, let W_j be the weight of connection j , it is calculated as:

$$W_j = Q_j / \sum_{j=1}^J Q_j \quad (3.21)$$

where Q_j is the queue length of the j^{th} connection in unit of bytes. As packets in BE service class do not have deadline constraints, they could be dropped in a queue due to the queue length exceeding the maximum queue size. The queue length of each BE connection decides its weight. The scheduler serves a BE connection in proportion to its weight in turn. The connection with long queue length will be given high weight, and is allocated more bandwidth. As a result, the queue length of connections in the BE service class can be maintained stable.

3.2.4 Tier 3 Scheduling Scheme

Tier 3 scheduling is a type of packet scheduling for the burst PS and BE traffic. Considering the burst nature of the traffic, each burst in a flow has its different individual time constraint. Moreover, each burst, either in PS or BE traffic flows, has its individual burst size. Effective scheduling on the variable size packets (bursts) in a traffic flow is necessary to produce a more efficient usage of the bandwidth. Different scheduling algorithms have been applied to PS and BE service flows with burst traffic based on their traffic characteristics:

- Packet scheduling in PS burst traffic flows: EDF [6] algorithm is applied to schedule the transmission of the bursts in PS burst traffic flows. The packets with the earliest deadline will be transmitted first.
- Packet scheduling in BE burst traffic flows: SPLF [7] scheduling algorithm is applied. The packet with the shortest length will be transmitted first.

Since the scheduler at each SS has the updated information on the status of each connection, tight QoS guarantees could be achieved by Tier 2 scheduling algorithm and Tier 3 scheduling algorithm.

3.3 System Queueing Model and Packet Delay Analysis

The communication between SSs and the BS has two directions of UL and DL. The thesis focuses on the UL channel analysis. The UL channel of WiMAX TDMA PMP systems can be modeled as a multi-class priority TDMA queueing system as shown in Figure 3.2.

The UL channel can be considered as a multiple-access communication channel shared by N SSs. The UL sub-frame consists of M consecutive time slots. Each time slot is normalized to be of unit length τ . Let the duration of a time frame be T_F . So that $T_F = M\tau$. The TDMA scheme for inter-SS is Max-Min Fair Sharing (MMFS) [69] scheduling, in which station i is allocated n_i time slots per frame, uniformly distributes in UL sub-frame.

Figure 3.2 WiMAX UL channel queueing model: Multi-class priority TDMA queueing system

The TDMA scheme for inter-class is the strict priority queue scheme in which higher priority packets are queued ahead of lower priority ones. Packets of the same priority class that arrive at different slots are served on a first-come-first-served basis. Packets of the same priority class that arrive during the same slot are randomly ordered for transmission. Each station can transmit its packets only during its dedicated time slots, which is synchronized and governed by the BS via DL/UL_Map.

Packets arriving at each SS belong to one of the three priority classes: UGS (priority-1, high priority), Polling Service (PS) (priority-2) or BE (priority-3, lowest priority). The rtPS and nrtPS service classes are combined under the Polling Service (PS) as suggested in [41]. At each station, packet arrival is a Poisson arrival process, so that λ_j is the average arrival rate of class j packets, $j=1, 2, \dots, J$. Each packet is transmitted in different number of time slots according to its length. The number of time slots to transmit the k^{th} arriving packet of priority class j is denoted by B_k^j ($j = 1, 2, \dots, J$). The time taken to transmit the k^{th} packet

belonging to class j is represented as σ_k^j .

The transmission requests can be expressed by integer multiples of a time slot. The probability of collision of the request contention in the BE is assumed to be zero in the queueing model. The information of bandwidth allocation are carried by DL/UL_Map and broadcasted to all SSs. When the UL channel becomes available, any waiting priority- i packet is served before any priority- j one, if $i < j$.

Assume that each SS is identical in general. Hence, the packet delay and queue-size behavior of any station is statistically independent of that of any other station. Consequently, the packet delay behavior of an arbitrary single station, say Station 1, is investigated under a p -priority model where p is equal to 3. Note that the steady-state average packet delay for each priority class would be invariant to the intra-slot/intra-frame scheduling scheme. The queueing model analysis focuses on the average packet delay of each priority class regardless of intra-slot/intra-frame scheduling. The delay of a packet is defined as the total time spent by the packet to get through the UL channel. Denoted by D_k^j , the delay of the k^{th} packet of priority class j , is expressed as:

$$D_k^j = W_k^j + \sigma_k^j + P_k^j + F_L \quad (3.22)$$

where W_k^j denotes the packet waiting-time, σ_k^j represents the total time to transmit a priority- j k^{th} packet. F_L denotes the packet frame latency. P_k^j denotes the packet propagation delay. The average packet delay of j class is $\overline{D_j}$, given by [70]:

$$\overline{D_j} = W_{j,1} + \left\{ E[B_k^j] - \frac{M-1}{M} \right\} \left[\frac{M}{n_i} \right] \tau + F_L + P^j \quad (3.23)$$

where

$$W_{j,1} = \frac{\sum_{i=1}^J \lambda_i b_{i,2} + (1-\eta_j) \left\lfloor \frac{M}{n_i} \right\rfloor \tau}{2(1-\eta_j)(1-\eta_{j-1})} \quad (3.24)$$

$\left\lfloor \frac{M}{n_i} \right\rfloor$ represents the biggest integer not larger than $\frac{M}{n_i}$.

$b_{j,2}$ is the second moment of service time, and is given by:

$$b_{j,2} = \int_0^{\infty} x^2 dB_j(x) \quad (3.25)$$

$$\rho_j = \lambda_j b_{j,1}$$

$$\eta_j = \sum_{i=1}^j \rho_i \quad j=1, 2, \dots, J$$

For UGS traffic, $J = 1$. For PS traffic, $J = 2$. For BE traffic, $J = 3$.

3.4 Performance Evaluation

3.4.1 Simulation Design

Following the signaling mechanism specified by the IEEE 802.16d standard, the WiMAX simulation model has been developed by using ExtendSim [71] software. The scheduling algorithms have been developed by using C language. The network has been designed with one BS and eight SSs. The positions of the SSs are assumed to be independent and distributed identically. The distance from the BS to each SS is set to 5 km. Each SS has UGS, PS and BE service classes. The following assumptions are made for the traffic model.

- Connection is already established.
- The inter-arrival time of the UGS packets is constant. The packet length

is fixed to 200 bytes. The deadline is set to 50 ms.

- Every PS and BE connection is a burst traffic flow. Both PS and BE packet arrival is a Poisson process with the packet inter-arrival time exponentially distributed. Both PS and BE packet length is exponentially distributed. The mean packet length is set to 500 bytes. The deadline of the PS packet is exponentially distributed with a mean value of 80 ms. There is no deadline set to the BE packets.
- The frame size is set to 10 ms, 500 time slots per frame. The duration of each time slot is 20 μ s. The duration of an uplink sub-frame is pre-set to 5 ms. The time slots in UL sub-frame M is set to 250. The UL channel bandwidth C is pre-set to 20 Mbps.

3.4.2 Numerical Calculation of Average Packet Delay

According to the configuration of the WiMAX system mentioned above, $n_i = 250/8 = 31$ slots/frame as all SSs are identical. Noting that $\overline{B}_1 = 4$, $\overline{B}_2 = 10$ and $\overline{B}_3 = 10$ according to the average packet length of 3 different service classes. Table 3.1 shows the numerical results of the average packet delay of UGS, PS and BE classes.

The numerical results evaluate and compare with the simulation results of two-tier model, which refers to scheduling framework and algorithms in [21], three-tier model [72] and three-tier with DRR model.

TABLE 3.1 NUMERICAL CALCULATION OF MEAN DELAY

<i>Load</i>	<i>UGS</i> (<i>Mbps</i>)	<i>PS</i> (<i>Mbps</i>)	<i>BE</i> (<i>Mbps</i>)	λ_1 (<i>mean</i>) <i>Packet/s</i>	$\lambda_{2,3}$ (<i>mean</i>) <i>Packet/s</i>	<i>UGS</i> <i>Delay</i> (<i>ms</i>)	<i>PS</i> <i>Delay</i> (<i>ms</i>)	<i>BE</i> <i>Delay</i> (<i>ms</i>)
0.1	1.37	0.315	0.315	107.03	9.84	6.14	7.91	8.72
0.2	1.37	1.315	1.315	107.03	41.09	6.14	10.48	13.98
0.3	1.37	2.315	2.315	107.03	72.34	6.14	13.09	19.44
0.4	1.37	3.315	3.315	107.03	103.59	6.14	15.73	25.12
0.5	1.37	4.315	4.315	107.03	134.84	6.14	18.41	31.02
0.6	1.37	5.315	5.315	107.03	166.09	6.14	21.11	37.15
0.7	1.37	6.315	6.315	107.03	197.34	6.14	23.85	43.53
0.8	1.37	7.315	7.315	107.03	228.59	6.14	26.63	50.17
0.9	1.37	8.315	8.315	107.03	259.84	6.14	29.45	57.07

3.4.3 Experimental Results

Two sets of simulation experiments have been carried out to evaluate the performance of the three-tier with the DRR framework and proposed scheduling algorithms. The performances of the three scheduling schemes have been compared. The results show that the proposed scheduling scheme can not only effectively support real-time traffic to make real-time packets meet their delay bound but also provide fairness to non-real-time traffic and serve it with reasonable performance.

The simulation focuses on the effect of uplink scheduling algorithms on the traffic in the network. 1.37 Mbps bandwidth is set aside for the UGS service class.

The traffic load for the PS service class is 50% of the remaining traffic load. The major metrics for evaluating PS service class are packet loss rate, average packet delay and throughput. The major metric for evaluating BE service class is throughput.

3.4.3.1 Performance Evaluation by Simulation

In the first set of simulation, experiments are designed in which a real-time packet could be dropped if its deadline requirement cannot be met. The performance of two-tier, three-tier and three-tier with DRR models are evaluated in terms of packet loss rate and throughput of the PS connections, throughput of the BE connections and throughput of the UL channel. The simulation results are shown in Figure 3.3, 3.4, 3.5, and 3.6.

Figure 3.3 depicts the relationship between the packet loss rates of the PS traffic and different traffic loads. The packet loss rates of the three-tier with DRR model are much lower than that of two-tier model. The reasons are:

- The proposed DRR algorithm is able to allocate bandwidth to UL sub-channel and DL sub-channel based on the traffic pattern and loads.
- The PQLW scheduler can allocate more bandwidth to the PS connections when the PS traffic increases.
- The EDF scheduler gives more transmission opportunities to the packets with a tighter deadline in the PS connections; thus, the packet loss rate of the PS connections would not increase sharply while traffic load increases.

Figure 3.3 PS packet loss rate vs. traffic load for PS traffic

Figure 3.4 and Figure 3.5 show the performances of the throughputs of the PS service class and the throughputs of the BE service class with different traffic loads for three scheduling schemes respectively. It is observed that the throughputs of the PS service class by the proposed three-tier with DRR solution have been improved as compared to the two-tier model. The reason is the packet loss rates of the PS connections reduce dramatically by the contribution of the combination of DRR, PQLW and EDF schedulers.

It is observed that the throughputs of the BE service class of the proposed three-tier with DRR solution are slightly lower than that of the three-tier model. However, the throughputs of the PS service class of the proposed three-tier with DRR solution are higher than that of the three-tier model. The trade-off is set to ensure QoS provisioning to real-time applications.

Figure 3.4 PS throughput vs. traffic load

Figure 3.5 BE throughput vs. traffic load

Figure 3.6 shows the performances of the UL throughputs with different traffic loads for three scheduling schemes respectively. The three-tier with DRR model outperforms among the three scheduling schemes. The main reason is as follows. The bandwidths of the UL channel and DL channel have been fixed at 50% of the total network bandwidth in the two-tier model [21]. The Pre-scale Dynamic Resource Reservation (PDRR) algorithm is used in the three-tier model [72]. However, the proposed Dynamic Resource Reservation (DRR) scheme can allocate bandwidth to UL sub-frame and DL sub-frame dynamically in the three-tier with DRR scheme. The UL bandwidth is allocated proportionally to the ratio of UL traffic to the total traffic by the DRR algorithm. When the UL traffic load increases, the UL channel will be allocated more than 50% of the total bandwidth. As a result, the UL channel throughput increases when the traffic load increases gradually.

Figure 3.6 UL throughput vs. traffic load

3.4.3.2 Performance Comparison between Simulation and Theoretical Analysis

The second set of experiments are carried out to compare the performance of two-tier model, three-tier model and three-tier with DRR model with the result of theoretical queueing model without packet dropping due to a packet violating its time constraint. Comparisons of the performance on the average packet delays of the UGS, PS and BE service classes are shown in Figure 3.7, 3.8 and 3.9

Figure 3.7 shows the average packet delays of UGS of three scheduling schemes and the numerical results obtained by the queueing model.

Figure 3.7 UGS packet delay vs. traffic load

As the UGS traffic flows are allocated with a fixed bandwidth each according to their bandwidth requirements based on the QoS framework specified by the IEEE 802.16d standard, it is observed that average packet delays of three

scheduling schemes are very close to each other. It is also observed that the average packet delays of the UGS traffic flows obtained by the queueing model are similar to the simulation results of the three scheduling schemes. The simulation model has been validated by the queueing model. It shows that UGS is delay guaranteed in all scheduling schemes. The results comply with IEEE 802.16d standard.

Figure 3.8 shows the average packet delays of the PS connections of the three scheduling schemes based on simulation without packet dropping and the numerical calculation results of the queueing model.

Figure 3.8 PS packet delay vs. traffic load (model without packet dropping)

It is observed that the average packet delays of two-tier and three-tier scheduling schemes are very close to each other. The average packet delays of the three scheduling schemes are lower than the numerical result when the traffic

load is greater than 0.2.

When the traffic load is light, for example, less than 0.2, the signaling and scheduling overhead of the WiMAX system have a major contribution to the average packet delay in the simulation model. However, the signaling and scheduling overhead are not considered in the queueing model. The average packet delays calculated by the queueing model are lower than that in the simulation model. Thus, it is observed that the curve crosses over at around 0.2. When the traffic load increases, the signaling and scheduling overhead of the WiMAX system have a less impact on the average packet delay in the simulation model. The queueing delay and transmission delay have a major contribution to the average packet delay. It is observed that the theoretical results of the queueing model can serve as an upper bound for the PS traffic.

The average packet delays of the PS connections of the three-tier with DRR model are lower than that of two-tier model. Three-tier with DRR scheme has the best performance. The reasons are as follows. The bandwidth of UL channel and DL channel has been fixed at 50% of the total network bandwidth in two-tier model while Pre-scale Dynamic Resource Reservation (PDRR) algorithm is used in three-tier model [72]. However, UL bandwidth is allocated proportionally to the ratio of UL traffic to total traffic by DRR algorithm in three-tier with DRR scheme. When UL traffic load increases, UL channel will be allocated more bandwidth. As a result, the average packet delay of the PS connections increases slightly when traffic load increases gradually.

Figure 3.9 shows the average packet delays of the BE service class of the three scheduling schemes based on simulation without packet dropping, and the numerical calculation results of the queueing model.

It is observed that when the traffic load is light, e.g. less than 0.5, the average delays of BE of the three scheduling schemes are similar to the results obtained by the queueing model. They are close to each other. However, when traffic load increases, the average packet delay of the BE of the two-tier model increases sharply. This is because the overall bandwidth allocation is done in a strict priority way at level one scheduling in the two-tier model.

Figure 3.9 BE packet delay vs. traffic load

Since the priority of BE is lower than that of the PS service class, a less bandwidth is allocated to the BE service class. The three-tier model with DRR scheme has the best performance in terms of the average delay of the BE service class. The proposed DRR scheduler is to allocate the overall bandwidth to the DL and UL sub-channel dynamically according to the change of traffic loads from different SSs. Furthermore, the PQLW scheduling algorithm is applied to allocate the overall UL bandwidth excluding the bandwidth for UGS to PS and BE service

classes.

The proposed solution can improve the performance of the PS service class in terms of average packet delay, loss rate and throughput while provide reasonable performance for the BE service class. At least a small portion of bandwidth is allocated to the BE service class by PQLW scheduler. The starvation problem of the BE service class has been overcome effectively.

3.5 Summary

In this chapter, a three-tier with DRR QoS framework and scheduling schemes is proposed to enhance QoS support in WiMAX WirelessMAN-SC PMP systems. Associated with the framework, the proposed DRR and scheduling algorithms with hierarchical structure can effectively and efficiently provide QoS to the integrated traffic consisting of UGS, PS and BE service flows with or without time constraints. A queueing model is also developed to evaluate the performance of the proposal in the WiMAX system. The simulation and analytical studies show that the proposed solution can not only provide QoS to the UGS service class, improve the performance of the QoS for the PS service class in terms of packet loss rate, packet delay and throughput but also overcome the starvation problem of the BE service class.

In the next chapter, the works on dynamic opportunistic scheme, Adaptive Power Control scheme are presented to further enhance QoS support in WiMAX WirelessMAN-SC PMP systems with cross-layer optimization design techniques.

Chapter 4 Cross-Layer MAC Protocol and Holistic Opportunistic Scheduling with APC Scheme

Wireless access networks have unique characteristics, which are time-varying channel conditions and multi-user diversity. The medium access control protocol and scheduling solutions have to be developed specially for this environment [45]. Opportunistic MAC (OMAC) [45] is a novel view of communication over spatiotemporally varying wireless link, whereby the multi-user diversity is exploited rather than combated to maximize bandwidth efficiency or system throughput. It requires a cross-layer MAC protocol design approach. Protocols can be designed by violating the reference architecture, for example, by allowing direct communication between protocols at nonadjacent layers or sharing variables between layers. Such violation of a layered architecture is a cross-layer design with respect to the reference architecture [36].

In cross-layer MAC protocol, Channel Specification (Cspec) carrying the estimated instantaneous channel information can be fed to the MAC layer from the physical (PHY) layer and Traffic Specification (Tspec) carrying traffic QoS

related information can be fed to the MAC layer from higher layers such as the network or application layer. The Cspec feedback includes the information on the estimated instantaneous Signal-to-Interference and Noise Ratio (SINR), supportable data rates $R(t)$, Received Signal Strength Indications (RSSI), and Bit Error Rates (BER) of a link. The Tspec feedback includes the information on traffic maximum latency (ML) constraint and maximum sustained traffic rate (MSTR) and the instantaneous length of queues at a station. The cross-layer nature of OMAC with opportunistic scheduling scheme and Adaptive Power Control (APC) scheme has the potential to revolutionize the design of broadband wireless access networks from the physical to the networking layer.

In paper [73], the authors proposed a cross-layer system architecture to support QoS of video streaming services over the mobile WiMAX system. It was shown that the implementation of multiple connections with feedback information on the available transmission bandwidth is critical for supporting scalable video streaming, in which the transmission packets can be further separated into multiple levels of importance. In paper [74], a heuristic cross-layer mechanism was proposed to improve the performance of real-time applications in IEEE 802.16 networks. The main idea of the mechanism is to coordinate the adaptive modulation capability of the physical layer and the multi-rate data-encoding capability of modern real-time applications in order to avoid inefficiencies caused by their independent operation. In paper [75], a fair and QoS guaranteed scheduling approach with fuzzy controls (FQFCs) was proposed for WiMAX OFDMA systems. The proposed controllers adjust the priority and transmission opportunity for each WiMAX connection according to its QoS requirement and service class. In paper [76], the authors proposed an optimized scheduling scheme

in OFDMA-based WiMAX networks. The authors suggested that the optimized resource allocation can be processed first under different constraints to achieve different performance metrics, then different traffic sessions at each user can be scheduled further with respect to a proper admission control mechanism.

As discussed in Chapter 2.3.3, cross-layer QoS support mechanisms and opportunistic scheduling designs tailored for WiMAX have been proposed [54-57].

However, the existing scheduling schemes proposed for WiMAX systems only take channel characteristics and one of the traffic characteristics like delay or queue length into account. Furthermore, the priority coefficient setting for rtPS and nrtPS traffic connections is rigid while in fact, the traffic load and average long-term channel conditions vary dynamically.

In this chapter, with a cross-layer optimizing design approach, a cross-layer MAC protocol and QoS support framework associated with a two-stage opportunistic scheduling scheme and an adaptive power control scheme are proposed to provide QoS support to the heterogeneous traffic in single carrier WiMAX PMP systems. The two-stage opportunistic scheduling algorithm is termed as Holistic Opportunistic Scheduling (HOS) with features of channel-awareness, queue-awareness and traffic QoS-awareness. The proposed Adaptive Power Control (APC) scheme can optimize transmission power and keep the data transmission rate for rtPS traffic flows at a higher level.

This chapter is organized as follows. In Section 4.1, the graph of BER vs. SINR is plotted. A lookup table reflecting AMC vs. SINR as well as receiver sensitivity (RSS) requirement is tabulated. In Section 4.2, the proposed cross-layer MAC protocol and QoS support framework are described and elaborated. In

Section 4.3, the HOS algorithm is elaborated in detail. In Section 4.4, APC scheme is illustrated. Section 4.5 shows the simulation model design and the simulation results. Finally, Section 4.6 gives the summary of the chapter.

4.1 AMC Schemes vs. SINR in WiMAX

WiMAX transceivers support different transmission modes with different modulation and coding schemes corresponding to different data transmission rates. For each modulation scheme, there is one relationship between the theoretical Bit Error Rate (BER) and E_b/N_0 . E_b/N_0 is classically defined as the ratio of Energy per Bit (E_b) to the Spectral Noise Density (N_0). E_b/N_0 is one type of measurement of signal to noise ratio for a digital communication system. Different forms of modulation -- BPSK, QPSK, QAM, etc. have different relationships of theoretical BER versus E_b/N_0 . Table 4.1 lists BER versus E_b/N_0 equations corresponding to various modulation schemes [77].

TABLE 4.1 BER EQUATION VS. E_b/N_0 FOR VARIOUS MODULATION SCHEMES

<i>Modulation Scheme</i>	<i>BPSK</i>	<i>QPSK</i>	<i>16-QAM</i>	<i>64-QAM</i>
<i>BER</i>	$\frac{1}{2} \operatorname{erfc} \left[\frac{E_b}{N_0} \right]^{\frac{1}{2}}$	$\frac{1}{2} \operatorname{erfc} \left[\frac{E_b}{N_0} \right]^{\frac{1}{2}}$	$2 \operatorname{erfc} \left[0.4 \frac{E_b}{N_0} \right]^{\frac{1}{2}}$	$2 \left[1 - \frac{1}{6} \right] \operatorname{erfc} \left[\left(\frac{3}{6^2 - 1} \right) \frac{E_b}{N_0} \right]^{\frac{1}{2}}$

In order to derive the relationship among modulation and coding schemes, the corresponding requested SINR and the corresponding achievable data rate at desirable BER in WiMAX WirelessMAN-SC PMP systems, the relationship of BER vs. E_b/N_0 is first converted into the relationship of BER vs. SINR at the

receiver using the bit rate R_b . The bit rate R_b reflects the different modulation and coding schemes adapted at certain bandwidth W . Then, the graph of theoretical BER vs. SINR relationship for different modulation and coding schemes is plotted. After that, the range of the received SINR values will be classified into seven non-overlapping regions for adaptive modulation and coding index $AMC(q)$ where $q=1,2,\dots,7$ on a targeted prescribed BER e.g. $BER \leq 10^{-5}$ in the WiMAX system. Further complying with the receiver sensitivity RSS requirement specified by IEEE 802.16d standard [1], a lookup table can be tabulated to show the relationship among AMC, SINR and RSS requirement with its corresponding channel data rate. The details of derivation are as follows.

E_b/N_0 in BER equations in Table 4.1 can be expressed in term of $SINR$ as [77]:

$$\frac{E_b}{N_0} = SINR \times \frac{W}{R_b} \quad (4.1)$$

where W is the channel bandwidth, R_b is the transmission bit rate of a transmission mode corresponding to a modulation and coding scheme. Substituting (4.1) into BER equations in Table 4.1, the theoretical BER vs. SINR relationship for different modulation and coding schemes with different channel bandwidth W and transmission bit rate R_b can be obtained.

The channel bandwidth versus transmission bit rate for different modulation schemes can be obtained according to IEEE 802.16d standard [1] as shown in Table 4.2. For example, if channel bandwidth W is 25 MHz, the corresponding channel data rate can be 40 Mbps, 80 Mbps and 120 Mbps according to QPSK, 16-QAM and 64-QAM modulation scheme, respectively. R_b for each modulation and coding scheme can be calculated when channel bandwidth W is 25 MHz.

TABLE 4.2 CHANNEL BANDWIDTH VS. BIT RATE WITH VARIOUS MODULATION
SCHEMES

<i>Channel Bandwidth (MHz)</i>	<i>Bit Rate (Mbps) (QPSK)</i>	<i>Bit Rate (Mbps) (16-QAM)</i>	<i>Bit Rate (Mbps) (64-QAM)</i>
20	32	64	96
25	40	80	120
28	44.8	89.6	134.4

Substitute W and R_b into (4.1) and further substitute (4.1) into BER equations corresponding to different modulation schemes in Table 4.1, the graph of theoretical BER vs. SINR relationship is plotted by using MATLAB [78] when PHY channel bandwidth is set to 25 MHz in single carrier WiMAX PMP systems as shown in Figure 4.1.

Figure 4.1 Theoretical BER vs. SINR relationship for various AMC schemes

Over a communication link, the transmission from node i to the corresponding receiver node j can be successful at a data transmission rate by using a modulation and coding scheme at a certain targeted BER only when the communication link meets the following requirements:

- The received SINR at node j , $SINR_{ij}$, is greater than or equal to the minimum required SINR threshold γ_q of the receiver for the corresponding modulation and coding scheme.
- The received signal strength RSS at node j is above the minimum receiving sensitivity (dBm) requirement of the receiver for the corresponding modulation and coding scheme.

Based on Figure 4.1, when BER is kept below a certain value e.g. $BER < 1 \times 10^{-5}$, a range of SINR thresholds can be obtained for each modulation and coding scheme while each modulation and coding scheme can provide a corresponding channel's data transmission rate.

The receiver sensitivity of the BS, which is the threshold of the receiving signal strength, changes with various modulation schemes used and data rates for the transmission. Table 337 in IEEE 802.16d standard [1] lists the receiver minimum input level sensitivity (dBm) for different modulation and coding schemes with various bandwidths from 1.5 MHz to 20 MHz. The bandwidth is set to 25 MHz in the simulation model in this chapter. According to the standard, the receiver minimum input level sensitivity for different modulation and coding schemes with bandwidth 25 MHz can use the closest values of the receiver minimum input level sensitivity, which is the receiver minimum input level sensitivity for bandwidth set to 20 MHz from Table 337 in IEEE 802.16d

standard. Table 4.3 shows the receiver minimum input level sensitivity (dBm) for different modulation and coding schemes with bandwidth set to 25 MHz.

TABLE 4.3 RECEIVER MINIMUM INPUT LEVEL SENSITIVITY (dBm)

<i>Bandwidth</i>	<i>BPSK</i>	<i>QPSK</i>		<i>16-QAM</i>		<i>64-QAM</i>	
<i>(MHz)</i>	1/2	1/2	3/4	1/2	3/4	2/3	3/4
25	-80	-80	-78	-73	-71	-66	-65

Combining Figure 4.1 and Table 4.3, a lookup table can be derived as shown in Table 4.4 to show the relationship among AMC, SINR and receiver sensitivity RSS requirement as well as its corresponding channel data rate when BER is set to 1×10^{-5} . It is clear that the data transmission rate can be dynamically changed from 20 Mbps to 180 Mbps.

TABLE 4.4 AMC VS. RECEIVER SINR & MIN SENSITIVITY REQUIREMENT

<i>AMC</i> <i>Index</i> <i>AMC(q)</i>	<i>Receiver</i> <i>SINR(dB)</i> $\gamma(q)$	<i>Receiver</i> <i>Min Sensitivity</i> <i>RSS(q) (dBm)</i>	<i>Modulation/</i> <i>Coding</i> <i>Scheme</i>	<i>Channel Bit</i> <i>Rate</i> <i>R(q) (Mbps)</i>
<i>AMC(1)</i>	8.3-11.6	- 80	BPSK1/2	20
<i>AMC(2)</i>	11.7-13.2	- 80	QPSK1/2	40
<i>AMC(3)</i>	13.3-18.9	- 78	QPSK3/4	60
<i>AMC(4)</i>	19.0-21.9	- 73	16-QAM1/2	80
<i>AMC(5)</i>	21.0-28.0	-71	16-QAM3/4	120
<i>AMC(6)</i>	28.1-29.1	- 66	64-QAM2/3	160
<i>AMC(7)</i>	≥ 29.2	-65	64-QAM3/4	180

In summary, each data transmission rate corresponds to a SINR requirement and a receiver minimum sensitivity RSS requirement. A high data rate requires a high SINR and a high receiver sensitivity RSS in order to keep the BER under certain level. For a particular link, the receiver's SINR and sensitivity RSS depend on the transmitter's transmission power as well as interference resulting from other links' transmission and noise. Higher SINR and higher receiver sensitivity RSS require higher transmission power under the same channel condition.

4.2 Proposed MAC-PHY Cross-Layer Framework and MAC Protocol

As shown in Figure 2.2, IEEE 802.16d frames out AMC and the conceptual power control scheme. It left the details of the AMC and the adaptive power control algorithm undefined. IEEE 802.16d MAC frames out class-based QoS guarantees and per-flow resource assignment. It left the details of the QoS based packet scheduling algorithms undefined. This chapter focuses on the cross-layer OMAC with an opportunistic scheduling scheme and an Adaptive Power Control (APC) scheme. The cross-layer OMAC has the potential to revolutionize the design of broadband wireless access networks from the physical to the networking layer.

The performance of the MAC protocol and the functionality of the scheduling algorithm are influenced by the time-varying wireless channel in WiMAX systems. With the cross-layer approach, some new interfaces need to be designed.

- Considering the impact of air interface on scheduling algorithm, a new

interface of AMC_Controller _Info needs to be designed to link up the AMC scheme and the scheduler from the PHY to the MAC layer.

- APC can increase the transmission power hence it might cause interference in the system. A new interface needs to be designed to link up APC, AMC and the scheduler so that APC can perform its functions effectively.

With a cross-layer approach, the proposed QoS provisioning architecture aims at providing scheduling modules, AMC and APC modules to complete the missing parts in the QoS architecture specified in the IEEE 802.16d standard as shown in Figure 2.2. An efficient cross-layer MAC protocol and QoS support framework with a two-stage opportunistic scheduling algorithm with APC are proposed as shown in Figure 4.2.

Figure 4.2 Proposed cross-layer MAC protocol with QoS support framework

The proposed cross-layer MAC protocol with QoS support framework supports both UL and DL transmission. The UL transmission is used as an example to explain its functionality as follows.

1. Considering the impact of air interface on MAC layer protocol, a wireless Channel Condition Estimator (CCE) is designed at the PHY layer at the BS (as well as SS_m , where $m=1,2,3,\dots,M$. M is the number of SSs in the WiMAX system). CCE monitors instantaneous channel condition status like its transmitter transmission power $P(BS, m)$ to transmit its PHY symbols to SS_m , as well as the receiver received signal strength $RSS(m, BS)$ and SINR $\gamma(m, BS)$ when the BS receives signal from SS_m . At the same time, CCE also indicates long term channel condition by using the statistic parameters including $Max\_ \gamma(m, BS)$, $Average\_ \gamma(m, BS)$ and $Min\_ \gamma(m, BS)$ over the execution period. Based on the information of a channel's status provided by CCE, FEC, Symbol Mapper and AMC controller at the BS select a FEC scheme, modulation and coding scheme $AMC(m, BS, q)$ (hence the data transmission rate) for UL data transmission from SS_m to the BS at modulation and coding level q according to Table 4.4. According to the algorithm described later in Section 4.4, APC computes a feasible requested SS_m transmission power $Req\_P(m, BS, q)$ to fulfill both $\gamma(m, BS)$ and receiver sensitivity $RSS(m, BS)$ requirements of $AMC(m, BS, q)$ at the BS. By sensing the maximum value of q in $AMC(m, BS, q)$ at all M SSs in the system, Symbol Mapper and AMC controller at the BS can determine the maximum data transmission rate $Max\_R(t)$ at time t in the system.

The PHY Symbol Rate Controller at the BS then forms $AMC\_Controller\_Info\{P(BS, m), \gamma(m, BS), RSS(m, BS), AMC(m, BS, q),$

- $Req\_P(m, BS, q), Max\_R(t)\}$. The $AMC\_Controller\_Info\{P(BS, m), \gamma(m, BS), RSS(m, BS), AMC(m, BS, q), Req\_P(m, BS, q), Max\_R(t)\}$ is forwarded to MAC layer via its PHY Service Access Point (SAP) for scheduling and APC. Similarly, the PHY Symbol Rate Controller at SS_m forms $AMC\_Controller\_Info\{P(m, BS), \gamma(BS, m), RSS(BS, m), AMC(BS, m, q), Req\_P(BS, m, q)\}$ and forwards it via its PHY SAP for scheduling and APC.
2. The first stage of scheduling scheme operates at each SS. The scheduler at SS_m uses the Tspec of each connection and $AMC\_Controller\_Info\{P(BS, m), \gamma(m, BS), RSS(m, BS), AMC(m, BS, q), Req\_P(m, BS, q), Max\_R(t)\}$ to compute four scheduling parameters which are to determine a dynamic Scheduling Priority (SP) for every packet in each connection by the proposed HOS algorithm. Then, the Scheduling Priority (SP) will be sorted in descending order. Then SS_m takes the maximum value of the calculated SP, $Max\_SP(m)$, to represent itself to compete with other SSs for the order of UL transmission. SS_m embeds its $Max\_SP(m)$, its total $BW\_request(m)$ and $AMC\_Controller\_Info\{P(m, BS), \gamma(BS, m), RSS(BS, m), AMC(BS, m, q), Req\_P(BS, m, q)\}$ in a $Polling\_response(m)$ and sends the $Polling\_response(m)$ to the BS when SS_m is polled.
  3. The second stage of scheduling scheme at the BS sorts $Max\_SP(1), Max\_SP(2), \dots, Max\_SP(M)$ in descending order and selects SS_m with the highest $Max\_SP(m)$ to allocate time slots according to its $BW\_request(m)$ in an UL sub-frame excluding the time slots for UGS. If $BW\_request(m)$ is less than the number of available time slots in the UL sub-frame, the remaining available time slots will be allocated to SS n with the second highest $Max\_SP(n)$ according to its $BW\_request(n)$. This procedure iterates until all

the available time slots in the UL sub-frame are used up. The $AMC\_Controller\_Info\{P(BS, m), \gamma(m, BS), RSS(m, BS), AMC(m, BS, q), Req\_P(m, BS, q), Max\_R(t)\}$ and the result of scheduling, which is the UL sub-frame time slots allocations for SSs, are embedded in a DL/UL_Map. The DL/UL_Map is broadcasted by the BS to all SSs for UL transmission.

4. The scheduler at SS_m extracts the information of its allocated transmission time slots and the $AMC\_Controller\_Info\{P(BS, m), \gamma(m, BS), RSS(m, BS), AMC(m, BS, q), Req\_P(m, BS, q)\}$ from the DL/UL_Map and selects packets from the highest Scheduling Priority to the lowest Scheduling Priority from different connections. SS_m transmits the selected packets to the BS in the allocated time slots at a requested transmission power $Req\_P(m, BS, q)$ and at the data rate determined by $AMC(m, BS, q)$ of BS receiver. When SS_m transmits a packet from rtPS connections, the scheduler will start APC to increase its transmission power in order to improve the received signal $\gamma(m, BS)$ and $RSS(m, BS)$ at the BS receiver if one of the following conditions is true:
  - normalized time-delay satisfaction index $NTDSI(m, i, k, t)$ of the packet is close to 1 or;
  - q in $AMC(m, BS, q)$ is less than 7, meaning that the transmission data rate of SS_m is not at the maximum rate.

This results in the Symbol Rate Controller at the BS using a higher q level of modulation and coding scheme $AMC(m, BS, q)$. SS_m thus can transmit packets belonging to rtPS service class at a higher data transmission rate to prevent the real-time packet violating its deadline.

4.3 Holistic Opportunistic Scheduling Algorithm

As mentioned in Chapter 2.3.3, the proposed opportunistic scheduling schemes in [54-57] were shown to achieve satisfied performance in the given network conditions with the help of several additional parameters for the preference metrics, which are the priority of connections or users to be determined by the opportunistic scheduler for its transmission service.

However, the existing scheduling schemes proposed for WiMAX systems only take channel characteristics and one of traffic characteristics like delay or queue length into account. Some common performance metrics such as packet loss rate have not been evaluated. The scheduling schemes have not been applied to packet scheduling in each traffic flow. The studies have not explored the metrics of throughput and delay under different traffic load. In addition, the priority coefficient setting for rtPS and nrtPS traffic connections is rigid while in fact, traffic load and average long-term channel conditions vary dynamically.

Associated with the cross-layer MAC protocol and QoS support framework, the Holistic Opportunistic Scheduling algorithm is proposed. The proposed scheduling algorithm takes the information of instantaneous wireless channel condition and the real-time traffic condition as well as traffic QoS specification to set the Scheduling Priority (SP) for each individual packet from rtPS or nrtPS service classes for the UL transmission. Since the UGS connections have been allocated with fixed bandwidth (fixed time duration) based on their fixed bandwidth requirement in the framework of the IEEE 802.16d standard, the proposed scheduling is only applied to the rtPS and nrtPS service classes. The Scheduling Priority of each packet is determined by four key scheduling

parameters, namely Dynamical Priority Index (DPI), Channel Specification Index (CSI), Normalized Time-delay Satisfaction Index (NTDSI) and Normalized Predictive Starvation Index (NPSI). In the scheduling period starting at time t , SS_m (or BS scheduling for DL traffic) decides the scheduling priority (SP) of the k^{th} packet from connection i of either rtPS or nrtPS traffic flows based on the following equation:

$$SP(m, i, k, t) = \{DPI(m, i, t) \times CSI(m, i, t) \times \exp[NTDSI(m, i, k, t)] \times NPSI(m, i, t)\} \quad (4.2)$$

where $DPI(m, i, t)$ is the dynamic priority index of connection i at SS_m . One of the typical WiMAX features is class-based QoS that provides differentiated traffic treatment. WiMAX system defines four different priorities of service classes where by the priority of UGS > the priority of rtPS > the priority of nrtPS > the priority of BE. The role of $DPI(m, i, t)$ is to provide different priorities for different QoS classes dynamically. The priority coefficients were fixed, the priority of rtps = 1.0 > the priority of nrtps = 0.8 > the priority of BE = 0.6, in paper [56]. In contrast, the priority index (priority coefficient in paper [56]) of service classes is set dynamically on a channel-aware approach in my proposal. The procedure to set the value of $DPI(m, i, t)$ is described as follows.

1. A range of [1.0, 0.8] for $DPI(m, i, t)$ is set up for rtPS traffic connections. The SS firstly assigns 0.8 to $DPI(m, i, t)$ for rtPS traffic connection i and 0.6 to $DPI(m, j, t)$ for nrtPS traffic j connection.
2. After the system runs over a period of time, SS_m is able to derive statistical channel condition parameters like $Max_{\gamma}(m, BS)$, $Average_{\gamma}(m, BS)$ and $Min_{\gamma}(m, BS)$ based on $\gamma(m, BS)$ in $AMC\_Controller\_Info\{P(BS, m), \gamma(m, BS), RSS(m, BS), AMC(m, BS, q), Req\_P(m, BS, q), Max\_R(t)\}$ from the BS.

The $DPI(m, i, t)$ for each packet belonging to rtPS service class can be updated dynamically according to the linear function:

$$DPI(m, i, t) = 0.8 + 0.2 \times [\gamma(m, BS) - Min\_ \gamma(m, BS)] \wedge [Max\_ \gamma(m, BS) - Min\_ \gamma(m, BS)] \quad (4.3)$$

By using (4.3), $DPI(m, i, t)$ is set up dynamically from 0.8 to 1.0 for rtPS service class based on channel SINR $\gamma(m, BS)$, while it is fixed 0.6 for nrtPS service class. It makes sure that the priority of rtPS traffic is always higher than that of nrtPS traffic. It gives favor treatment to rtPS service class to enhance the performance of QoS for real-time traffic connections when the channel condition is good.

$CSI(m, i, t)$ in (4.2) is the channel specification index for connection i at SS_m at time t . It calculates as:

$$CSI(m, i, t) = \frac{R(m, i, t)}{Max\_ R(t)} \quad (4.4)$$

where $R(m, i, t)$ is the data transmission rate of connection i at SS_m at time t . It is set according to the $AMC(m, BS, q)$ dynamically based on the channel condition at PHY layer. $Max\_ R(t)$ is the maximum of $R(m, i, t)$ over all m SSs and all of connection $i, i=1, 2, 3, \dots, I$ at time instant t .

Based on the definition of the channel specification index defined in (4.4), $CSI(m, i, t)$ is in fact a MCS[47] scheduling scheme. $CSI(m, i, t)$ is the prototype of throughput efficient opportunistic scheduling. This policy is to allocate resources to SSs with the best channel condition. In theory, $CSI(m, i, t)$ policy results in a high system throughput. In order to achieve maximum system throughput, $CSI(m, i, t)$ plays a role as a channel-awareness scheduling parameter

in the proposed HOS. However, this channel-aware policy monopolizes all resources to good-channel users that are usually located close to the Base Station. Because the SSs are usually distributed across the entire WiMAX system, this unfair scheduling parameter alone is not beneficial for anybody but local SSs.

In order to overcome this unfair problem caused by the side effect of $CSI(m, i, t)$ and to provide delay bound guarantees for individual connection in order to support delay-sensitive rtPS traffic, two new scheduling parameters are introduced in the HOS scheduling function. One is the traffic QoS-aware scheduling parameter that is termed as normalized time-delay satisfaction index, $NTDSI(m, i, k, t)$. Another one is the queue-aware scheduling parameter that is termed as normalized predictive starvation index, $NPSI(m, i, t)$. They are explained in detail as follows.

$NTDSI(m, i, k, t)$ in (4.2) is the normalized time-delay satisfaction index. This parameter is to capture packet latency requirement if it is applicable. It is defined as:

$$NTDSI(m, i, k, t) = 1 - TDSI(m, i, k, t) / \underset{i \in I}{Max}(TDSI(m, i, k, t)) \quad (4.5)$$

Packet timeout is defined as the waiting time of the packet in a queue is over its maximum latency $ML(m, i, k)$ or if its waiting time is over the TCP retransmission threshold. Time-delay satisfaction index (TDSI) for the k^{th} packet from rtPS connection i at SS_m , $TDSI(m, i, k, t)$, is set as:

$$TDSI(m, i, k, t) = \begin{cases} ML(m, i, k) - [t - VC(m, i, k)] \\ 0, & \text{if } \{ML(m, i, k) - [t - VC(m, i, k)]\} \leq 0 \end{cases} \quad (4.6)$$

$ML(m,i,k)$ is the maximum latency of the k^{th} packet. Each packet is stamped with a virtual clock (VC) [79]. VC monitors the average transmission rate of statistical data flows and provides every flow with guaranteed throughput and low queue delay. The idea behind the VC algorithm is that each packet is assigned a VirtualTime, which represents when it would be sent in a time division multiplexing (TDM) system. The virtual clock $VC(m, i, k)$ is set for the k^{th} packet from connection i at SS_m according to the algorithm:

$$VC(m, i, k) = \text{Max}[VC(m, i, k-1), \text{real time}] + PL(m, i, k)/R(m, i, t-1) \quad (4.7)$$

where $R(m, i, t-1)$ is the transmission rate of connection i at SS_m at time $(t-1)$. $PL(m, i, k)$ is the length of the k^{th} packet. Since a non-real-time packet in nrtPS connections has no maximum latency requirement, the maximum waiting time for a non-real-time packet is TCP retransmission time-out value TO which is 400 ms for wireless channel [80]. The $TDSI(m, i, k, t)$ for the k^{th} packet from nrtPS connection i at SS_m , $TDSI(m, i, k, t)$, is set as:

$$TDSI(m,i,k,t) = \begin{cases} TO - [t - VC(m,i,k)] \\ 0, \text{ if } \{TO - [t - VC(m,i,k)]\} \leq 0 \end{cases} \quad (4.8)$$

$NPSI(m, i, t)$ in (4.2) is the normalized predictive starvation index. It is calculated as follows:

$$NPSI(m,i,t) = \frac{PSI(m,i,t)}{\text{Max}_{i \in I}(PSI(m,i,t))} \quad (4.9)$$

where $PSI(m, i, t)$ is the predictive starvation index (PSI) which indicates queue status in terms of length of a queue and urgency of a packet in the queue that is to

be transmitted successfully.

$$PSI(m, i, t) = \text{the number of packets in queue } i \text{ at } SS_m \text{ at time } t / \overline{PSR(m, i, t)} \quad (4.10)$$

where $\overline{PSR(m, i, t)}$ is average packet success rate. It is defined as follows:

A packet is correctly transmitted from connection i at SS_m at time t , only if it is not dropped from the queue with probability $[1 - P_d(m, i, t)]$ and is correctly transmitted through the wireless channel with probability $[1 - PER(m, i, t)]$. Hence, the average packet loss rate for connection i at SS_m at time t is expressed as [81]:

$$\overline{\varepsilon(m, i, t)} = 1 - [1 - \overline{P_d(m, i, t)}][1 - PER(m, i, t)] \quad (4.11)$$

So average packet success rate $\overline{PSR(m, i, t)}$ for the k^{th} packet from connection i at SS_m at time t is calculated as:

$$\overline{PSR(m, i, t)} = 1 - \overline{\varepsilon(m, i, t)} = [1 - \overline{P_d(m, i, t)}][1 - PER(m, i, t)] \quad (4.12)$$

Let $\overline{P_d(m, i, t)}$ denote the average packet dropping (due to the queue being full, the packet delay exceeding its ML or waiting longer than TCP re-transmission threshold) probability of the queue for connection i at SS_m at time t .

$$\overline{P_d(m, i, t)} = \frac{E_{m,i}\{D\}}{E_{m,i}\{A\}} \quad (4.13)$$

$\overline{P_d(m, i, t)}$ is the ratio of the average number of dropped packets $E_{m,i}\{D\}$ over the average number of arrived packets $E_{m,i}\{A\}$ during the sliding window or scheduling period.

To simplify the AMC design, packet error rate ($PER(m, i, t)$) of wireless PHY

transmission channel for connection i at SS_m at time t can be approximately expressed as [81][82]:

$$PER(m, i, t) \approx \begin{cases} 1, & \text{if } \gamma(m, BS) \leq \gamma_{BS\_Min} \\ a_n \exp(-g_n \gamma(m, BS)) & \text{if } \gamma(m, BS) \geq \gamma_{BS\_Min} \end{cases} \quad (4.14)$$

γ_{BS_Min} is the minimum SINR requirement of the BS receiver. Parameters a_n , g_n are dependent on the modulation and coding mode. They can be determined by the modulation and coding scheme.

4.4 Adaptive Power Control Scheme

Associated with the proposed MAC protocol and Holistic Opportunistic Scheduling scheme, the APC scheme has been designed for a SS or the BS to perform two functions. The first function of APC is to optimize requested transmission power $Req\_P(m, BS, q)$ (or $Req\_P(BS, m, q)$ of the BS for DL transmission). The second function of APC is power control function. The function is to increase the transmission power $P(m, BS)$ of SS_m in order to improve the received signal $\gamma(m, BS)$ and $RSS(m, BS)$ at the BS receiver (or $P(BS, m)$ of the BS for DL transmission) when source SS_m transmits packets from rtPS service class to the BS. With the APC, source SS_m can transmit rtPS packets at a higher data transmission rate to improve QoS for real-time traffic in UL transmission.

4.4.1 The First Function of APC: Requested Transmission Power Optimization

At the PHY layer, the packet is sent successfully from source SS_m to the BS only if

- the power of transmitter at source SS_m is high enough to make received signal $\gamma(m, BS)$ at the BS greater than or equal to the minimum SINR requirement threshold, e.g. $\gamma(I)$; and
- the received signal strength $RSS(m, BS)$ is higher than the receiver sensitivity $RSS(I)$ (dBm) at the BS as shown in Table 4.4.

The received SINR $\gamma(m, BS)$, which is detected by CCE at the BS when the BS receives messages from source SS_m , is expressed as:

$$\gamma(m, BS) = \frac{P(m, BS)G(m, BS)}{N_0 + \sum_{\substack{k=1 \\ k \neq m}}^M P(k, BS)G(k, BS)}$$

$$m=1, 2, 3, \dots, M, \quad k=1, 2, 3, \dots, M, \quad k \neq m \quad (4.15)$$

where $P(m, BS)$ denotes the transmission power of source SS_m , $G(m, BS)$ denotes the link gain from source SS_m to the BS, which depends on the distance between SS_m and the BS, antenna gains, system loss, etc. N_0 denotes the thermal noise of the BS.

Let the average link gain from source SS_m to the BS be:

$$\overline{G(m, BS)} = \frac{G(m, BS)}{N_0 + \sum_{\substack{k=1 \\ k \neq m}}^M P(k, BS)G(k, BS)} \quad (4.16)$$

Substituting (4.16) into (4.15), $\gamma(m, BS)$ is expressed as:

$$\gamma(m, BS) = \overline{G(m, BS)}P(m, BS) \quad (4.17)$$

As the WiMAX system under study is in PMP topology, the distance between SS_m and the BS is fixed. It is assumed that the wireless transmission channel varies from frame to frame so that the link gain $\overline{G(m,BS)}$ is constant over a frame. Then, equation (4.17) states that $\gamma(m,BS)$ is directly proportional to the transmission power $P(m,BS)$ over a frame under the same transmission path in the WiMAX PMP system. Let $P_{\gamma(q)}(m,BS)$ be the transmission power of source SS_m that can achieve SINR $\gamma(m,BS,q)$ corresponding to $AMC(q)$ over a frame as shown in Table 4.4. So, $P_{\gamma(1)}(m,BS)$ is the transmission power of source SS_m that can achieve $\gamma(m,BS,1)$ corresponding to $AMC(1)$. By using (4.17), the relationship of $P_{\gamma(q)}(m,BS)$ and $P_{\gamma(1)}(m,BS)$ over a frame can be expressed as:

$$P_{\gamma(q)}(m,BS) = P_{\gamma(1)}(m,BS) \times \frac{\gamma(m,BS,q)}{\gamma(m,BS,1)} \quad (4.18)$$

The $RSS(m,BS)$ is a measurement on the power level of a received radio signal at the BS receiver by the given transmission power $P(m,BS)$ to the signal transmitted from source SS_m transmitter. Theoretically, $RSS(m,BS)$ can be expressed as Friis Transmission Equation [83]:

$$RSS(m,BS) = P(m,BS) \times AG(m) \times AG(BS) \left(\frac{\lambda}{4\pi d} \right)^2 \quad (4.19)$$

where $AG(m)$ and $AG(BS)$ are the antenna gain of the transmitting SS_m and receiving BS antennas, respectively. λ is the radio wave length, d is the distance between the BS and SS_m .

Let

$$\overline{AG(m,BS)} = AG(m) \times AG(BS) \left(\frac{\lambda}{4\pi d} \right)^2 \quad (4.20)$$

Substituting (4.20) into (4.19), $RSS(m,BS)$ is expressed as:

$$RSS(m, BS) = \overline{AG(m, BS)} \times P(m, BS) \quad (4.21)$$

Similarly, let $P_{RSS(q)}(m, BS)$ be the transmission power of source SS_m , which can achieve $RSS(m, BS, q)$ at the BS corresponding to $AMC(q)$ over a frame as shown in Table 4.4. So, $P_{RSS(1)}(m, BS)$ is the transmission power of source SS_m which can achieve $RSS(m, BS, 1)$ corresponding to $AMC(1)$. By using (4.21), the relationship of $P_{RSS(q)}(m, BS)$ and $P_{RSS(1)}(m, BS)$ over a frame can be expressed as:

$$P_{RSS(q)}(m, BS) = P_{RSS(1)}(m, BS) \times \frac{RSS(m, BS, q)}{RSS(m, BS, 1)} \quad (4.22)$$

In order to maximize the efficiency of the transmission power of the SSs in the WiMAX system, a higher q level modulation and coding schemes $AMC(m, BS, q)$ should be chosen by giving a certain amount of transmission power to the signal transmitted from source SS_m transmitter. The first function of APC scheme at the BS is to compute an optimal requested transmission power $Req\_P(m, BS, q)$ of source SS_m to fulfill both $\gamma(m, BS)$ and $RSS(m, BS)$ requirements at the BS receiver when q level modulation and coding scheme $AMC(m, BS, q)$ is selected. This principle can be applied for DL channel transmission similarly.

The operation of optimizing the requested transmission power is described as follows:

Corresponding to adaptive modulation and coding scheme $AMC(m, BS, 1)$, $AMC(m, BS, 2)$, $AMC(m, BS, q)$, ..., $AMC(m, BS, 7)$, the receiver at the BS must meet both SINR requirement, $\gamma(m, BS, q)$ as well as BS minimum receiver sensitivity requirement, $RSS(m, BS, q)$, $q=1,2,...,7$, respectively, as shown in Table 4.4. When the BS receives the $Polling\_response(m)$ message from SS_m ,

APC at the BS extracts $P(m, BS)$ from $AMC\_Controller\_Info\{P(m, BS), \gamma(BS, m), RSS(BS, m), AMC(BS, m, q), Req\_P(BS, m, q)\}$ in the $Polling\_response(m)$ sent by SS_m . APC further retrieves received SINR, $\gamma(m, BS)$ and $RSS(m, BS)$ from its CCE at the PHY layer.

Equation (4.18) states the SINR requirement when $AMC(m, BS, q)$ is chosen. Similarly, equation (4.22) states the RSS requirement when $AMC(m, BS, q)$ is chosen. In order to fulfill both SINR and RSS requirements, the bigger value of $P(m, BS)$ from (4.18) and (4.22) is chosen for requested transmission power, $Req\_P(m, BS, q)$. Based on this policy, $Req\_P(m, BS, q)$ is calculated as:

$$Req\_P(m, BS, q) = \text{Max}[P(m, BS) \times \frac{\gamma(m, BS, q)}{\gamma(m, BS)}, P(m, BS) \times \frac{RSS(m, BS, q)}{RSS(m, BS)}] \quad (4.23)$$

Now, the maximum achievable q level of the modulation and coding scheme is going to be derived when the transmission power of SS_m , $P(m, BS)$ is given. Here, an approximation approach methodology is used to obtain the maximum achievable value of q .

Let the highest $AMC(m, BS, q)$, $q=7$ be selected. Substituting $q=7$ into (4.23), $Req\_P(m, BS, 7)$ is calculated. If $Req\_P(m, BS, 7) > P(m, BS)$, it means that the transmission power of SS_m , $P(m, BS)$ cannot support modulation and coding scheme $AMC(m, BS, 7)$.

Let $q = q-1$, re-calculate $Req\_P(m, BS, 6)$. This iteration continues until $Req\_P(m, BS, q) \leq P(m, BS)$.

The result is embedded in $AMC\_Controller\_Info\{P(BS, m), \gamma(m, BS), RSS(m, BS), AMC(m, BS, q), Req\_P(m, BS, q), Max\_R(t)\}$ and is broadcasted to all SSs.

The maximum achievable modulation and coding scheme $AMC(q)$, hence the

maximum achievable data transmission rate $R(q)$, is used for data transmission, with the need of a lower transmission power, the efficiency of transmission power is achieved.

4.4.2 The Second Function of APC: APC Algorithm for rtPS Traffic

The scheduler at SS_m extracts $AMC(m, BS, q)$ and $Req\_P(m, BS, q)$ from the $AMC\_Controller\_Info\{P(BS, m), \gamma(m, BS), RSS(m, BS), AMC(m, BS, q), Req\_P(m, BS, q)\}$ when SS_m receives the DL/UL_Map. The transmitter at SS_m uses those parameters to transmit a burst in a nrtPS connection without using the second function of APC scheme. When it transmits a burst in a rtPS connection, it performs its second function.

The objective of the second function of the APC scheme is to enhance QoS to rtPS service class by increasing the transmission power of SS_m . Once $P(m, BS)$ increases, it results in the increase of the received SINR at the BS, $\gamma(m, BS)$. The level of modulation and coding scheme is stepped up by the AMC controller both in the BS and SS_m based on $\gamma(m, BS)$. As a result, the data transmission rate increases to transmit the burst from rtPS service class.

The detailed procedure is as follows.

When APC at SS_m senses that q in $AMC(m, BS, q)$ is equal to 7, it is aware that maximum data transmission rate is achieved by transmission power $Req\_P(m, BS, q)$. It maintains the transmission power.

The APC increases the transmission power of SS_m to its maximum transmission power based on the following conditions:

- When APC senses that q is less than 7, it is aware that the data transmission rate for rtPS service class is not at the highest rate.
- and if the normalized time-delay satisfaction index of the k^{th} packet from rtPS connection i at time t , $NTDSI(m, i, k, t)$, is greater than 0.8, which indicates that the k^{th} packets is going to be dropped if it is not transmitted immediately, the APC is aware of the urgency of the k^{th} packet.

SS_m uses its maximum transmission power to provide the feasible highest data transmission rate for the burst to avoid the k^{th} packet to be dropped from the queue.

AMC controller steps $AMC(m, BS, q)$ one level up and the requested transmission power of SS_m is re-calculated when APC senses that q is less than 7 but the normalized time-delay satisfaction index of the k^{th} packet at time t , $NTDSI(m, i, k, t)$ is less than 0.8.

Under maximum transmission power constraint of SS_m , a feasible higher transmission power that is able to fulfill both SINR requirement and RSS requirement for higher AMC index is set to transmit the k^{th} packet from rtPS service class.

The operation of the APC algorithm to adjust transmission power of SS_m , $P(m, BS)$, for transmitting packets from rtPS service class is described as follows:

Begin

Gets $\gamma(m, BS)$, $AMC(m, BS, q)$ and $Req\_P(m, BS, q)$ from

$AMC\_Controller\_Info\{P(BS, m), \gamma(m, BS), RSS(m, BS), AMC(m, BS, q), Req\_P(m, BS, q), Max\_R(t)\}$;

Gets timeslot allocation information for the rtPS from HOS scheduler in DL/UL_Map;

If ($q=7$ is true or $Req\_P(m, BS, q) = \text{maximum transmission power of } SS_m$)

$$P(m, BS) = Req\_P(m, BS, q);$$

Else if (timeslots allocated for rtPS burst is true && $q < 7$)

If ($NTDSI(m, i, k, t) > 0.8$)

** The k^{th} packet is going to be dropped.

$$P(m, BS) = \text{maximum transmission power of } SS_m;$$

Else

$$p = q + 1; \quad \text{** Step } AMC(q) \text{ one level up.}$$

calculates

$$P(m, BS) = \text{Max}[Req\_P(m, BS, q) \times \frac{\gamma(m, BS, p)}{\gamma(m, BS, q)}, Req\_P(m, BS, q) \times \frac{RSS(m, BS, p)}{RSS(m, BS, q)}]$$

If ($P(m, BS) \geq \text{maximum transmission power of } SS_m$)

**This is the maximum transmission power constraint of SS_m .

$$P(m, BS) = Req\_P(m, BS, q);$$

$$AMC(m, BS, p) = AMC(m, BS, q);$$

Else

$$P(m, BS) = \text{Max}[Req\_P(m, BS, q) \times \frac{\gamma(m, BS, p)}{\gamma(m, BS, q)}, Req\_P(m, BS, q) \times \frac{RSS(m, BS, p)}{RSS(m, BS, q)}]$$

$$AMC(m, BS, p) = AMC(m, BS, p);$$

End

The transmitter of SS_m uses updated $P(m, BS)$ as well as $AMC(m, BS, p)$ and corresponding data transmission rate to transmit traffic burst from rtPS service class. It results in the enhancement of QoS provisioning to the traffic from rtPS service class.

4.5 Performance Evaluation

A series of simulation experiments have been carried to evaluate the performance of the proposed cross-layer MAC protocol and the corresponding scheduling algorithm as well as APC scheme. The experiments have been conducted in two stages. In the first stage, the simulation for the HOS algorithm without APC has been carried out. In the second stage, the simulation of the HOS algorithm with APC has been carried out.

Following the signaling mechanism specified by the IEEE 802.16d standard, the WiMAX simulation model has been developed by using ExtendSim [71] software. The scheduling algorithms have been developed by using C language.

The performances of the proposed scheduling schemes for UGS, rtPS and nrtPS connections have been compared with that of MCS [47], Exponential Rule Scheduler (EXPRule) in [53] and Priority Function Scheduler (PRFS) in [56]. The simulation results show that the proposed scheduling scheme can not only effectively support rtPS traffic to make real-time packets meeting their delay bounds but also provide fairness to nrtPS traffic and serve it with reasonable performance.

4.5.1 Simulation Design

In the simulation experiments, it is assumed that the WiMAX system is operating as the WirelessMAN-SC in TDD mode. For wireless fading channel in the simulation model, the channel quality can be captured by a parameter, γ . Since the transmission channel varies from frame to frame, the general Nakagami- m model is used to describe γ statistically [84]. The received SINR,

γ per frame is thus a random variable with a Gamma probability density function:

$$p_{\gamma}(\gamma) = \frac{m^m \bar{\gamma}^{m-1}}{\bar{\gamma}^m \Gamma(m)} \exp\left(-\frac{m\gamma}{\bar{\gamma}}\right) \quad (4.24)$$

where $\bar{\gamma}$ is the average received SINR, $\Gamma(m) := \int_0^{\infty} t^{m-1} e^{-t} dt$ is the Gamma function and m is the Nakagami fading parameter ($m \geq 1/2$). The channel is assumed to be quasi-static so that the channel gains are constant over a frame. However, they are allowed to vary from frame to frame. The average received SINR, $\bar{\gamma}$ can be expressed as:

$$\bar{\gamma} = P_t - L_i - P_N - L_p \quad (4.25)$$

where P_t (dBm) is the transmitter output power, L_i (dBm) is the implementation loss due to hardware connecting cables, antenna patterns, etc., P_N (dBm) is the receiver noise power, which is related to the hardware noise figure and bandwidth, and L_p (dBm) is the path loss due to radio propagation. The receiver sensitivity RSS for all receivers has been set at -65 dBm.

The network has been designed with PMP topology with one BS and eight SSs. The positions of the SSs are assumed to be independent and distributed randomly. The traffic parameters are selected from the supported range of values that fully cover the required values for multimedia services defined in Recommendation ITU-R M.1225 [85]. In the simulation experiments, the effect of the uplink scheduling algorithm and the APC scheme is considered. The integrated traffic consists of fixed load of UGS, different load of rtPS and nrtPS traffic from eight SSs to the BS in the network. The traffic load of UGS is fixed at 1.37 Mbps. The

traffic load for rtPS and nrtPS is 50% of the remaining entire traffic. The inter-arrival time of UGS packets is constant. It is set to 10.1 ms. Packet length is fixed to 200 bytes. The deadline is set to 50 ms. Both rtPS and nrtPS connections are burst traffic flows. Packet arrivals follow Poisson distribution with packet inter-arrival time exponentially distributed. Packet length is exponentially distributed. The mean packet length is set to 500 bytes. The mean deadline of rtPS is 80 ms and it is exponentially distributed. There is no deadline for nrtPS connections. The frame size is set to 10 ms, 500 time slots per frame. The duration of each time slot is 20 μ s. Channel bandwidth is assumed to be 25 MHz. The queue size is set big enough so that packets would not be dropped from a queue because the queue is full. So packet loss is due to the packet being dropped if its waiting time exceeds its maximum latency, or packet errors due to the wireless transmission channel in the simulation.

4.5.2 Experimental Results

As shown in Figure 4.3, 4.4, 4.5, 4.6, 4.7 and 4.8, the performance of five scheduling schemes are evaluated. The performance metrics are packet loss rate, average packet delay, UL throughput of the rtPS connections, UL throughput of the nrtPS traffic connections, system UL channel throughput and average packet delay of the UGS connections when the total traffic load, which consists of UGS, rtPS and nrtPS service classes, changes accordingly.

Figure 4.3 shows the relationship between the packet loss rates of rtPS traffic and the various traffic loads. It is clear that the packet loss rates of rtPS traffic in the HOS are lower than that of MCS, PRFS and EXPRule schemes when the traffic intensity increases. The reason is that the MCS has not considered any

delay satisfaction factor for rtPS traffic. The PRFS scheme has only considered the delay satisfaction indicator for rtPS traffic. The EXPRule scheme has only considered starvation index for rtPS traffic. However, the proposed HOS algorithm has considered both normalized time-delay satisfaction index $NTDSI(m, i, k, t)$ and normalized predictive starvation index $NPSI(m, i, t)$ to determine the priority of the transmission.

Figure 4.3 Loss rate vs. traffic load for rtPS traffic

It is also observed that the proposed HOS and HOS with APC have similar performance when the traffic load is light. The HOS with APC scheme has a better performance when the traffic load is heavy. The proposed APC scheme can perform two functions, one is to optimize requested transmission power, and the other one is power control function.

The proposed APC scheme adjusts SS transmission power for rtPS traffic to achieve a higher data transmission rate for rtPS traffic classes when

- the normalized time-delay satisfaction index $NTDSI(m, i, k, t)$ of the packet is close to 1 or;
- q in $AMC(m, BS, q)$ is less than 7, meaning that the transmission data rate of SS_m is not at the maximum rate.

This results in the Symbol Rate Controller at the BS using a higher q level of modulation and coding scheme $AMC(m, BS, q)$. SS_m thus can transmit packets belonging to rtPS service class at a higher data transmission rate to prevent the real-time packet violating its deadline. Thus, the proposed HOS with APC solution can achieve a much lower packet loss rate for the rtPS traffic connections.

Figure 4.4 shows the relationship between the average packet delays of the rtPS connections and the traffic loads for five scheduling schemes.

Figure 4.4 rtPS delay vs. traffic load

Based on Figure 4.3 and 4.4, it is observed that MCS has poor performance in terms of rtPS loss rate and delay. The MCS [47] scheduling scheme is the prototype of the throughput efficient opportunistic scheduling. This policy is to allocate resources to SSs with the best channel condition. MCS is a channel-awareness scheduling scheme. However, this channel-aware policy monopolizes all resources to good-channel users that are usually located close to the Base Station. The SSs will be allocated fewer transmission opportunities if they are far from the BS.

It is also observed that the performance of EXPRule scheduler degrades dramatically when the traffic load is heavy.

However, the HOS with APC scheme outperforms among the five schemes. The reason is that the delay requirement of a packet from a rtPS connection has been included to determine its Scheduling Priority by $\exp(NTDSI(m, i, k, t)) \in [0.368, 1]$. If $ML(m, i, k) - [t - VC(m, i, k)] \leq 0$, the highest value of $NTDSI(m, i, k, t)$ is set, which leads to a higher Scheduling Priority. The expression $\exp(NTDSI(m, i, k, t))$ makes the Scheduling Priority assignment much more sensitive to the delay bound of a packet from the rtPS traffic.

It is observed that the HOS with APC scheme has a superior performance in terms of rtPS packet delay. The proposed APC scheme can adjust the SS transmission power to a higher level and AMC can select the highest feasible AMC index $AMC(q)$ when SS transmits rtPS packets. The packet delay and the QoS guarantee to rtPS traffic can be capped at a certain level.

Figure 4.5 depicts the relationship between UL throughputs of the rtPS connections and the traffic loads for the five scheduling schemes. It can be observed that the HOS and the HOS with APC scheme can achieve a much

higher UL channel throughput. The reasons are:

- The DPI assignment mechanism of the HOS algorithm for rtPS traffic burst assigns higher DPI when channel condition is good. This dynamic priority assignment mechanism provides more transmission opportunity for rtPS traffic burst.
- The APC increases the transmission power for rtPS traffic burst and thus a higher AMC index can be used, leading to a higher data transmission rate.

Figure 4.5 rtPS UL throughput vs. traffic load

Figure 4.6 shows the UL throughputs of nrtPS connections with changes in traffic loads. It is clear that MCS [47] has the best performance for nrtPS traffic class, as MCS is the prototype of throughput efficient opportunistic scheduling

for nrtPS traffic class that does not have ML requirement. It is observed that the proposed HOS and HOS with APC scheme have similar performance as MCS. It means that the proposed solution can achieve the maximum system throughput as $CSI(m, i, t)$ plays a role as a channel-awareness scheduling parameter in the proposed HOS and HOS with APC scheme.

Figure 4.6 nrtPS UL throughput vs. traffic load

The proposed solution can achieve a moderately higher UL throughput than PRFS and EXPRule. The reason is that predictive starvation index $NPSI(m, i, t)$ is considered in the opportunistic scheduling design. If $DPI(m, i, t)$, $NTDSI(m, i, k, t)$ and $CSI(m, i, t)$ are the same for all the connections, $NPSI(m, i, t)$ becomes the major factor to determine the Scheduling Priority. When the queue length of one of the nrtPS connections increases, it leads to a high value of $NPSI(m, i, t)$ so that the starvation problem of nrtPS connections can be mitigated. A certain level

of fairness in the transmission service is achieved. The queue length of each connection will be restricted to a certain level to make the system stable. This results in that the UL throughput of nrtPS connections can be improved while the QoS provisioning can be ensured to the real-time traffic from the rtPS connections.

Figure 4.7 shows the system UL channel throughputs with changes of traffic loads. Although MCS is the prototype of throughput efficient opportunistic scheduling, the policy merely is to allocate resources to SSs with the best channel condition. It does not consider delay satisfaction factor for rtPS traffic. This causes a higher packet loss rate of rtPS as shown in Figure 4.3. Thus, MCS cannot achieve the best overall system UL channel throughput as the integrated traffic load consists of UGS, rtPS and nrtPS.

Figure 4.7 System UL throughput vs. traffic load

It is observed that the proposed HOS and HOS with APC proposal can achieve a higher system UL throughput. The proposed scheduling algorithm takes the information of instantaneous wireless channel condition and the real-time traffic condition as well as traffic QoS specification to set the Scheduling Priority (SP) for each individual packet from rtPS or nrtPS service classes for the UL transmission.

Furthermore, the connection-based scheduler could waste bandwidth, if the BW_request of a scheduled connection is less than the available bandwidth in a frame. Since the HOS scheduler is a packet-based scheduler, all available bandwidth is used up. The proposed solution can achieve higher bandwidth utilization.

Figure 4.8 shows the relationship between average packet delays of the UGS connections and the traffic loads for the five scheduling schemes. Since the UGS connections have been allocated with a fixed bandwidth (fixed time duration) based on their fixed bandwidth requirement in the framework of the IEEE 802.16d standard, it is observed that average delays of the UGS are similar to each other in the five scheduling schemes.

Figure 4.8 UGS delay vs. traffic load

4.6 Summary

In this chapter, firstly, a cross-layer MAC protocol and QoS support framework has been proposed to enhance QoS provisioning in the WiMAX WirelessMAN-SC PMP system. The proposed cross-layer MAC protocol and QoS support framework takes the impact of the air interface on the MAC layer protocol into account to enhance the system efficiency.

Secondly, the HOS algorithm associated with the proposed MAC protocol and QoS support framework has been proposed. The proposed HOS has features of channel-awareness, queue-awareness and traffic QoS-awareness. It determines the dynamic Scheduling Priority of each packet by its four key scheduling

parameters, namely the dynamical priority index, the channel specification index, the normalized time-delay satisfaction index and the normalized predictive starvation index. The proposed HOS is a type of opportunistic scheduling algorithm in view of communication over spatiotemporally varying wireless link whereby the multi-user diversity is exploited to maximize bandwidth efficiency and system throughput. With the MAC-PHY cross-layer approach, the proposed HOS algorithm with two-stage structure can enhance the QoS to rtPS service class with less scheduling complexity and scheduling overhead.

Thirdly, the Adaptive Power Control scheme has been proposed associated with the HOS scheduling algorithm. The APC scheme can not only perform transmission power optimization to enhance the system power efficiency, but also control the transmission power adaption for rtPS traffic with QoS-oriented consideration.

The following features can be observed from the design of cross-layer MAC protocol and QoS support framework with Holistic Opportunistic Scheduling algorithm and APC scheme:

1. Each SS monitors and computes the scheduling priority of packets from its own connections. A frame time is sufficient for every SS to compute four key scheduling parameters for a packet that could be transmitted in the next frame. The first stage scheduler at SS only sorts the Scheduling Priority of packets in its own queues, the complexity of the algorithm at SS is $O(N\log N)$ where N is the number of packets in the queues. Similarly, the second stage scheduler at BS only sorts the SS's $Max\_SP(m)$, where $m=1,2,...,M$. The scope of the sorting list is limited to the total number of SSs, the complexity

- of the algorithm at BS is $O(N \log N)$ where N is the number of SSs. This means that the overhead of scheduling is divided by two stages and it is distributed and shared by all SSs. The efficiency of the scheme is achieved.
2. When a channel condition is good and both SINR at the BS and SS_m are high, $DPI(m, i, t)$ will be set a high value for rtPS service class. Thus, maximum system throughput can be achieved. When $R(m, i, t) = 0$, the channel is in deep fading and the capacity is zero, so connection i at SS_m should not be served regardless of its delay performance. Efficient bandwidth utilization can be achieved.

The proposed HOS scheduling algorithm and APC scheme can effectively and efficiently schedule and manage the transmission service to the integrated traffic consisting of UGS, rtPS and nrtPS connections. The simulation results show that the proposed solution can improve QoS significantly for rtPS traffic connections while achieving high system UL throughput and channel utilization.

In the next chapter, two new CAC solutions will be presented to enhance QoS support in WiMAX system with a cross-layer optimization design.

Chapter 5 Proposed CAC Solutions to Support QoS

As presented in Chapter 3 and 4, the proposed QoS provisioning architecture and frameworks together with packet scheduling schemes as well as an APC scheme are able to provide QoS provisioning in WiMAX system.

Both CAC and scheduling algorithm have an important role since they can manage and guarantee the QoS requirements of connections. A single scheduling algorithm cannot guarantee all the QoS requirements of traffic without the support of a suitable CAC and vice versa. The CAC is essential to guarantee the users' QoS requirements as it regulates the number of connections or users in the system so as to control the congestion, connection-level QoS satisfaction and packet-level QoS satisfaction while achieving the system efficiency through optimizing the network resources.

Therefore, it appears to be essential to develop a novel CAC scheme in tandem with the packet scheduling scheme to provide QoS support in a flexible and efficient way in the WiMAX system.

The importance of the combination of scheduling policy with an efficient CAC scheme to guarantee QoS in WiMAX systems has been recognized [14, 21, 42].

Although those CAC policies have been proposed to combine with an uplink scheduling algorithm to provide QoS in WiMAX, very few of them were catered for WiMAX systems with a tight combination of an opportunistic scheduling algorithm.

In the first part of this chapter, a novel two-stage cross-layer QoS support framework and a cross-layer CAC scheme in tandem with the Holistic Opportunistic Scheduling (HOS) are presented in a single carrier (SC) WiMAX PMP network. The proposed CAC scheme is termed as Elastic CAC (ECAC). The proposed Holistic Opportunistic Scheduling (HOS) has been presented in Chapter 3.

Furthermore, in the second part of this chapter, a novel cross-layer Cognitive Radio-based QoS support framework and Cognitive Radio-based CAC (CRCAC) scheme are presented in WiMAX WirelessHUMAN^{TM}-OFDM PMP systems. The proposed CRCAC is able to expand the system capacity while providing QoS to real-time traffic. By using a cross-layer approach, the proposed solution can intelligently explore unused spectrums and spread to non-active spectrums to improve the capacity of the system significantly and provide guaranteed QoS to real-time traffic.

5.1 Cross-Layer Elastic CAC

As discussed in Chapter 2.5, both CAC and the scheduling algorithm have an important role since they can manage and guarantee the QoS requirements of the connections. A single scheduling algorithm cannot guarantee all the QoS requirements of traffic without the support of a suitable CAC and vice versa.

CAC schemes tailored for WiMAX networks have been proposed in [12-25]. The combination of a scheduling policy with an efficient CAC scheme to guarantee QoS in WiMAX networks has become an important issue. Although some CAC policies have been proposed to combine with an uplink scheduling algorithm [14, 21, 25] to provide QoS support in WiMAX systems, very few of them were catered for WiMAX systems with a tight combination of an opportunistic scheduling algorithm.

In this section, a novel two-stage cross-layer QoS support framework and a cross-layer CAC scheme termed as Elastic CAC (ECAC) are proposed in tandem with the Holistic Opportunistic Scheduling (HOS) in WiMAX WirelessMAN-SC PMP systems. The proposed solution has unique features. The proposed cross-layer CAC has been tightly combined with the HOS algorithm. It can control the number of connections to be admitted into the system based on the predefined QoS requirements of different connections to achieve cost-effective revenue in the WiMAX system. When the system resource becomes scarce, the ECAC can automatically and dynamically adjust or compress the bandwidth consumed by the existing connections, so that the QoS requirements of on-going connections can be guaranteed, and as many rtPS connections as possible will be admitted.

When the AMC scheme adjusts its transmission mode frame by frame, the transmission rate is an Adaptive Discrete Rate (ADR) [86] with different coding scheme. The system modeling of the proposed CAC is presented. Numerical analysis of the system effective bandwidth and the average packet delay of each service class are carried out.

5.1.1 QoS Framework with CAC in tandem with Scheduling Scheme

As shown in Figure 2.2, IEEE 802.16d left the details of the CAC algorithm undefined. This section focuses on the cross-layer CAC scheme that is in tandem with the Holistic Opportunistic Scheduling (HOS) in WiMAX WirelessMAN-SC PMP systems.

The performance of MAC protocol, the functionality of scheduler and CAC algorithm are influenced by time-varying wireless channel in WiMAX systems. With the cross-layer approach, a new interface of AMC_Control_Info needs to be designed to link up the AMC scheme and CAC from the PHY to the MAC layer.

With the cross-layer approach, the proposed QoS provisioning architecture aims at providing CAC module, scheduling module and AMC module to complete the missing parts in the QoS architecture specified in the IEEE 802.16d standard as shown in Figure 2.2. An efficient two-stage cross-layer QoS support framework with the HOS scheduling algorithm and a cross-layer CAC scheme are proposed, as shown in Figure 5.1.

Figure 5.1 Cross-layer QoS support framework with CAC in tandem with the HOS scheduling scheme

The detailed design objective and rationale for the proposed framework is described as follows.

1. Considering the impact of air interface on the MAC layer protocol, a wireless Channel Condition Estimator (CCE) is designed at the PHY layer at the BS as well as at the SSs. It not only monitors instantaneous channel condition status like RSS , $Current\_SINR$, γ , but also predicts long-term channel condition by using the statistic parameters including Max_SINR , $Average\_SINR$ and Min_SINR over the execution period.
2. Based on the information from CCE, the BS or SS PHY Symbol Rate Controller, which consists of FEC, Symbol Mapper and AMC Controller, computes a feasible modulation and coding scheme as well as a data transmission rate for the transmission service according to Table 4.4.

3. The output of the PHY Symbol Rate Controller is called $AMC\_Control\_Info_{SS}$ at a SS, or $AMC\_Control\_Info_{BS}$ at the BS. $AMC\_Control\_Info_{SS}$ consists of two elements namely, $Current\_SINR_{BSm}$, which is the SINR detected by the m^{th} SS when receiving UL/DL_Map from the BS, and AMC Index $AMC(q)$ for DL transmission based on its received $Current\_SINR_{BSm}$. $AMC\_Control\_Info_{BS}$ consists of three elements namely,
  - $Current\_SINR_{mBS}$, which is the SINR detected by the BS when receiving Polling_response from the m^{th} SS, where $m=1, 2, \dots, M$;
  - AMC Index $AMC(q)$ for each SS based on its received $Current\_SINR_{mBS}$ at the BS and
  - $Max\_R(t)$, which is the system maximum data transmission rate at time t determined by the BS when the maximum modulation and coding level q in $AMC(q)$ is used in the system.

They are forwarded to the MAC layer via PHY Service Access Point (SAP) for CAC and packet scheduling.

4. The Bandwidth Allocation Estimation Module (BAEM) in the CAC at the BS takes in $AMC\_Control\_Info_{SS}$ information from Polling_response message and bandwidth allocation information from DL/UL_Map. After the system executes for the duration of a few DL/UL frames, the CAC conducts a statistical analysis of bandwidth allocation information from DL/UL_Maps, enabling it to compute the bandwidth consumed by every SS. Then the CAC estimates the effective UL bandwidth of the system $\overline{BW}_{uplink}$ based on (5.4) described later in Section 5.1.4.1 and computes the average packet delay of UGS and rtPS service class $\overline{D}_{ugs}$, $\overline{D}_{rtPS}$ by using (5.7) described later in

- Section 5.1.4.2. The CAC determines whether a new request of a connection can be granted or not based on the proposed cross-layer CAC algorithm, which is described in detail in Section 5.1.2. The BS maintains the ACL and updates it periodically.
5. In tandem with the CAC, the scheduler serves the connections in an ACL. The first stage of scheduling operates at each SS. It uses the Traffic Specification (TSpec) of each connection, AMC_Control_Info_{SS} and queue status to compute four priority indices and determine the Scheduling Priority of packets in each connection by the HOS algorithm. The Scheduling Priority will be sorted in descending order. Then the SS takes the maximum value of the calculated priority as the Scheduling Priority (SP) of the SS. The SS loads its SP, its total BW_request and its AMC_Control_Info_{SS} in Polling_response and sends the Polling_response to the BS when it is polled.
  6. The second stage of scheduling at the BS sorts SP values of all SSs in descending order and selects the SS with the highest SP to assign UL time slots to it according to its BW_request in the UL sub-frame exclusive of the time slots for UGS. If the number of time slots allocated to the SS is less than the total number of available time slots in the UL sub-frame, the remaining available time slots will be assigned to a SS with the second highest SP according to its BW_request. This procedure iterates until all available time slots in the UL sub-frame is used out. The result of the time slots assignment and AMC_Control_Info_{BS} will be embedded in the DL/UL_Map to be broadcasted from the BS to all SSs.
  7. The scheduler at each SS extracts the information of its allocated transmission time slots and AMC_Control_Info_{BS} from the UL_Map. The SS

transmits its packets, in order of the highest Scheduling Priority to the lowest Scheduling Priority, to the BS in the allocated time slots at the data rate specified in $AMC\_Control\_Info_{BS}$.

5.1.2 Cross-Layer Elastic CAC Algorithm

From the network's point of view, the network must support the connections of all service classes. Before the total network resource is consumed, all connections should be admitted with equal priority as soon as they arrive based on a First-Come-First-Served basis if the network can support both the bandwidth requirement and the delay constraint of both on-going connections and new arrival connections.

A WiMAX network is a sophisticated wireless broadband access network that has UGS, rtPS, nrtPS and BE service classes. Each service class has a different QoS requirement. A CAC policy must be able to differentiate service classes based on their different QoS criteria. In order to avoid the starvation problem of lower priority service classes like nrtPS and BE service class, a threshold is set for each service class based on Optimal Revenue algorithm [17] before the execution of the network. The Optimal Revenue algorithm is used as it can provide connection level QoS to real-time and non-real-time traffic.

Wireless channel is time-varying channel where the channel capacity or bandwidth is no long fixed. It is dynamically changed based on the SINR of the wireless channel. Hence, the effect of AMC on the proposed CAC scheme must be considered.

Based on the above considerations, a cross-layer CAC algorithm that is tightly combined with the HOS opportunistic scheduling scheme is proposed. It is

termed as Elastic Connection Admission Control (ECAC) algorithm. The ECAC consists of three unique modules, namely

- Bandwidth Allocation Estimation Module (BAEM);
- Bandwidth Elastic Module (BEM) and
- QoS Control Module (QoS_CM).

The illustrative flowchart is shown in Figure 5.2.

Figure 5.2 ECAC flowchart

When a traffic flow sends its request to join one sector of the network, the ECAC extracts its MSTR and ML, which are two QoS parameters from its AdmittedQoSParameterSet [1], and makes a decision on whether to reject or to

admit the flow into the network according to the following procedure:

Step 1. In order to avoid starvation of lower priority classes, Bandwidth Allocation Estimation Module initially sets a threshold for each service class based on Optimal Revenue algorithm [17] in an off-line manner. They are four parameters: Th_{UGS} , Th_{rtPS} , Th_{nrtPS} , Th_{BE} and $Th_{UGS} + Th_{rtPS} + Th_{nrtPS} + Th_{BE} \leq BW_{uplink}$, where BW_{uplink} is the total bandwidth of uplink. The potential revenue of a WiMAX network system is determined by the number of connections of UGS, rtPS, nrtPS and BE in the network and the price of service rer^{UGS} , rer^{rtPS} , rer^{nrtPS} and rer^{BE} applied to each UGS, rtPS, nrtPS and BE connection, respectively.

An optimization problem is formulated to maximize the network revenue while the connection-level QoS (i.e., connection blocking probability) is maintained at a target level. The average number of on-going connections $Avg\_Connection\_UGS$, $Avg\_Connection\_rtPS$, $Avg\_Connection\_nrtPS$ and $Avg\_Connection\_BE$ that can be admitted in the system given the connection blocking probability BP_{UGS} , BP_{rtPS} , BP_{nrtPS} and BP_{BE} which are defined as the functions of the corresponding threshold Th_{UGS} , Th_{rtPS} , Th_{nrtPS} and Th_{BE} .

The optimization formulation is expressed as follows:

Maximize potential revenue:

$$\begin{aligned} \text{Max } [& rer^{UGS} \times (Avg\_Connection\_UGS | Th_{UGS}) + rer^{rtPS} \times \\ & (Avg\_Connection\_rtPS | Th_{rtPS}) + rer^{nrtPS} \times \\ & (Avg\_Connection\_nrtPS | Th_{nrtPS}) + rer^{BE} \times \\ & (Avg\_Connection\_BE | Th_{BE})] \end{aligned} \quad (5.1)$$

Subject to: $Th_{UGS} + Th_{rtPS} + Th_{nrtPS} + Th_{BE} \leq BW_{uplink}$

$$BP_{UGS} / Th_{UGS} \leq Target\_BP_{UGS}$$

$$BP_{rtPS} / Th_{rtPS} \leq Target\_BP_{rtPS}$$

$$BP_{nrtPS} / Th_{nrtPS} \leq Target\_BP_{nrtPS}$$

$$BP_{BE} / Th_{BE} \leq Target\_BP_{BE}$$

where $Target\_BP_{UGS}$, $Target\_BP_{rtPS}$, $Target\_BP_{nrtPS}$ and $Target\_BP_{BE}$ are the target connection blocking probabilities for the UGS, rtPS, nrtPS and BE connections, respectively.

The ECAC at the BS takes AMC information from $AMC\_Control\_Info_{ss}$ and DL/UL_Map from the scheduler. It computes the effective UL channel bandwidth $\overline{BW}_{uplink}$ based on (5.4) described later in Section 5.1.4.1, and the bandwidth occupied by each class BW_{UGS} , BW_{rtPS} , BW_{nrtPS} and BW_{BE} as well as the average packet delay of UGS and rtPS service class $\overline{D}_{ugs}$, $\overline{D}_{rtPS}$ by (5.7).

Step 2. When a request of a new flow has been sent from the m^{th} SS to the BS, the ECAC calculates the remaining uplink bandwidth:

$$UL\_BW_{remain} = BW_{uplink} - BW_{UGS} - BW_{rtPS} - BW_{nrtPS} - BW_{BE} \quad (5.2)$$

Then UL_BW_{remain} is converted to an effective remaining uplink bandwidth based on $AMC\_Control\_Info_{ss}$ of the m^{th} SS.

The effective remaining uplink bandwidth is calculated as follows:

$$Effective\_UL\_BW_{remain} = UL\_BW_{remain} \times E(q) \quad (5.3)$$

where $E(q)$ is the efficiency of the modulation and coding scheme used in the m^{th} SS based on its channel SINR as shown in Table 4.4.

Step 3. The QoS Control Module of ECAC performs its first function, $QoS\_CM(1)$, to examine the bandwidth requirement of the new flow.

$Effective\_UL\_BW_{remain}$ is compared with the MSTR requirement of the new flow. If $Effective\_UL\_BW_{remain} < MSTR$, go to Step 5 to check on the possibility of borrowing some bandwidth from a lower service class. If $Effective\_UL\_BW_{remain} > MSTR$, go to Step 4 to check up the maximum latency requirement of both on-going connections and the new flow.

Step 4. In fact, just a bandwidth checkup is not sufficient to guarantee the whole ActiveQoSParameterSet as wireless communication channel is a time-varying channel. After the request of a new flow has successfully passed the bandwidth requirement checkup, the QoS Control Module performs its second function, QoS_CM(2), to examine the maximum latency requirements of existing connections and the new flow.

Assume that the new flow could be admitted in the network, it would join one of the UGS, rtPS, nrtPS and BE service classes that it belongs to. $\overline{D}_{ugs}$, $\overline{D}_{rtPS}$ will be updated as the traffic pattern and traffic load have changed. The second function of QoS Control Module, QoS_CM(2), will check the ML constraint of each existing UGS and rtPS connection and the possible admitted new flow. The function of QoS_CM(2) will ensure that the ML of all admitted UGS traffic connections together with the possible admitted new connection is not smaller than the calculated $\overline{D}_{ugs}$ or $\overline{D}_{ugs} \leq Min\_ML(UGS)$, where $Min\_ML(UGS)$ is the minimum Maximum Latency requirement among all admitted UGS connections plus the possible admitted new connection. The function of QoS_CM(2) will also achieve the result, which is the ML of all admitted rtPS traffic

connections with the possible admitted new connection is not smaller than the calculated $\overline{D}_{rtPS}$ or $\overline{D}_{rtPS} \leq Min\_ML(rtPS)$, where $Min\_ML(rtPS)$ is the minimum Maximum Latency requirement among all admitted rtPS connections and the possible admitted new connection. If both conditions are true, go to Step 7, the new flow will be admitted. Otherwise, go to Step 8, the new flow will be blocked.

The corresponding average delay of UGS or rtPS obtained by the queueing analytical model is less than the delay requirement of each admitted UGS or rtPS traffic connection, respectively. The trade-off in the proposed admission control design is to reach the delay guaranteed real-time service commitments as much as possible while keeping high network utilization.

Step 5. The Bandwidth Elastic Module of ECAC sets the bandwidth-borrowing algorithm so that a higher service class can borrow bandwidth from a lower service class when the lower service class consumes the bandwidth more than its threshold. The detail is as follows:

Firstly, the ECAC determines the type of service class that the new flow belongs to. Secondly, the ECAC checks whether the bandwidth consumed by all connections belonging to this type of service class is greater than its threshold or not. If it is so, go to Step 8, the ECAC denies the request of the new flow. Otherwise if there is a lower service class using more bandwidth than its threshold, the ECAC calculates BW_{Low} which is the actual bandwidth used by this service class subtracted by its threshold. BW_{Low} represents the amount of bandwidth the ECAC can borrow from the lower service class. The ECAC can compress the bandwidth used by

the lower service classes by re-allocating MRTR to the connections. Update $UL\_BW_{remain} = BW_{Low} + UL\_BW_{remain}$, then re-calculate the effective remaining uplink bandwidth $Effective\_UL\_BW_{remain} = UL\_BW_{remain} \times E(q)$. After that, go to Step 6.

Step 6. This procedure is similar as Step 3, but it is slightly different. After $Effective\_UL\_BW_{remain}$ is updated by the Bandwidth Elastic Module of ECAC as stated in Step 5, it is compared with the MSTR requirement of the new flow. If $Effective\_UL\_BW_{remain} < MSTR$, go to Step 8, the ECAC denies the request of the new flow. If $Effective\_UL\_BW_{remain} > MSTR$, go to Step 4 to check up Maximum Latency requirements of all connections.

Step 7. The request of the new flow is granted. The new flow is admitted in the ACL, and the UL_BW_{remain} will be updated.

Step 8. The request of the new flow is rejected.

The ECAC manages the ACL so that a new flow can be accepted when its MSTR and ML requirements are met and the QoS requirements of the existing connections are not violated.

The key advantage of the proposed cross-layer ECAC mechanism is its ability to catch the PHY condition like AMC, and real-time network performance like average packet delay and effective bandwidth by using the AMC information from the PHY layer and bandwidth allocation assignment from the scheduler. Thus, the ECAC can improve the accuracy of the decision. A better accuracy and higher network utilization can be achieved while network overloading can be prevented.

The proposed ECAC is also equipped with the Bandwidth Elastic Module to

perform the bandwidth borrowing function to enhance the connection-level QoS of the higher priority service classes like UGS and rtPS.

5.1.3 The Cooperation between Opportunistic Scheduling and ECAC

In the previous sections, the HOS algorithm and the ECAC scheme are discussed separately. Nevertheless, the HOS and the ECAC work cooperatively to provide cross-layer QoS support. As shown in Figure 5.1, the proposed HOS scheduling module and ECAC module can interact with each other in order to obtain enough information of network status. On one hand, HOS needs the traffic load pattern in the WiMAX system, in short, the information of the ACL. The ECAC can manage and regulate the ACL. It can provide the traffic load information like the number of connections and their bandwidth requirements since it receives requests of the traffic flows and corresponding QoS parameters from the application layer. On the other hand, the criteria used by the ECAC are $Effective\_UL\_BW_{remain}$, $\overline{D}_{ugs}$ and $\overline{D}_{rtPS}$. When the ECAC module makes an admission decision, it has to track the bandwidth assignment scheme that is decided by the HOS algorithm. It also has to track the current data transmission rate of downlink or uplink that is decided by the AMC module. From (5.4), $\overline{BW}_{uplink} = \sum_{m=1}^M BW_m \sum_{q=1}^Q E(q)p_q$, it is obvious that $\overline{BW}_{uplink}$ and $Effective\_UL\_BW_{remain}$ are determined by the HOS. Furthermore, $\overline{D}_{ugs}$, $\overline{D}_{rtPS}$ is a function of $\overline{BW}_{uplink}$ in (5.7). As a result, the ECAC depends on the HOS algorithm to make the CAC decision accurately to provide QoS provisioning to

the heterogeneous traffic in the WiMAX system.

5.1.4 Queueing Analytical Model

5.1.4.1 Effective Uplink Channel Bandwidth Analysis for Wireless Channel with AMC

It is assumed that the bandwidth of wireless channel in the network remains constant within a frame but may vary from frame to frame. The AMC scheme adjusts its transmission mode on a frame basis. The transmission rate is an Adaptive Discrete Rate (ADR) [86] with different coding scheme.

Lemma 5.1: In the case of using the Adaptive Discrete Rate AMC with possible $AMC(q)$ based on different channel SINR regions, the effective UL channel bandwidth can be expressed as:

$$\overline{BW}_{uplink} = \sum_{m=1}^M BW_m \sum_{q=1}^Q E(q) p_q \quad (5.4)$$

where BW_m is the bandwidth allocated to the m^{th} SS by the proposed HOS algorithm or other bandwidth allocation schemes. $E(q)$ is the efficiency of modulation and coding scheme as shown in Table 4.4. p_q is the probability that SINR falls into the region $[\gamma_q, \gamma_{q+1})$ ($q=1,2,\dots,Q$) where $AMC(q)$ with $E(q)$ is selected.

Proof:

First, the probability of a given modulation and coding scheme $AMC(q)$ to be selected is calculated.

For the wireless fading channel in study, the channel quality can be captured

by a parameter, Current_SINR, γ . Since the transmission channel varies from frame to frame, the general Nakagami- m model [84] is used to describe γ statistically. The received Current_SINR, γ per frame is thus a random variable with a Gamma probability density function:

$$p_\gamma(\gamma) = \frac{m^m \bar{\gamma}^{m-1}}{\bar{\gamma}^m \Gamma(m)} \exp\left(-\frac{m\gamma}{\bar{\gamma}}\right) \quad (5.5)$$

where $\bar{\gamma}$ is the average received SINR, $\Gamma(m) := \int_0^\infty t^{m-1} e^{-t} dt$ is the Gamma function

and m is the Nakagami fading parameter ($m \geq 1/2$).

Suppose that a target BER, e.g. BER_0 is set. The region boundaries or switching thresholds are then set to the SINR requirement to achieve the BER_0 as shown in Table 4.4. There are Q possible modulation and coding schemes, numbered from $\text{AMC}(1)$ to $\text{AMC}(Q)$, where $Q=7$. The probability that SINR falls in the region $[\gamma_q, \gamma_{q+1})$ can be expressed as:

$$p_q = \int_{\gamma_q}^{\gamma_{q+1}} p_\gamma(\gamma) d\gamma \quad (5.6)$$

The average AMC efficiency is $\bar{E} = \sum_{q=1}^Q E(q) p_q$ over Nakagami- m fading channel, which is just the sum of the AMC efficiency associated with the individual q region weighted by the probability p_q .

It is assumed that M SSs in the WiMAX network. When a connection transmits packets from the m^{th} SS to the BS, it is assigned to one of the $\text{AMC}(q)$ modulation and coding schemes depending on the SINR received at the BS. The scheduler at the BS will allocate BW_m to the m^{th} SS so that the effective UL

channel bandwidth can be the sum of the bandwidth allocated to the individual SS associated with its average AMC efficiency $\overline{E}$. Then Lemma 5.1 is proved.

The effective UL channel bandwidth $\overline{BW}_{uplink}$ given by (5.4) is determined by two parameters, BW_m and $E(q)$. BW_m in turn depends on the decision of the scheduler. $E(q)$ depends on the choice of the modulation and coding scheme at the PHY layer. It defines a cross-layer interaction from PHY to MAC layer.

5.1.4.2 Analysis of Average Packet Delay for Each Service Class

The WiMAX system under study can be modeled as a multi-class priority TDMA queueing system. The thesis focuses on the UL channel analysis. The UL channel can be considered as a multiple-access communication channel shared by M SSs. The UL sub-frame consists of N consecutive time slots. Each time slot is normalized to unit length τ . Let the duration of a time frame be T_F , so that $T_F = N\tau$. The TDMA scheme for inter-SS applies in the system in which the m^{th} SS is allocated n_m time slots/frame ($m=1, 2, 3, \dots, M$) uniformly distributed in the UL sub-frame. The TDMA scheme for inter-class at a SS becomes a strict priority queue scheme where higher priority packets will be queued ahead of lower priority ones. The packets with the same priority class that arrive at different slots are served on a First-Come-First-Served basis. Each SS can transmit its packets only during its dedicated time slots that are synchronized and governed by the BS via the DL/UL_Map.

Packets arriving at each SS belong to one of the four priority classes: UGS (priority-1), rtPS (priority-2), nrtPS (priority-3) or BE (priority-4). At each SS, packet arrival is a Poisson arrival process so that λ_j is the average arrival rate of

class j packets, $j = 1, 2, 3, 4$. Each packet is transmitted in different number of time slots according to its length. The number of time slots to transmit the k^{th} arriving packet of priority class j is denoted by B_k^j (where $j = 1, 2, 3, 4$). The time taken to transmit the k^{th} packet belonging to class j is represented as σ_k^j .

Assume that the probability of the collision of requests competing for the BE transmission service is zero in the queueing model. When the UL channel becomes available, any waiting packet with priority- i will be served before anyone with priority- j , if $i < j$.

Assume that each SS is identical in general. Hence, the packet delay and queue-size behavior of any SS is statistically independent of that of any other SS. Consequently, the packet delay behavior of an arbitrary single SS, say SS_I , by a p -priority model where p is equal to 4, is investigated in this study.

The average packet delay of class j , $\overline{D_j}$, as stated as (3.23) presented in Chapter 3.3, is expressed as:

$$\overline{D_j} = W_{j,1} + \left\{ E[B_k^j] - \frac{N-1}{N} \right\} \left\lfloor \frac{N}{n_m} \right\rfloor \tau + F_{L+} P^j \quad (5.7)$$

where

$$W_{j,1} = \frac{\sum_{i=1}^J \lambda_i b_{i,2} + (1 - \eta_J) \left\lfloor \frac{N}{n_m} \right\rfloor \tau}{2(1 - \eta_j)(1 - \eta_{j-1})} \quad (5.8)$$

$\left\lfloor \frac{N}{n_m} \right\rfloor$ represents the biggest integer not larger than $\frac{N}{n_m}$.

$b_{j,2}$, the second moment of service time, is given by

$$b_{j,2} = \int_0^\infty x^2 dB_j(x) \quad (5.9)$$

$$\rho_j = \lambda_j b_{j,1}$$

$$\eta_j = \sum_{i=1}^j \rho_i \quad j=1, 2, \dots, 4$$

where F_L is the frame latency, P^j is the propagation delay.

5.2 Cross-Layer Cognitive Radio-Based Framework and CAC Scheme

There is a large amount of research results on CAC scheme to provide QoS provisioning to the traffic with Maximum Latency constraint in WiMAX systems [16-24]. However, all those proposed CAC schemes operate under the Fixed Spectrum Allocation (FSA) scheme with a pre-defined fixed spectrum with fixed channel bandwidth or fluctuating channel bandwidth if AMC scheme is used. No matter how efficient the CAC is, the number of connections or users that can be admitted in the system would be limited under the limited bandwidth constraint in the pre-defined fixed spectrum. The legacy of FSA has recently come under much investigation. Evidence by various academic studies shows that FSA is finally recognized to result in great under-utilization in a seemingly scarce wireless medium. Deregulating spectrum access, creating a shift to Dynamic Spectrum Allocation (DSA) is motivated. Facilitating this shift is the technology of Cognitive Radios (CR), which are opportunistic radios that exercise substantial intelligence to recognize and exploit unused spectrum portions [87].

In this section, a novel cross-layer CR-based QoS support framework and a

cross-layer CR-based CAC (CRCAC) is proposed in WiMAX WirelessHUMAN^{TM}-OFDM PMP systems. The proposed CRCAC has advanced features of cross-layer approach and intelligent spectrum spreading ability. It can recognize and use one of the unused spectrum portions, hence multiplying the system capacity without the fixed spectrum bandwidth constraint.

The proposed CRCAC can provide QoS provisioning to the heterogeneous traffic. It consists of three modules, namely, Bandwidth Allocation Estimation Module (BAEM), QoS Control Module (QoS_CM) and Bandwidth Spreading Module (BSM). With a cross-layer approach, it can intelligently exploit unused spectrums and spread over them. Thus, the WiMAX system can operate under more than one spectrum. The system capacity and QoS provisioning to real-time traffic can be enhanced significantly in the system below 11 GHz of unlicensed frequency band.

The proposal can guarantee the connection-level requirement, i.e., connection blocking probability and the packet-level QoS requirements, i.e., ML for different types of service classes while multiplying the system capacity. The proposed CRCAC is equipped with an intelligent Bandwidth Spreading Module (BSM). The BSM links to CR-based Intelligent Spectrum Management Module (ISMM) at the PHY layer. It exploits unused spectrum portions. When traffic load increases, the system can operate under more than one spectrum. As a result, the system throughput can be improved significantly.

The proposed CRCAC makes a decision to admit or reject the request of a new flow based on two QoS parameters, namely, Maximum Sustained Traffic Rate (MSTR) and Maximum Latency (ML), which are extracted from the AdmittedQoSParameterSet of a flow. Hence, both the connection-level and the

packet-level QoS requirements of real-time traffic can be satisfied.

Firstly, the basic concept of Intelligent Spectrum Management is briefly introduced in this sub-section.

Then, the novel cross-layer Cognitive Radio-based QoS support framework is presented.

After that, the Cognitive Radio-based CAC (CRCAC) scheme is presented in WiMAX WirelessHUMAN^{TM}-OFDM PMP systems. The proposed CRCAC aims at expanding the system capacity while providing QoS support to real-time traffic. Analytical modeling is carried out to analyze the effective channel transmission rate and the average packet delay of UGS and rtPS service class.

5.2.1 Intelligent Spectrum Management

Recently, the traditional approaches for spectrum management have been reconsidered, taking into account the actual use of spectrum. FCC's (Federal Communications Commission) Spectrum Policy Task Force has reported plentiful temporal and geographic variations in the usage of allocated spectrum [88]. One way of increasing spectrum utilization is to reuse the spectrum when their primary users are non-active. With this concept called Opportunistic Spectrum Sharing, secondary users are allowed to access to the radio frequency (RF) bands without agreement from the primary users. The unlicensed bands play a key role in this wireless system since the deployment of applications in these bands is unencumbered by regulatory delays, which resulted in a rich increase of new applications.

CR is a promising technology for enabling unlicensed devices to use White Spaces, which are referred to as the parts of the spectrum not in active use by the

primary users, efficiently [89]. It is an intelligent wireless communication system that is aware of its surrounding environment, uses the method of understanding-by-building to learn from the environment, and adapts its internal states to statistical variations in the incoming RF stimulation by making corresponding changes of its parameters in time [90]. The radio dynamically identifies portions of the spectrum that are not in use by primary users, and configures the radio to operate in the appropriate White Space. Figure 5.3 shows an example of intelligent spectrum access WSs in temporal-spectrum domain [91].

Figure 5.3 Intelligent Spectrum Access White Spaces in Temporal-Spectrum domain

The CR-based wireless system is able to utilize the large amount of unused spectrum in an intelligent way while not interfering with other incumbent devices in the frequency bands already licensed for specific uses. For detailed overview of the dynamic spectrum access schemes for cognitive radio networks, one may refer to the new comprehensive surveys in [91, 92].

WirelessHUMAN^{TM} PHY not only complies with the WirelessMAN-SCa PHY,

the WirelessMAN-OFDM PHY and the WirelessMAN-OFDMA PHY, but also further complies with the Dynamical Frequency Selection (DFS) protocols. WirelessHUMAN^{TM} PHY is implemented for license-exempt frequencies below 11 GHz. The license-exempt nature introduces new mechanisms such as DFS to detect and avoid interference. Hence, it provides the platform whereby Cognitive Radio can perform.

5.2.2 Cross-Layer Cognitive Radio-Based QoS Support Framework

As shown in Figure 2.2, IEEE 802.16d left the details of the CAC algorithm undefined. The existing proposed CAC schemes operate under the Fixed Spectrum Allocation (FSA) scheme with a pre-defined fixed spectrum and a fixed channel bandwidth. In order to overcome this problem, the function modules of Cognitive Radio need to be embedded in the current QoS architecture in WiMAX systems. The performance of the MAC protocol, the functionality of the scheduler and the CAC algorithm are influenced by the time-varying wireless channel in WiMAX systems. With a cross-layer approach, a new interface of Symbol_Control_Info needs to be designed to link up AMC scheme and CAC from the PHY to the MAC layer.

With a cross-layer approach, the QoS provisioning architecture is proposed as shown in Figure 5.4. It aims at providing CAC control functionality and spectrum spreading functionality in the WiMAX WirelessHUMAN^{TM}-OFDM system.

The proposed cross-layer CR-based QoS support framework with the corresponding CR-based CAC (CRCAC) scheme is described as follows.

Figure 5.4 Cross-layer CR-based QoS support framework for WiMAX

1. Considering the impact of PHY air interface on MAC layer protocol, a wireless Channel Condition Estimator (CCE) is designed at the PHY layer at the BS as well as at SS_u , where $u=1,2,3,\dots,U$. It monitors instantaneous channel condition status, SINR, $\gamma(u, BS)$, when the BS receives signal from SS_u or SINR, $\gamma(BS, u)$, when SS_u receives a message from the BS.
2. Considering dynamic spectrum allocation and spectrum spreading functionality of the system, a CR-based Intelligent Spectrum Management Module (ISMM) is designed at the PHY layer at the BS as well as at all SSs. It scans White Spaces and monitors the status of $WSS(BS, x)$ or $WSS(u, x)$ which is a list of the spectrums not used by primary users around the BS or SS_u , where x is non-active spectrum index.

3. Based on the information of channel's status provided by CCE, the BS or SS PHY Symbol Rate Controller, which consists of FEC, Symbol Mapper and AMC Controller, selects a feasible modulation and coding scheme $AMC(u, BS, q)$ for each sub-carrier for UL data transmission from SS_u to the BS. Modulation and coding level q is set according to SINR, $\gamma(u, BS)$. It then computes a data transmission rate over a frame based on the algorithm described later in Section 5.2.3. The PHY Symbol Rate Controller at the BS forms $Symbol\_Control\_Info_{BS} \{\gamma(u, BS), AMC(u, BS, q), WSS(BS, x)\}$. It is forwarded to MAC layer via PHY Service Access Point (SAP) for CAC. Similarly, the PHY Symbol Rate Controller at SS_u forms $Symbol\_Control\_Info_{SS} \{\gamma(BS, u), AMC(BS, u, q), WSS(u, x)\}$. SS_u embeds it in a $Polling\_response(u)$ and sends the $Polling\_response(u)$ to the BS when SS_u is polled.
4. The BAEM in the CRCAC at the BS takes in $Symbol\_Control\_Info_{SS}$ information from $Polling\_response(u)$ message and bandwidth allocation information from DL/UL_Map. After the system executes a few frames of time, the CRCAC conducts a statistic analysis of bandwidth allocation information from DL/UL_Maps, enabling it to estimate the bandwidth consumed by every SS. Then the CRCAC estimates the effective UL bandwidth of the system $\overline{BW}_{uplink}$ based on (5.21) and computes the average packet delay of UGS and rtPS service class, $\overline{D}_{ugs}$, $\overline{D}_{rtPS}$ based on (5.23) described later in Section 5.2.4. The CRCAC makes a decision on whether a new request of a flow can be admitted in the ACL or not based on the proposed cross-layer CRCAC algorithm described in detail in Section 5.2.3.

The BS maintains the ACL and updates it periodically.

5. The BSM at the BS takes in White Space Status $WSS(u, x)$ information from the $Polling\_response(u)$ message. It compares its $WSS(BS, x)$ with every $WSS(u, x)$ of SS_u where $u=1,2,3,\dots,U$ to find the same value of x which means that the common non-active spectrum is detected by BS and all SSs. The BSM at the BS sends Bandwidth_Spreading_Info to CR-based Intelligent Spectrum Management Module (ISMM) at the BS and all SSs via DL/UL_Map. If the channel bandwidth in a spectrum is used out, the CRCAC triggers CR-based ISMM. The CR-based ISMM makes spectrum x active. The system can use its own spectrum together with spectrum x for data transmission. In other words, the system now operates under two spectrums.

5.2.3 Cross-Layer Cognitive Radio-Based CAC Scheme

From the system's point of view, the system must admit the connections of all service classes as much as possible in order to maximize the system revenue. The system should be able to exploit the unused spectrum portions and spread spectrums over the White Spaces when the spectrums around the BS and SSs are in low-usage.

Wireless channel is a time-varying channel in which the channel capacity or bandwidth dynamically fluctuates according to the SINR of PHY. Thus, the effect of AMC on CAC scheme must be considered.

Based on the above considerations, the cross-layer CR-based CAC (CRCAC) scheme is proposed. A flowchart to illustrate the procedure of CRCAC is shown in Figure 5.5.

Figure 5.5 CRCAC flowchart

The procedure of CRCAC is elaborated based on the UL channel. However, the CRCAC can be easily extended to the DL channel. When a traffic flow sends its request to join the system, the CRCAC extracts its MSTR and ML two QoS requirement parameters from its AdmittedQoSParameterSet and makes a decision on whether to reject or to admit the flow into the network according to the following procedure:

Step 1. Initially, the system operates under its own spectrum, say, Spectrum 1.

The BAEM of CRCAC at the BS takes $AMC(u, BS, q)$ information from $Symbol\_Control\_Info_{BS} \{\gamma(u, BS), AMC(u, BS, q), WSS(BS, x)\}$ and bandwidth allocation for every SS from DL/UL_Map and computes the effective UL bandwidth $\overline{BW}_{uplink}$, $\overline{D}_{ugs}$ and $\overline{D}_{rtPS}$ for Spectrum 1.

Step 2. When a request of a new flow is sent from SS_u to the BS, the CRCAC calculates the remaining effective uplink bandwidth:

$$Effective\_UL\_BW_{remain} = \overline{BW}_{uplink} - BW_{UGS} - BW_{rtPS} - BW_{nrtPS} - BW_{BE} \quad (5.10)$$

Where BW_{UGS} , BW_{rtPS} , BW_{nrtPS} and BW_{BE} is the bandwidth occupied by each service class.

Step 3. The QoS_CM performs its first function QoS_CM(1) (QoS Control Module Function 1) to examine the bandwidth requirement of the new flow. $Effective\_UL\_BW_{remain}$ is compared with the MSTR requirement of the new flow. If $Effective\_UL\_BW_{remain} > MSTR$, go to Step 5 to check up the ML requirement of the new flow. If $Effective\_UL\_BW_{remain} < MSTR$, go to Step 4 to check on the possibility of spectrum spreading in order to expand the system capacity.

Step 4. The CR-based ISMM provides the instant information of White Space status to the BSM. The BSM at the BS extracts White Space status $WSS(u, x)$ information of all SSs from their $Symbol\_Control\_Info_{ss}$. The CRCAC at the BS has the information of non-active spectrums around each SS. Then CRCAC compares $WSS(BS, x)$ at the BS with $WSS(u, x)$ of each SS_u to find the same value of x , which means that the common non-active spectrum is detected by BS and all SSs. If the BS can identify the

common non-active spectrum, then the system can use spectrum x together with its own spectrum for data transmission. In other words, the system can operate in spectrum x when the capacity of its own spectrum 1 is used out.

Theoretically, the potential range of spectrums can be huge as long as they are non-active, but if too many spectrums are used by one network system, interference can be caused by other surrounding networks or primary users. So the number of spectrum x is limited to one in this study in order to avoid the possible interference.

If the BSM cannot find a common non-active spectrum x , or the WiMAX system already occupied one more common non-active spectrum, the CRCAC will deny the flow. Otherwise, the CRCAC at the BS embeds spectrum spreading instruction in the DL/UL_Map and broadcasts it to all SSs. The system spreads to new spectrum x . The CRCAC makes a decision to reject or admit a new flow in spectrum x following the same CAC procedure in Spectrum 1. The newly arrived flow can be transmitted in spectrum x if it passes the QoS checkups.

Step 5. In fact, just a bandwidth checkup is not sufficient to guarantee the whole ActiveQoSParameterSet that is defined by the standard as a wireless channel is a time-varying channel. The QoS_CM performs its second function, QoS_CM(2) (QoS Control Module Function 2) to examine the ML requirement for both the new flow and the existing connections after the request of a new flow has successfully passed the bandwidth requirement checkup. Assuming that the new flow would be admitted in the network, it would join one of the UGS, rtPS, nrtPS and BE service

classes that it belongs to. $\bar{D}_{ugs}, \bar{D}_{rtPS}$ will be updated as traffic pattern and traffic load has changed. QoS_CM(2) will check the ML constraint of each existing UGS and rtPS connection and the possible admitted new flow. The function of QoS_CM(2) will achieve the result that the ML of all admitted UGS traffic connections plus the possible admitted new connection is not smaller than the calculated $\bar{D}_{ugs}$ or $\bar{D}_{ugs} \leq Min\_ML(UGS)$, where $Min\_ML(UGS)$ is the minimum ML requirement among all admitted UGS connections plus the possible admitted new connection. QoS_CM(2) will also achieve the result that the ML of all admitted rtPS traffic connections and the possible admitted new connection is not smaller than the calculated $\bar{D}_{rtPS}$ or $\bar{D}_{rtPS} \leq Min\_ML(rtPS)$, where $Min\_ML(rtPS)$ is the minimum ML requirement among all admitted rtPS connections and the possible admitted new connection. If both conditions are fulfilled, the new flow will be admitted. UL_BW_{remain} will be updated. Otherwise, the new flow will be blocked.

The corresponding average delay of UGS or rtPS obtained by queueing analytical model is less than the delay requirement of each admitted UGS or rtPS traffic connection respectively. The trade-off in the admission control design is to reach the delay guaranteed real-time service commitments as much as possible while keeping the network utilization high.

The proposed CRCAC has two advantages.

- Firstly, the proposed CRCAC mechanism is able to catch the PHY condition like AMC and global system performance like average packet

delay and effective bandwidth by using the AMC information from the PHY layer and bandwidth allocation assignment from the scheduler. Thus, the CRCAC can improve the accuracy of the decision.

- Secondly, the proposed CRCAC has CR-based intelligent spectrum management ability. It has the potential to utilize the large amount of unused spectrums in an intelligent way while not interfering with other networks. With such a spectrum spreading ability, the CRCAC can double the capacity of the WiMAX system.

5.2.4 Queueing Analytical Model for WiMAX OFDM System

To analyze a packet-level QoS in real-time and non-real-time data transmission, the system performance parameters like channel's data transmission rate (channel bandwidth) and average packet delay need to be investigated. A queueing analytical model can be used in an off-line manner to obtain those performance parameters of the network system for CAC to meet the desired QoS requirements, such as MSTR and ML. In this sub-section, queueing analytical modeling is carried out to analyze the WiMAX OFDM system. Based on the queueing model, an effective channel transmission rate and the average packet delay of UGS and rtPS service class are obtained.

5.2.4.1 WiMAX OFDM System Queueing Model

Based on the proposed CRCAC algorithm, the WiMAX system using OFDM with TDD duplex technique may operate on two spectrums. Without loss of

generality, the WiMAX system is modeled based on one spectrum. It is easy to extend to the system under two spectrums. As shown in Figure 5.6, the UL transmission in the system is focused [1]. The UL channel of the WiMAX OFDM system is a multiple-access communication channel shared by U SSs in TDMA mode. The UL sub-frame consists of a total of L OFDM symbols. Each symbol consists of N_{FFT} sub-carriers.

Figure 5.6 Data transmission in WiMAX OFDM with TDD duplex mode

Assume that the scheduler is to allocate each SS a certain number of symbols. All of N_{FFT} sub-carriers are allocated to a SS over an OFDM symbol in this study. The interval of an OFDM symbol is T_s seconds. Let the duration of an UL sub-frame be T_F . So that $T_F = L \times T_s$.

The CCE shown in Figure 5.4 provides the channel SINR to the BS and all SSs. The total transmission channel bandwidth per spectrum is C MHz, so that each sub-carrier has a bandwidth of $\Delta f = C/N_{FFT}$ MHz.

The number of bits transmitted during a symbol time can be different depending on the AMC scheme applied to each sub-carrier. By using the AMC scheme based on a sub-carrier SINR and a pre-set target BER, the maximum number of bits per symbol per Hz that sub-carrier i for SS_u can transmit to the BS, denoted by $E(i, u, BS)$, can be expressed as [86][93]:

$$E(i, u, BS) = \left\lfloor \log_2 \left(1 + \frac{-1.5}{\ln(5P_{ber})} \gamma(i, u, BS) \right) \right\rfloor \quad (5.11)$$

where $\gamma(i, u, BS)$ is the BS received instantaneous SINR of sub-carrier i corresponding to source SS_u at a symbol time. P_{ber} is the target BER.

For the wireless fading channel in this study, the channel quality can be captured by SINR. Since the transmission channel varies from frame to frame, the general Nakagami- m model is used to describe SINR statistically. For a Nakagami- m fading channel [84], the received SINR $\gamma(i, u, BS)$ per frame is thus a random variable with a Gamma probability density function:

$$p[\gamma(i, u, BS)] = \frac{m^m \gamma(i, u, BS)^{m-1}}{\gamma(i, u, BS)^m \Gamma(m)} \exp\left(-\frac{m\gamma(i, u, BS)}{\gamma(i, u, BS)}\right) \quad (5.12)$$

where $\overline{\gamma(i, u, BS)}$ is the average BS received SINR of sub-carrier i from source

SS_u . $\Gamma(m) := \int_0^\infty t^{m-1} e^{-t} dt$ is the Gamma function and m is the Nakagami fading

parameter ($m \geq 1/2$).

The AMC scheme adjusts its modulation and coding scheme; hence, data

transmission rate, on a frame-by-frame basis. The modulation and coding scheme controlled by AMC is independent from SSs and sub-carriers. It depends on a sub-carrier SINR and a pre-set target BER. The data transmission rate is an Adaptive Discrete Rate (ADR) [86] with different modulation and coding schemes. There are Q possible modulation and coding schemes available, numbered from $AMC(i, u, BS, 1)$ to $AMC(i, u, BS, Q)$. Let $E(i, u, BS, q)$ be the efficiency of modulation and coding scheme which represents the number of bits per Hz in a PHY symbol when modulation and coding scheme $AMC(i, u, BS, q)$ is chosen in sub-carrier i from source SS_u to the BS. The SINR can be partitioned into $Q + 1$ consecutive non-overlapping regions with boundary points denoted by $\gamma(i, u, BS, q)$ where $q = 0, 1, \dots, Q$.

Suppose that a target BER P_{ber} is set. Using (5.11), the region boundaries or switching thresholds to achieve the BER can be obtained as follows.

$$\gamma(i, u, BS, q) = \frac{(2^{E(i, u, BS, q)} - 1) \ln(5P_{ber})}{-1.5} \quad (5.13)$$

When $\gamma(i, u, BS, q) \leq \gamma(i, u, BS) \leq \gamma(i, u, BS, q+1)$, the modulation and coding scheme $AMC(i, u, BS, q)$ is applied to sub-carrier i of source SS_u . As a result, the modulation and coding efficiency $E(i, u, BS, q)$ can be achieved. Obviously, in order to avoid possible transmission loss, no packet will be transmitted if the modulation and coding scheme is $AMC(i, u, BS, 0)$.

The probabilities $p(i, u, BS, q)$ to choose a modulation and coding scheme $AMC(i, u, BS, q)$, hence $E(i, u, BS, q)$ for sub-carrier i of source SS_u in the region $[\gamma(i, u, BS, q), \gamma(i, u, BS, q+1))$, can be expressed as:

$$p(i, u, BS, q) = \int_{\gamma(i, u, BS, q)}^{\gamma(i, u, BS, q+1)} p[\gamma(i, u, BS)] d\gamma(i, u, BS) \quad (5.14)$$

5.2.4.2 Effective UL Channel Bandwidth Analysis for WiMAX OFDM Wireless Channel with AMC

In the WiMAX OFDM system, the packets are served by slotted OFDM sub-carriers. In an OFDM symbol time T_s , the number of bits $r(i, u, BS, q)$ that can be transmitted for sub-carrier i of source SS_u is generally defined as a function of the modulation and coding scheme efficiency $E(i, u, BS, q)$:

$$r(i, u, BS, q) = T_s \times \Delta f \times E(i, u, BS, q) \quad (5.15)$$

As $T_s = \frac{1}{\Delta f}$, $r(i, u, BS, q)$ is expressed as:

$$r(i, u, BS, q) = E(i, u, BS, q) \quad (5.16)$$

For source SS_u , a set of possible bit transmission rates during an interval of an OFDM symbol in sub-carrier i can be obtained as follows:

$$r(i, u, BS) = \{E(i, u, BS, 0) \times p(i, u, BS, 0), E(i, u, BS, 1) \times p(i, u, BS, 1), \dots, E(i, u, BS, Q) \times p(i, u, BS, Q)\} \quad (5.17)$$

The average bit transmission rate $\overline{r(i, u, BS)}$ that can be transmitted in sub-carrier i from source SS_u to the BS in the interval of an OFDM symbol, T_s , is:

$$\overline{r(i, u, BS)} = \sum_{q=1}^Q E(i, u, BS, q) \times p(i, u, BS, q) \quad (5.18)$$

In each OFDM symbol, the total bit rate $R(u, BS)$ (in the unit of bits/sec) that can be transmitted from source SS_u to the BS can be easily obtained by scaling the total number of sub-carriers N_{FFT} over an OFDM symbol time T_s . Thus, $R(u, BS)$ can be calculated as:

$$R(u, BS) = \left(\sum_{i=1}^{N_{FFT}} \sum_{q=1}^Q E(i, u, BS, q) \times p(i, u, BS, q) \right) / T_s \quad (5.19)$$

Assume that all sub-carriers have the same SINR during an interval of an OFDM symbol from a source SS_u as AMC function is based on a frame-by-frame basis. The same $AMC(u, BS, q)$ is chosen based on the BS received SINR from source SS_u . Hence, $R(u, BS)$ can be expressed as:

$$R(u, BS) = \frac{N_{FFT}}{T_s} \sum_{q=1}^Q E(u, BS, q) \times p(u, BS, q) \quad (5.20)$$

where $p(u, BS, q)$ is the probability to choose $AMC(u, BS, q)$, hence $E(u, BS, q)$ for source SS_u in the BS received SINR region $[\gamma(u, BS, q), \gamma(u, BS, q+1))$.

Assume that the UL channel bandwidth is 50% of the total bandwidth per spectrum. The effective UL channel bandwidth is influenced by inter-SS scheduling algorithm. Assume that the UL channel bandwidth is equally distributed to all U SSs by inter-SS schedule algorithm like Round Robin scheme. The effective UL channel bandwidth of the system is estimated as:

$$\overline{BW}_{uplink} = \frac{1}{2} \sum_{u=1}^U R(u, BS) / U \quad (5.21)$$

5.2.4.3 Average Packet Delay Analysis for Each Service Class in WiMAX OFDM System

Although the given results by the analysis of the M/G/1 model for wireless OFDM system are readily extensible to priority queues [94], the result of delay from the model is only referred to as queueing delay. A new approach is adopted to model the WiMAX OFDM system in this study. The WiMAX OFDM system is modeled as a multi-class priority TDMA queueing system since the WiMAX

system can provide QoS to service classes like UGS, rtPS, nrtPS and BE of different priority.

In order to calculate the average packet delay of each service class, each OFDM symbol is further divided into virtual mini-time slots theoretically just for delay analysis purpose. The UL sub-frame of WiMAX system can be modeled as a frame consisting of N consecutive virtual mini-time slots. Each virtual mini-time slot is normalized to unit length τ . So that $T_F = L \times T_s = N \times \tau$. The TDMA scheme for inter-SS applied in the system in which SS_u is allocated n_u virtual mini-time slots/frame ($u=1, 2, 3, \dots, U$) uniformly distributed in the UL sub-frame. The TDMA scheme for inter-class at a SS becomes a strict priority queue scheme in which higher priority packets are queued ahead of lower priority ones. The packets with the same priority class that arrive at different slots are served on a First-Come-First-Served basis. Each SS can transmit its packets only during its dedicated symbol (in virtual mini-time slots) which are synchronized and governed by the BS via the DL/UL_Map.

Packets arriving at each SS belong to one of the four priority classes: UGS (priority-1), rtPS (priority-2), nrtPS (priority-3) or BE (priority-4). At each SS, packet arrival is a Poisson arrival process so that λ_j is the average arrival rate of class j packets. The packet length of UGS is fixed. For tractability, it is assumed that the packet length of rtPS, nrtPS and BE is independent and identically distributed (i.i.d.). The average packet lengths of rtPS, nrtPS and BE service classes are fixed. All sub-carriers are used to transmit a packet. The packet can be transmitted in a different number of virtual mini-time slots according to its length. The number of virtual mini-time slots to transmit the k^{th} arriving packet of priority class j is denoted by B_k^j ($j = 1, 2, \dots, 4$). The time taken to transmit the k^{th}

packet belonging to class j is represented as σ_k^j .

Assume that the probability of collision of BE_requests, which compete for the BE transmission service, is zero in the queueing model. When the UL channel becomes available, any waiting packet with priority- a will be served before anyone with priority- b , if $a < b$. Assume that each SS is identical in general, the packet delay and queue-size behavior of any SS is statistically independent of that of any other SS. The packet delay behavior of an arbitrary single SS, say SS_I , by a 4-priority model, is investigated.

Denoted by D_k^j , the delay of the k^{th} packet with priority j , is expressed as:

$$D_k^j = W_k^j + \sigma_k^j + P_k^j + F_L \quad (5.22)$$

where W_k^j denotes the packet waiting-time, σ_k^j represents the total time to transmit the priority- j k^{th} packet. F_L denotes the packet frame latency. P_k^j denotes the packet propagation delay. Then the average packet delay of class j , $\overline{D_j}$, which can serve as the packet upper delay bound of each service class is given by [70]:

$$\overline{D_j} = W_{j,1} + \left\{ E[B_k^j] - \frac{N-1}{N} \right\} \left\lfloor \frac{N}{n_m} \right\rfloor \tau + P^j + F_L \quad (5.23)$$

where

$$W_{j,1} = \frac{\sum_{i=1}^J \lambda_i b_{i,2} + (1 - \eta_j) \left\lfloor \frac{N}{n_m} \right\rfloor \tau}{2(1 - \eta_j)(1 - \eta_{j-1})} \quad (5.24)$$

$$\left\lfloor \frac{N}{n_m} \right\rfloor \text{ represents the biggest integer not larger than } \frac{N}{n_m}.$$

$$\rho_j = \lambda_j b_{j,1} \quad (5.25)$$

$$\eta_j = \sum_{i=1}^j \rho_i \quad j=1, 2, 3, 4$$

$b_{j,1}$ and $b_{j,2}$ is the first moment and the second moment of the priority- j class packet service time in unit of virtual mini-time slots respectively. Now, $b_{j,1}$ and $b_{j,2}$ are going to be derived.

Using (5.15), the number of bits that can be transmitted for i sub-carrier from source SS_u during an interval of a virtual mini-time slot τ is:

$$r(i, u, BS, q, \tau) = \tau \times \frac{1}{T_s} \times E(i, u, BS, q, \tau) \quad (5.26)$$

As all sub-carriers are allocated for source SS_u to transmit its packets, let $R(u, BS, q, \tau)$ be the total number of bits that can be transmitted by all sub-carrier N_{FFT} from source SS_u based on $AMC(u, BS, q)$ during the interval of τ . As the distance between SS_u and the BS is fixed, it is assumed that the BS received SINR of every sub-carrier is the same so that the same $AMC(u, BS, q)$ is chosen in every sub-carrier. Let $E(u, BS, q, \tau)$ be the modulation and coding scheme efficiency when the modulation and coding scheme $AMC(u, BS, q)$ is chosen for source SS_u during the interval of τ . As the AMC scheme adjusts its modulation and coding scheme on a frame-by-frame basis, $E(u, BS, q, \tau)$ is equal to $E(u, BS, q)$ over a frame.

$R(u, BS, q, \tau)$ is expressed as:

$$R(u, BS, q, \tau) = \sum_{i=1}^{N_{FFT}} r(i, u, BS, q, \tau) = \frac{N_{FFT} \times \tau}{T_s} \times E(u, BS, q) \quad (5.27)$$

The number of virtual mini-time slots needed to transmit the k^{th} arriving packet of priority class j , $B_k^j(q)$, is:

$$B_k^j(q) = L_k^j / R(u, BS, q, \tau) = \frac{L_k^j \times Ts}{N_{FFT} \times \tau \times E(u, BS, q)} \quad (5.28)$$

where L_k^j is the packet length of the k^{th} arriving packet of priority class j in unit of bit.

For SS_u , a set of B_k^j can be obtained as follows:

$$B_k^j = \left\{ \frac{L_k^j \times Ts}{N_{FFT} \times \tau \times E(u, BS, 1)} \times p(u, BS, 1), \right. \\ \left. \frac{L_k^j \times Ts}{N_{FFT} \times \tau \times E(u, BS, 2)} \times p(u, BS, 2), \dots, \right. \\ \left. \frac{L_k^j \times Ts}{N_{FFT} \times \tau \times E(u, BS, Q)} \times p(u, BS, Q) \right\} \quad (5.29)$$

The average number of virtual mini-time slots needed to transmit the priority - j k^{th} class packet, $b_{j,k,1}$, is :

$$b_{j,k,1} = \sum_{q=1}^Q B_k^j(q) \times p(u, BS, q) \quad (5.30)$$

$$E(B_k^j) = \sum_{q=1}^Q \frac{L_k^j \times Ts}{N_{FFT} \times \tau \times E(u, BS, q)} \times p(u, BS, q) \quad (5.31)$$

$$b_{j,k,2} = \sum_{q=1}^Q (B_k^j(q))^2 \times p(u, BS, q) \quad (5.32)$$

$$b_{j,k,2} = E((B_k^j)^2) = \sum_{q=1}^Q \left(\frac{L_k^j \times Ts}{N_{FFT} \times \tau \times E(u, BS, q)} \right)^2 \times p(u, BS, q) \\ q=1, \dots, Q \quad (5.33)$$

In order to simplify the computing, the average packet delay of each service class is calculated. The average packet length of priority- j class, L^j , is used for

calculation. Average $b_{j,1}, b_{j,2}$ of the priority- j class packet service time can be obtained. Thus, the theoretical packet upper delay bound for each service class is obtained.

5.3 Performance Analysis

In this chapter, two novel connection admission control schemes, which are termed as ECAC and CRCAC respectively, are proposed. In order to evaluate the performance of the proposed schemes, the Complete Sharing (CS) and the Reservation-based CAC (R_CAC) [12] schemes are adopted as the counterparts for the purpose of comparison. The CS is a traditional CAC scheme in which the BS accepts a connection request whenever there is sufficient resource, and blocks the connection request otherwise. All the service types are treated equally. The BS accepts or rejects a connection request simply based on the available bandwidth in a First-Come-First-Served manner. The R-CAC is characterized by defining thresholds of an allocated bandwidth at the BS for each service class based on the priority of the service. C_u is the bandwidth exclusively reserved for UGS service, which is allocated with the highest priority in the four service classes supported in IEEE 802.16d systems. C_r is the bandwidth exclusively reserved for UGS and rtPS to mitigate the bandwidth competition coming from the other two types of service. The residual bandwidth (i.e., $C - C_u - C_r$) is the only part that a BS can assign to the nrtPS connection requests to meet their minimum bandwidth requirements, and it can also be assigned to UGS and rtPS service classes.

A series of simulation experiments have been carried out to evaluate the performance of the two proposed CAC schemes. The experiments have been

performed in two stages. In the first stage, the performance of ECAC and CRCAC in terms of connection blocking probability and the average number of on-going connections has been evaluated. In the second stage, the case studies of the ECAC and the CRCAC scheme have been carried out.

5.3.1 Performance Analysis for the Proposed ECAC Admission Control Scheme

5.3.1.1 Simulation Design

Following the signaling mechanism specified by the IEEE 802.16d standard, the WiMAX simulation model has been developed by using ExtendSim [71] software. The CAC schemes have been developed by using C language. The WiMAX network under study operates as the WirelessMAN-SC (Single Carrier) in TDD mode.

For wireless fading channel in the simulation model, the channel quality can be captured by a parameter, γ . Since the transmission channel varies from frame to frame, the general Nakagami- m model is adopted to describe γ statistically [84]. The received SINR, γ per frame is thus a random variable with a Gamma probability density function:

$$p_{\gamma}(\gamma) = \frac{m^m \bar{\gamma}^{m-1}}{\bar{\gamma}^m \Gamma(m)} \exp\left(-\frac{m\gamma}{\bar{\gamma}}\right) \quad (5.34)$$

where $\bar{\gamma}$ is the average received SINR, $\Gamma(m) := \int_0^{\infty} t^{m-1} e^{-t} dt$ is the Gamma function and m is the Nakagami fading parameter ($m \geq 1/2$). The channel is

assumed to be quasi-static so that the channel gains are constant over a frame. However, they are allowed to vary from frame to frame. The average received SINR, $\bar{\gamma}$ can be expressed as:

$$\bar{\gamma} = P_t - L_i - P_N - L_p \quad (5.35)$$

where P_t (dBm) is the transmitter output power, L_i (dBm) is the implementation loss due to hardware connecting cables, antenna patterns, etc., P_N (dBm) is the receiver noise power, which is related to the hardware noise figure and bandwidth, and L_p (dBm) is the path loss due to radio propagation.

The network has been designed in PMP topology with one BS and eight SSs. The positions of SSs are assumed independent and distributed randomly. Since there is no Maximum Latency requirement for BE traffic flows, only UGS, rtPS and nrtPS traffic flows will be considered in the simulation study. The threshold Th_{UGS} is set to 6.85% of total uplink bandwidth, Th_{rtPS} and Th_{nrtPS} are set to 46.57% of total uplink bandwidth. The traffic parameters are selected from the supported range of values which fully follow the required values for multimedia services defined in the Recommendation ITU-R M.1225 [85].

The inter-arrival time of UGS packets is constant. Packet length is fixed to 200 bytes. The deadline is set to 50ms.

The rtPS and nrtPS connections are burst traffic flows. The rtPS and nrtPS connection arrivals follow a Poisson process and the connection holding time is exponentially distributed. In the performance evaluation scenarios, the connection arrival rate for both rtPS and nrtPS connections varies in the range of 1-24 connections per second. The average connection holding time is set to 1.250 S. The mean Maximum Sustained Traffic Rate is set to 371.5 Kbps for both rtPS

and nrtPS connections and it is exponentially distributed. The mean MRTR is set to 297.2 Kbps for both rtPS and nrtPS connections and it is exponentially distributed. The mean Maximum Latency (ML) is set to 50 ms for rtPS connections and it is exponentially distributed. There is no ML requirement of nrtPS connections. The packet arrival in rtPS and nrtPS connections is a Poisson process. Packet length is exponentially distributed with mean value of 500 bytes.

The frame size is set to 10 ms with 500 time slots per frame. The duration of each time slot is 20 μ s. Channel bandwidth is assumed to be 25 MHz.

The parameters for R_CAC scheme are set as follows.

- C_U is set to 1.37 Mbps. It is exclusively reserved for UGS.
- C_r is set to 50% of $(C - C_U)$. It is reserved for rtPS.

Based on the above parameters set for the simulation, the numerical results of the average packet delay for the three service classes can be calculated by using (5.7). The results are tabulated as shown in Table 5.1.

5.3.1.2 Simulation Results

Since the UGS connections have been allocated with a fixed bandwidth (fixed time duration) based on their fixed bandwidth requirement in the framework of the IEEE 802.16d standard, the simulation results of rtPS and nrtPS with CS_CAC, R_CAC [12] and the ECAC schemes are analyzed. The performance metrics for the evaluation are connection blocking probability and the average number of on-going connections. The results are shown in Figure 5.7, 5.8, 5.9, 5.10, and 5.11.

TABLE 5.1 AVERAGE DELAY OF UGS, rtPS AND nrtPS

<i>Load</i>	<i>UGS</i> (<i>Mbps</i>)	<i>rtPS</i> (<i>Mbps</i>)	<i>nrtPS</i> (<i>Mbps</i>)	λ_1 (<i>mean</i>) <i>Packet/s</i>	$\lambda_{2,3}$ (<i>mean</i>) <i>Packet/s</i>	<i>UGS</i> <i>Delay</i>	<i>rtPS</i> <i>Delay</i>	<i>nrtPS</i> <i>Delay</i>
0.1	1.37	0.315	0.315	107.03	9.84	6.14	7.91	8.72
0.2	1.37	1.315	1.315	107.03	41.09	6.14	10.48	13.98
0.3	1.37	2.315	2.315	107.03	72.34	6.14	13.09	19.44
0.4	1.37	3.315	3.315	107.03	103.59	6.14	15.73	25.12
0.5	1.37	4.315	4.315	107.03	134.84	6.14	18.41	31.02
0.6	1.37	5.315	5.315	107.03	166.09	6.14	21.11	37.15
0.7	1.37	6.315	6.315	107.03	197.34	6.14	23.85	43.53
0.8	1.37	7.315	7.315	107.03	228.59	6.14	26.63	50.17
0.9	1.37	8.315	8.315	107.03	259.84	6.14	29.45	57.07
0.96	1.37	8.915	8.915	107.03	278.59	6.14	31.12	61.35

Figure 5.7 shows the relation of the connection blocking probability (CBP) versus the arrival rates of the connection requests for rtPS service class. As expected, the connection blocking probability of the three CAC schemes increases as the connection arrival rate increases. However, CBP of rtPS with the R_CAC scheme is lower than that with the CS scheme. It is because with the R_CAC scheme, a fraction of the total bandwidth at BS is reserved for the rtPS traffic. The rtPS request has a higher priority than that of the nrtPS request, so a fraction of the total bandwidth will be reserved for the rtPS request but not for the nrtPS request. Therefore, a rtPS request has a higher probability of being accepted than a nrtPS request.

Figure 5.7 rtPS CBP vs. connection arrival rate

It can be observed that the proposed ECAC outperforms the other two CAC schemes. This is due to the function of Bandwidth Elastic Module (BEM) in the proposed ECAC scheme. The Bandwidth Elastic Module of ECAC sets the bandwidth-borrowing algorithm so that a higher service class can borrow bandwidth from a lower service class when the lower service class consumes the bandwidth more than its threshold. When nrtPS service class uses more bandwidth than its threshold, the ECAC can compress the bandwidth used by the nrtPS service class by re-allocating MRTR to the nrtPS connections. The ECAC can borrow the bandwidth of BW_{Low} from nrtPS. The ECAC can allocate the BW_{Low} to rtPS. Therefore, the CBP of rtPS with the proposed ECAC scheme is lower than that with the R_CAC scheme.

Figure 5.8 shows the relation of the CBP versus the arrival rates of the connection requests for the nrtPS service class. The CBP of nrtPS with the R-

CAC scheme has the highest CBP among the three CAC schemes.

Figure 5.8 nrtPS CBP vs. connection arrival rate

Since a portion of the total bandwidth is reserved for the service class with a higher priority, (i.e., UGS or rtPS), the bandwidth for nrtPS is less. Furthermore, when the connection arrival rate of rtPS increases, some nrtPS connections are terminated in order to release the bandwidth for rtPS service class. It results in a higher CBP of nrtPS with the R_CAC scheme. However, the bandwidth-borrowing mechanism in the ECAC is triggered only when the nrtPS service class consumes the bandwidth more than its threshold. The nrtPS connections are not terminated, instead, the connections are maintained by allocating bandwidths based on their MRTR requirements. In short, the ECAC can compress the bandwidth used by the nrtPS service class by re-allocating MRTR to the connections. Once the bandwidth consumed by nrtPS is compressed, there is

more room to accept rtPS and nrtPS connections. Hence, the CBP of nrtPS with the ECAC scheme is lower and crosses over that with the CS_CAC scheme when the bandwidth used by the nrtPS service class exceeds its threshold.

Figure 5.9 and 5.10 illustrate the relation of the average number of on-going connections versus the arrival rates of the connection requests for rtPS and nrtPS service classes, respectively. It is observed that the performance of the three CAC schemes is similar when the arrival rate of the connection request is light. It is true that when the traffic load is light, the network has enough resources to accommodate the traffic flows and to establish the connections.

Figure 5.9 Average number of rtPS on-going connections vs. connection arrival rate

It is observed that the average number of on-going connections of rtPS with the R_CAC scheme is slightly larger than that with the CS_CAC scheme when

the arrival rate of the connection request increases. It is also clear that the average number of on-going connections of nrtPS with the R_CAC scheme is slightly less than that with the CS_CAC scheme when the arrival rate of the connection request increases. The trade-off of R_CAC is to set the rtPS request at a higher priority than that of the nrtPS request, so a fraction of the total bandwidth will be reserved for the rtPS service class. The R_CAC would terminate nrtPS connections in order to admit more rtPS connections.

Figure 5.10 Average number of nrtPS on-going connections vs. connection arrival rate

Figure 5.11 illustrates the relation of the average number of on-going rtPS and nrtPS connections versus the arrival rates of the connection requests for both rtPS and nrtPS service classes. It is true that the average number of on-going connections of both rtPS and nrtPS with the R_CAC scheme is slightly less than

that with the CS_CAC scheme when the arrival rate of the connection request increases. The R_CAC can reserve a portion of the total bandwidth for a service class with a higher priority (i.e., UGS or rtPS). The average number of on-going connections of rtPS with the R_CAC scheme is slightly larger than that with the CS_CAC scheme.

Figure 5.11 Average number of rtPS& nrtPS on-going connections vs. connection arrival rate

Nevertheless, The R_CAC dynamically assigns and releases the existing nrtPS connections; consequently, the average number of on-going connections of nrtPS with the R_CAC scheme is reduced when the arrival rate of the connection request increases. This causes the total average number of on-going connections to become less. However, the ECAC has the largest total average number of on-

going connections among the three CAC schemes. The ECAC can compress the bandwidth used by the nrtPS service class by re-allocating MRTR to the nrtPS connections when the nrtPS service class uses more bandwidth than its threshold. The ECAC can assign the released bandwidth from the nrtPS service class to the rtPS service class. The ECAC can admit more rtPS connections while maintaining nrtPS connections.

5.3.1.3 Case Study for the ECAC Admission Control Scheme

In the second stage of the simulation, the case study for the ECAC scheme is carried out. It is to evaluate the cooperation between HOS opportunistic scheduling and ECAC. It also shows the functionality of the ECAC scheme in detail.

Since the proposed ECAC works closely with the HOS scheduling algorithm, an effective UL channel bandwidth can be obtained by (5.4). $Effective\_UL\_BW_{remain}$ can be obtained by DL/UL_Map which is determined by the HOS. Average packet delay $\overline{D}_{ugs}$, $\overline{D}_{rtPS}$ can be obtained by (5.7). With these two parameters, the admission control algorithm can estimate the available network bandwidth and the average packet delay of each class for both existing connections and the newly arrived one, and make a decision to admit or reject the request of the new connection.

It is clear that when the traffic load is light, the network has enough resources to accommodate the traffic flows and establish the connections. The functionality of the ECAC algorithm is evaluated when traffic load is heavy, e.g. 96% of the total effective UL channel bandwidth.

The case study consists of three experiments. Three experiments have been conducted to examine the proposed admission control algorithm when traffic load is 96% of the total effective UL channel bandwidth. Three experiments are designed with three scenarios, each one of the three experiments reflects one scenario. Initially, each SS has a total of eight traffic flows requesting transmission service. The detailed requirements of traffic MSTR, MRTR and ML are shown in Table 5.2.

TABLE 5.2 INITIAL INPUT TRAFFIC FOR THE ECAC SCHEME

<i>Traffic Flow</i>	<i>Maximum Sustained Traffic Rate(kbps)</i>	<i>Minimum Reserved Traffic Rate(kbps)</i>	<i>Maximum Latency (ms)</i>	<i>ECAC Decision</i>
SS1_UGS_SFID 1	85.6	85.6	50	Admitted
SS1_UGS_SFID 2	85.6	85.6	50	Admitted
SS1_rtPS_SFID 1	371.5	297.2	40	Admitted
SS1_rtPS_SFID 2	371.5	297.2	40	Admitted
SS1_rtPS_SFID 3	371.5	297.2	50	Admitted
SS1_nrtPS_SFID 1	371.5	297.2	400	Admitted
SS1_nrtPS_SFID 2	371.5	297.2	400	Admitted
SS1_nrtPS_SFID 3	371.5	297.2	400	Admitted

The effective UL channel bandwidth is 20 Mbps. Th_{UGS} is set to 1.37 Mbps. Th_{rtPS} and Th_{nrtPS} is pre-set to 9.315 Mbps. In scenario one, a new rtPS flow sends its request to join the network. In scenario two, a new nrtPS flow sends its request. Finally, a new rtPS flow sends its request in scenario three as shown in

Table 5.3.

TABLE 5.3 NEW ARRIVAL INPUT TRAFFIC FOR THE ECAC SCHEME

<i>Traffic Flow</i>	<i>Maximum Sustained Traffic Rate(kbps)</i>	<i>Minimum Reserved Traffic Rate(kbps)</i>	<i>Maximum Latency (ms)</i>	<i>ECAC Decision</i>
SS1_rtPS_SFID _new1	512	409.6	30	Rejected
SS1_nrtPS_SFID _new	512	409.6	400	Admitted
SS1_rtPS_SFID _new2	512	409.6	40	Admitted

The total bandwidth request is the sum of bandwidth requests of all traffic flows, which is 19.2 Mbps. $Effective\_UL\_BW_{remain}$ of the network is 0.8 Mbps. $\bar{D}_{ugs}$ is 6.14 ms, $Min\_ML(UGS)$ is 50 ms, so, $\bar{D}_{ugs} \leq Min\_ML(UGS)$. $\bar{D}_{rtPS}$ is 29.45 ms, $Min\_ML(rtPS)$ is 40 ms, $\bar{D}_{rtPS} \leq Min\_ML(rtPS)$. Both maximum sustained traffic rate and maximum latency requirements can be met, all traffic connections from the eight SSs can be admitted as shown in Table 5.2.

If a new rtPS flow, e.g. rtPS traffic SS1_rtPS_SFID_new1, sends its request to the BS to establish a connection, the ECAC will assume that the new traffic flow would be added in the ACL. New $Effective\_UL\_BW_{remain}$ is 0.288 Mbps, $\bar{D}_{rtPS}$ is 31.23 ms, $Min\_ML(rtPS)$ is 30 ms. $\bar{D}_{rtPS} > Min\_ML(rtPS)$. The ECAC makes a decision to reject the request.

Then if one new nrtPS traffic SS1_nrtPS_SFID_new sends its request, $Effective\_UL\_BW_{remain}$ is 0.288 Mbps, $\bar{D}_{ugs} \leq Min\_ML(UGS)$ and $\bar{D}_{rtPS} \leq$

$Min\_ML(rtPS)$. The new nrtPS flow can be admitted into the ACL.

Now there are a total of 65 connections in the ACL. $Effective\_UL\_BW_{remain}$ becomes 0.288 Mbps. If a new SS1_rtPS_SFID_new2 sends its request, $Effective\_UL\_BW_{remain}$ will be less than MSTR, which is 0.512 Mbps. Then the ECAC turns on Bandwidth Elastic Module to find that the service priority below that of the new flow is nrtPS service class. The ECAC further finds that the bandwidth used by nrtPS service class is 9.428 Mbps which exceeds its threshold Th_{nrtPS} . Then the ECAC can compress the bandwidth occupied by nrtPS service class by re-allocating the MRTR to its connections and borrow $BW_{low} = BW_{nrtPS} - Th_{nrtPS} = 0.512$ Mbps. Now, $Effective\_UL\_BW_{remain}$ is 0.8 Mbps. $\overline{D}_{ugs} \leq Min\_ML(UGS)$. $\overline{D}_{rtPS}$ is 31.23 ms, $Min\_ML(rtPS)$ is 40 ms, $\overline{D}_{rtPS} \leq Min\_ML(rtPS)$. SS1_rtPS_SFID_new2 can be admitted as shown in Table 5.3. Now, the total number of service connections in the ACL is 66.

Above simulation experiments have shown that the proposed ECAC in tandem with HOS can provide the delay guaranteed service to the real-time traffic while prevent service starvation to the nrtPS traffic.

5.3.2 Performance Analysis for the CRCAC Admission Control Scheme

5.3.2.1 Simulation Design

The simulation model for WiMAX system has been designed in PMP topology with one BS and eight SSs. The WiMAX system under study operates as WirelessHUMAN^{TM}- OFDM in TDD duplex mode. The positions of SSs are

assumed to be independent and distributed randomly. Since BE traffic does not have ML requirement, the simulation study focuses on UGS, rtPS and nrtPS traffic. The traffic parameters are selected from the supported range of values which fully cover the required values for multimedia services defined in Recommendation ITU-R M. 1225[85].

The inter-arrival time of UGS packets is constant. Packet length is fixed to 200 bytes. The total traffic arrival rate of UGS is fixed to 1.37 Mbps.

The rtPS and nrtPS traffic are burst traffic flows. The rtPS and nrtPS connection arrivals follow a Poisson process and the connection-holding time is exponentially distributed. The mean MSTR is set to 371.5 Kbps for both rtPS and nrtPS connections and it is exponentially distributed. The mean MRTR is set to 297.2 Kbps for both rtPS and nrtPS connections and it is exponentially distributed. The maximum number of rtPS and nrtPS connections, which the system can admit simultaneously in the UL channel of one spectrum, is assumed to be 50. In order to show the superior performance of the CRCAC, the connection arrival rates for both rtPS and nrtPS connections are overloaded in the UL channel of one spectrum. The connection arrival rates of both rtPS and nrtPS connections vary in the range of 1-38 connections per second in the performance evaluation scenarios. The average connection holding time is set to 1.250 S. The mean Maximum Latency (ML) is set to 50 ms for rtPS connections and it is exponentially distributed. Since a non-real-time packet in nrtPS connections has no maximum latency requirement, the maximum waiting time for a non-real-time packet can be set as the TCP retransmission time-out value TO at 400 ms for wireless channel [80].

The packet arrival in the rtPS and nrtPS flows is a Poisson process with packet

inter-arrival time exponentially distributed. The packet arrival rates of rtPS and nrtPS flows are the same and they are changed according to the different percentage of traffic loads. The mean packet length is set to 500 bytes for both rtPS and nrtPS flows.

The frame size is set to 10 ms, 500 virtual mini-time slots per frame. The duration of each virtual mini-time slot is 20 μ s.

The parameters for OFDM PHY layer in license-exempt frequency band below 11 GHz for simulation experiments are outlined in Table 5.4.

TABLE 5.4 SIMULATION PARAMETERS BASED ON IEEE 802.16D
WIRELESSHUMAN^{TM} OFDM SYSTEM

<i>Parameters</i>	<i>Value</i>
Bandwidth C per spectrum	20 MHz
Sampling rate F_s	23.04MHz
Number of sub-carrier N_{FFT}	256
Sub-carrier spacing Δf	23.04MHz/256
Symbol time	13.89 μ s
UL sub-frame length	5 ms
Adaptive Modulation Scheme	Squared M -QAM modulations ($M=4, 16, 64$)
Target BER P_{ber}	$\leq 10^{-5}$
Wireless channel model	Nakagami- m fading channel ($m=1$)

When $m = 1$, the received SINR is an exponential random variable, with the

following probability density function (*pdf*) in the Nakagami-m fading channel:

$$p[\gamma(i, u, BS)] = \frac{1}{\gamma(i, u, BS)} \exp\left(-\frac{\gamma(i, u, BS)}{\gamma(i, u, BS)}\right) \quad (5.36)$$

The AMC is limited to the squared M -QAM modulations where M is 4, 16 and 64. The discrete-rate adaptive modulation method is adopted in the simulation. Then, $E(u, BS, q) = 2 \times q$ as $M = 2^{2q}$ [86]. By substituting $q = 0, 1, 2, 3$, and the target BER $P_{ber} \leq 10^{-5}$ in (5.13), four region boundaries, $[\gamma(u, BS, q), \gamma(u, BS, q+1))$, where $q=0, 1, 2, 3$ to achieve the target BER $P_{ber} \leq 10^{-5}$ can be obtained.

By substituting the obtained four region boundaries $[\gamma(u, BS, q), \gamma(u, BS, q+1))$, $q=0, 1, 2, 3$ in (5.14), the probabilities $p(u, BS, q)$ of choosing a modulation and coding scheme $AMC(u, BS, q)$, hence $E(u, BS, q)$, for source SS_u in the region $[\gamma(u, BS, q), \gamma(u, BS, q+1))$ can be obtained. By substituting the parameters according to Table 5.4 and the result of probabilities $p(u, BS, q)$ for source SS_u in every region $[\gamma(u, BS, q), \gamma(u, BS, q+1))$, $q=0, 1, 2, 3$ in (5.20), the total data transmission rate (in the unit of bits/sec) for every SS during the interval of an OFDM symbol can be easily obtained. The effective UL channel bandwidth or data transmission rate can be obtained by (5.21), which is 18.45 Mbps. With different traffic loads in one spectrum, the average packet delay $\bar{D}_{ugs}, \bar{D}_{rtPS}$ can be obtained by (5.23). The results are tabulated in Table 5.5.

With $Effective\_UL\_BW_{remain}, \bar{D}_{ugs}, \bar{D}_{rtPS}$ three parameters, the proposed CRCAC algorithm can make a decision to admit or reject the request of a new flow.

TABLE 5.5 NUMERICAL RESULTS OF AVERAGE PACKET DELAY OF UGS AND rtPS
IN ONE SPECTRUM

<i>Load</i>	<i>UGS</i> (Mbps)	<i>rtPS</i> (Mbps)	<i>nrtPS</i> (Mbps)	<i>Delay</i> (UGS)(ms)	<i>Delay</i> (rtPS)(ms)
0.1	1.37	0.24	0.24	5.46	6.16
0.2	1.37	1.16	1.16	5.46	6.84
0.3	1.37	2.08	2.08	5.46	7.53
0.4	1.37	3.01	3.01	5.46	8.22
0.5	1.37	3.93	3.93	5.46	8.91
0.6	1.37	4.85	4.85	5.46	9.61
0.7	1.37	5.77	5.77	5.46	10.32
0.8	1.37	6.70	6.70	5.46	11.03
0.9	1.37	7.62	7.62	5.46	11.74
0.98	1.37	8.36	8.36	5.46	12.31

5.3.2.2 Simulation results for the CRCAC Scheme

Since the UGS connections have been allocated with a fixed bandwidth (fixed time duration) based on their fixed bandwidth requirement in the framework of the IEEE 802.16d standard, the simulation results of rtPS and nrtPS with CS_CAC, R_CAC [12] and CRCAC schemes are analyzed. The performance

metrics for the evaluation are the connection blocking probability and the average number of on-going connections. The results are shown in Figure 5.12, 5.13, 5.14, 5.15, and 5.16.

Figure 5.12 shows the relation of the connection blocking probability (CBP) versus the arrival rates of the connection requests for rtPS service class. As expected, the connection blocking probability of the three CAC schemes increases when the connection arrival rate increases. However, the CBP of rtPS with the R_CAC scheme is slightly lower than that with the CS scheme. It is because with the R_CAC scheme, C_r , which is a fraction of the total bandwidth at the BS, is reserved for the rtPS traffic. The rtPS request has a higher priority than that of the nrtPS request so a fraction of the total bandwidth will be reserved for the rtPS request but not for the nrtPS request. Therefore, a rtPS request has a higher probability of being accepted than a nrtPS request.

Figure 5.12 rtPS CBP vs. connection arrival rate

It is observed that both the CBP of rtPS with R_CAC and CS_CAC increase sharply when the arrival rate of requests of rtPS for UL transmission is more than 24 connections per second. This is due to the UL channel being overloaded when the arrival rate of requests of rtPS for UL transmission is more than 24 connections per second.

However, it is observed that the proposed CRCAC outperforms the other two CAC schemes. This is due to the function of the intelligent Bandwidth Spreading Module (BSM) in the proposed CRCAC. The CRCAC is equipped with an intelligent Bandwidth Spreading Module (BSM). The BSM links to CR-based Intelligent Spectrum Management Module (ISMM) at the PHY layer. It exploits unused spectrum portions. When traffic load increases, for example, the arrival rate of the requests of rtPS for UL transmission is more than 14 connections per second, the CRCAC triggers the CR-based ISMM. The CR-based ISMM makes spectrum x active. The system can use its own spectrum together with spectrum x for data transmission. In other words, the system now operates under two spectrums. The capacity of UL transmission is doubled. Therefore, the CBP of rtPS with the proposed CRCAC scheme is much lower than that with the R_CAC scheme.

Figure 5.13 shows the relation of the CBP versus the arrival rates of the connection requests for nrtPS service class. The CBP of nrtPS with the R-CAC scheme has the highest CBP among the three CAC schemes.

Since a portion of the total bandwidth is reserved for service with a higher priority, (i.e., UGS or rtPS), the bandwidth for nrtPS with the R_CAC scheme is less. Furthermore, when the connection arrival rate of rtPS increases, some nrtPS connections will be terminated in order to release the bandwidth for rtPS. This

results in a higher CBP of nrtPS by the R_CAC scheme.

However, it is clear that the proposed CRCAC outperforms the other two CAC schemes. The reason is similar to the explanation for Figure 5.12. When the system operates under two spectrums, its capacity of UL transmission will be doubled. Therefore, the CBP of nrtPS with the proposed CRCAC scheme is much lower than that with the R_CAC scheme.

Figure 5.13 nrtPS CBP vs. connection arrival rate

Figure 5.14 and 5.15 illustrate the relation of the average number of on-going connections versus the arrival rates of the connection requests for rtPS and nrtPS, respectively. It is observed that the performance of the three CAC schemes is similar when the arrival rate of the connection requests is light. It is true due to the fact: when the traffic load is light, the network has enough resources to

accommodate the traffic flows and to establish connections. It is observed that the average number of on-going connections of rtPS with the R_CAC scheme is slightly larger than that with the CS_CAC scheme when the arrival rate of the connection requests increases. It is also observed that the average number of on-going connections of nrtPS with the R_CAC scheme is slightly less than that with the CS_CAC scheme when the arrival rate of the connection requests increases. The trade-off of R_CAC is to set the rtPS request with a higher priority than that of the nrtPS request so a fraction of the total bandwidth will be reserved for the rtPS service class. The R_CAC would terminate nrtPS connections in order to admit more rtPS connections.

Figure 5.14 Average number of rtPS on-going connections vs. connection arrival rate

Figure 5.15 Average number of nrtPS on-going connections vs. connection arrival rate

When the arrival rate of the connection requests is more than 24 connections per second, the average number of on-going connections of rtPS or nrtPS remains constant. This is because the system has reached its UL transmission limit.

However, the CRCAC outperforms the other two CAC schemes. When the arrival rate of the connection requests is more than 14 connections per second, the CR-based system can operate under more than one spectrum. As a result, the system throughput can be improved significantly. Thus, the average number of on-going connections of rtPS or nrtPS with the CRCAC scheme increases by 90% compared with that with the CS_CAC scheme or the R_CAC scheme when the arrival rate of the connection requests is 38 connections per second.

Figure 5.16 illustrates the relation of the average number of on-going rtPS and

nrtPS connections versus the arrival rates of the connection requests for both the rtPS and the nrtPS service class. It is observed that the average number of on-going connections of both rtPS and nrtPS with the R_CAC scheme is slightly less than that with the CS_CAC scheme when the arrival rate of the connection requests increases. The R_CAC can reserve a portion of the total bandwidth for a service class with a higher priority (i.e., UGS or rtPS). The average number of on-going connections of rtPS with the R_CAC scheme is slightly larger than that with the CS_CAC scheme.

Figure 5.16 Average number of rtPS& nrtPS on-going connections vs. connection arrival rate

Nevertheless, the R_CAC dynamically accepts or releases the existing nrtPS connections; consequently, the average number of on-going connections of nrtPS

with the R_CAC scheme will be reduced when the arrival rate of the connection requests increases. This will cause the total average number of on-going connections to become less.

However, the CRCAC has the largest total average number of on-going connections. The total average number of on-going connections of rtPS and nrtPS with the CRCAC scheme increases by 90% compared that with the CS_CAC scheme or the R_CAC scheme, when the arrival rate of the connection requests is 38 connections per second. The reason is that the CR-based system can operate under two spectrums. The system throughput will be doubled. With the novel Cognitive Radio-based ISMM embedded, the CRCAC can enable flexible and efficient spectrum usage by adapting the radio's operating characteristics to the real-time conditions of the environment. It can utilize a large amount of unused spectrum in an intelligent way. The simulation results show that the proposed solution can expand the channel capability of the WiMAX OFDM system up to two times while providing connection-level and packet-level QoS to the heterogeneous traffic.

5.3.2.3 Case Study for the CRCAC Admission Control Scheme

When the traffic load is light, the network has enough resources to accommodate the traffic. The performance and functionality of the CRCAC algorithm is evaluated when the traffic load is heavy, e.g. 98% of the total effective UL channel bandwidth.

The case study has been conducted in two stages. In the first stage, the traffic

load is set to 98% of the total effective UL channel bandwidth of the system operating under its original spectrum e.g. Spectrum 1. The functionality of CRCAC and the functionality of CR-based ISMM are evaluated. In the second stage, it is assumed that the system operates in two spectrums. The functionality of CRCAC is evaluated when the traffic load for each of the two spectrums is 98% of the total effective UL channel bandwidth.

5.3.2.3.1 Experiment 1

Initially, each SS has eight traffic flows requesting transmission service. The detailed MSTR and ML requirements of each flow per SS are listed in Table 5.6.

TABLE 5.6 INITIAL INPUT TRAFFIC FLOWS FOR THE CRCAC SCHEME

<i>Service Flow</i>	<i>MSTR (kbps)</i>	<i>MRTR (kbps)</i>	<i>ML (ms)</i>	<i>CRCAC Decision</i>
SS1_UGS_SFID 1	85.6	85.6	50	Admitted
SS1_UGS_SFID 2	85.6	85.6	50	Admitted
SS1_rtPS_SFID 1	348.3	278.6	40	Admitted
SS1_rtPS_SFID 2	348.3	278.6	40	Admitted
SS1_rtPS_SFID 3	348.3	278.6	50	Admitted
SS1_nrtPS_SFID 1	348.3	278.6	400	Admitted
SS1_nrtPS_SFID 2	348.3	278.6	400	Admitted
SS1_nrtPS_SFID 3	348.3	278.6	400	Admitted

The total UL bandwidth consumed by all 64 flows is 18.09 Mbps. $Effective\_UL\_BW_{remain}$ is 0.36 Mbps. $Min\_ML(UGS)$ is 50 ms, $\overline{D}_{ugs}$ is 5.46 ms,

so, $\overline{D}_{ugs} \leq \text{Min\_ML}(UGS)$. $\text{Min\_ML}(rtPS)$ is 40 ms, $\overline{D}_{rtPS}$ is 12.31 ms, $\overline{D}_{rtPS} \leq \text{Min\_ML}(rtPS)$. Both MSTR and ML requirements of each flow can be met, so the 64 traffic flows from the eight SSs are admitted.

The detailed MSTR and ML requirements of new flows are listed in Table 5.7.

TABLE 5.7 NEW ARRIVAL INPUT TRAFFIC FLOWS FOR THE CRCAC EXPERIMENT 1
& 2

<i>Service Flow</i>	<i>MSTR</i> <i>(kbps)</i>	<i>MRTR</i> <i>(kbps)</i>	<i>ML</i> <i>(ms)</i>	<i>CRCAC</i> <i>Decision</i>
SS1_rtPS_SFID_Exp 1	348.3	278.6	30	Admitted
SS1_nrtPS_SFID_Exp 1	348.3	278.6	400	Admitted
SS1_rtPS_SFID_Exp 2	348.3	278.6	20	Admitted
SS1_nrtPS_SFID_Exp 2	348.3	278.6	400	Rejected

If a new rtPS flow, e.g. SS1_rtPS_SFID_Exp 1, sends its request for transmission service to the BS, $\text{Effective\_UL\_BW}_{\text{remain}}$ will be 0.012 Mbps, $\overline{D}_{ugs} \leq \text{Min\_ML}(UGS)$. $\text{Min\_ML}(rtPS)$ is 40 ms, $\overline{D}_{rtPS}$ is 12.56 ms, $\overline{D}_{rtPS} \leq \text{Min\_ML}(rtPS)$. Therefore, the CRCAC makes a decision to admit SS1_rtPS_SFID_Exp 1 flow.

Now, there are 65 connections in the ACL. If a new nrtPS flow, SS1_nrtPS_SFID_Exp 1, sends its request for transmission service to the BS, $\text{Effective\_UL\_BW}_{\text{remain}}$ will be 0.012 Mbps, which is less than the MSTR requirement of the new flow that is 0.348 Mbps. Then the CRCAC triggers the BSM. It sends Spectrum_Spreading_Info to CR-based ISMM. If a common non-active spectrum detected by the BS and all SSs e.g. Spectrum 2 exists, the

CRCAC will perform the same procedure of CAC mentioned above again in Spectrum 2. If both the QoS requirements of the new flow and the existing connections in Spectrum 2 can be met, the new flow will be admitted into the ACL and will be under the transmission service in Spectrum 2.

5.3.2.3.2 Experiment 2

Based on Experiment 1, there are 65 connections under service in Spectrum 1. Similarly, the same initial input traffic flows are used for Spectrum 2. Following the connection admission control procedure described in Experiment 1, if a new rtPS flow, e.g. SS1_rtPS_SFID_Exp 2, sends its request for transmission service to the BS, the admission control scheme will make a decision to admit it as both the QoS requirements of the new flow and the existing connections can be met. Now there are 65 connections under service in Spectrum 2. The $Effective\_UL\_BW_{remain}$ of Spectrum 2 is 0.012 Mbps. If a new nrtPS flow, e.g. SS1_nrtPS_SFID_Exp 2, sends its request to the BS, $Effective\_UL\_BW_{remain}$ will be less than the MSTR requirement of the new flow, which is 0.348 Mbps. Then the CRCAC triggers BSM. It sends Spectrum_Spreading_Info to CR-based ISMM again. However, the system already operates on two spectrums at this moment. The CRCAC makes a decision to reject the request. Now, the total number of connections under transmission service in the ACL is 130.

The above simulation experiments have shown that the proposed CRCAC can provide the delay guaranteed service to real-time traffic. Comparing with the system operating in one spectrum, the proposal has the ability to expand the channel capacity up to two times.

5.4 Summary

In this chapter, two novel CAC schemes to provide connection-level and packet-level QoS for WiMAX systems have been proposed. Their performances have been analyzed.

Firstly, a cross-layer MAC protocol and QoS support framework to enhance QoS provisioning specified by IEEE 802.16d standard in the WiMAX WirelessMAN-SC PMP system is proposed.

The proposed cross-layer MAC protocol and QoS support framework takes the impact of the air interface on the MAC layer protocol into account to enhance the system efficiency. As system using the Adaptive Discrete Rate AMC with possible $AMC(q)$ based on different channel SINR regions, the concept of the effective UL channel bandwidth is proposed.

Further, the proposed QoS support frame reflects the importance of the combination of scheduling policy with an efficient CAC scheme to provide QoS in WiMAX networks. Associated with the proposed MAC protocol and QoS support framework, a connection admission control scheme termed as ECAC scheme is proposed. The proposed ECAC can provide the connection-level and packet-level QoS (i.e., connection blocking probability, MSTR and ML) to the heterogeneous traffic. The criteria used by the ECAC are $Effective\_UL\_BW_{remain}$, $\overline{D}_{ugs}$ and $\overline{D}_{rtPS}$. The ECAC works cooperatively with the HOS scheduling scheme. It has an ability to catch the PHY condition like AMC and real-time network performance like average packet delay and effective bandwidth by using the AMC information from the PHY layer. The proposed ECAC can improve the

decision accuracy.

The proposed ECAC sets the bandwidth-borrowing algorithm so that a higher service class can borrow bandwidth from a lower service class when the lower service class consumes the bandwidth more than its threshold. With this function, the system can admit more real-time traffic to yield higher revenue.

In order to further expand the system capacity to maximize the system revenue and to enhance QoS provisioning for real-time and non-real-time traffic in WiMAX PMP OFDM system, one novel cross-layer Cognitive Radio-based QoS support framework and Cognitive Radio-based CAC scheme, called CRCAC, is introduced in the second part of this chapter.

The proposed QoS support frame is inspired by Dynamic Spectrum Allocation (DSA) function. The proposed solution is motivated by the technology of Cognitive Radios (CR). The proposed QoS framework is equipped with an intelligent Bandwidth Spreading Module (BSM). The BSM links to CR-based Intelligent Spectrum Management Module at the PHY layer. It can exploit unused spectrum portions. The system can operate under more than one spectrum. As a result, the system capacity can be doubled without a fixed spectrum bandwidth constraint.

Based on the proposed Cognitive Radio-based QoS framework, the queueing model analysis for the WiMAX system using OFDM with TDD duplex technique is carried out. The network performance parameters such as channel's data transmission rate (effective channel bandwidth) and average packet delay of each service class are investigated and calculated.

With the cross-layer approach by applying AMC in wireless channel, the

proposed CRCAC can improve the accuracy of a CAC decision. The proposed CRCAC consists of three modules, namely, Bandwidth Allocation Estimation Module (BAEM), QoS Control Module (QoS_CM) and Bandwidth Spreading Module (BSM). With the novel Cognitive Radio-based ISMM embedded, the CRCAC can enable flexible and efficient spectrum usage by adapting the radio's operating characteristics to the real-time conditions of the environment. It can utilize the large amount of unused spectrum in an intelligent way. The simulation results show the proposed solution can expand channel capability of WiMAX OFDM system up to two times while providing the connection-level and packet-level QoS to the heterogeneous traffic.

In the next chapter, a novel cross-layer CR-based QoS support framework and a joint self-optimizing sub-carrier allocation and symbol-duration scheduling cum mapping scheme (SOS²M) is presented. The proposed solution aims at providing QoS to the heterogeneous traffic in WiMAX WirelessHUMAN^{TM}-OFDMA PMP systems.

Chapter 6 Cognitive Radio-Based Framework and Self-Optimizing Temporal-Spectrum Block Scheduling

Recently, broadband wireless access has become ubiquitous. This makes the already heavily crowded radio spectrum much scarcer [88]. Cognitive radio is a promising technology to alleviate the increasing stress on the fixed radio spectrum. In cognitive radio based opportunistic spectrum access (OSA) networks, the secondary (unlicensed) users can periodically sense and identify available channels which are referred to as White Spaces (WS), the parts of the spectrum not in active use by the primary users. Based on the results of spectrum sensing, the secondary users dynamically tune their transceivers to the identified WSs to access the wireless channel without disturbing the communications of the primary users.

In previous Chapter 3 and 4, the novel three-tier scheduling and HOS opportunistic scheduling scheme have been proposed. These scheduling schemes aim at reducing both the average packet delay and the packet loss rate, as well as improving the system throughput. Their superior performances have been shown.

However, those scheduling decisions are made based solely on bandwidth requirements, system capacity, QoS of connections, queue status of a connection or PHY time-varying channel without considering the mapping constraint. In fact, Downlink (DL) and Uplink (UL) resource scheduling are done in two steps in WiMAX OFDMA systems. In the first step, the number of sub-carriers and the number of slots in units of symbol-duration are allocated to each SS, then those allocated sub-carriers and slots (symbol-durations) are further distributed to each connection at the SS. In the second step, the mapping algorithms can allocate the scheduled UL data bursts from each SS to two-dimensional rectangular “areas” [1] whereby the length of the area is the number of sub-carriers and the width is the number of OFDM symbols. It is assumed that the bins are in two dimensions in this study. This two-dimensional bin is defined as a Temporal-Spectrum Block (TSB) in the thesis. Now, a “tiling” problem has to be solved in which the objective is to fill a given area with TSBs.

The existing proposed scheduling solutions [46-57] and mapping strategies [60-63] tackle the QoS provisioning issue separately. From the system’s point of view, sub-carrier allocation scheme, symbol-duration scheduling scheme and mapping strategy have to be considered collaboratively based on PHY radio condition in order to provide QoS provisioning to the heterogeneous traffic in WiMAX systems.

Autonomic Computing (AC) [95] is a novel way to manage complexity. The goal of autonomic computing is to realize computer and software systems and applications that can manage themselves in accordance with high-level guidance from humans [96].

Motivated by the technology of Cognitive Radios (CR) to tackle the fixed

spectrum-allocation problem and inspired by the Autonomic Computing (AC) system to manage complexity, a novel cross-layer CR-based QoS support framework and a self-optimizing scheduling scheme are proposed. The proposed scheduling scheme is a joint self-optimizing sub-carrier allocation and symbol-duration scheduling cum mapping scheme (SOS²M). The proposed solution is unique. It is an intelligent self-optimizing scheduling scheme, considering the mapping constraint. It has an intelligent spectrum spreading ability to recognize and use one of unused spectrum portions. The solution can improve the system capacity without the constraint of the fixed spectrum bandwidth. It can provide guaranteed QoS to the heterogeneous traffic in WiMAX WirelessHUMAN^{TM}-OFDMA PMP systems.

This chapter is organized as follows. In Section 6.1, research background is presented. In Section 6.2, the proposed cross-layer CR-based QoS support framework is presented firstly. Then, the proposed CR-based self-optimizing joint sub-carrier allocation and symbol-duration scheduling cum mapping scheme is elaborated. In Section 6.3, a queueing model analysis on the proposed solution is carried out. The average packet delay of UGS, rtPS and nrtPS service classes are obtained theoretically. The effective transmission rates for the four service classes are derived. In Section 6.4, the simulation model design and the simulation results are presented. Finally, in Section 6.5, the chapter is concluded with a summary.

6.1 Research Background

6.1.1 Autonomic Computing

There is no easy way to achieve optimality with simple computations to fill a given area with TSBs. Fortunately, an Autonomic Computing (AC) paradigm [95-97] has a mechanism whereby changes in its essential variables can trigger changes in the behavior of the computing system such that the system can be brought back into equilibrium with respect to the environment. This state of stable equilibrium is a necessary condition for the survivability of the organism. In an Autonomic Computing system, the survivability of the system is the ability to protect itself, recover from faults, reconfigure as required by changes in the environment, and always maintain its operations at a near optimal performance.

The vision of Autonomic Computing [98] is to improve the management of complex IT systems by introducing self-management systems. Multiple interacting Autonomic Elements (AEs) form the building blocks of the AC system. The AC system monitors data sensed by sensors, analyzes them, and adjusts its operations according to policies, thus reducing complexity [95].

IBM has defined an Autonomic Computing architecture [99] that focuses on enabling self-management functionality. This has shifted the burden of managing systems from people to technologies based on an enterprise view of appropriate policies. The architecture is built based on an intelligent control loop that monitors, analyzes, plans and executes based on a perception of the current environment [100].

Autonomic applications and systems are composed from autonomic elements,

and are capable of managing their behaviors and their relationships with other systems/ applications in accordance with high-level policies. Autonomic systems/applications exhibit eight defining characteristics [96]:

- **Self-Awareness:** An autonomic application/system “knows itself” and is aware of its state and its behaviors.
- **Self-Configuring:** An autonomic application/system should be able to configure and reconfigure itself under varying and unpredictable conditions.
- **Self-Optimizing:** An autonomic application/system should be able to detect sub-optimal behaviors and optimize itself to improve its execution.
- **Self-Healing:** An autonomic application/system should be able to detect and recover from potential problems and continue to function smoothly.
- **Self-Protecting:** An autonomic application/system should be capable of detecting and protecting its resources from both internal and external attacks and maintaining overall system security and integrity.
- **Context-Aware:** An autonomic application/system should be aware of its execution environment and be able to react to changes in the environment.
- **Open:** An autonomic application/system must function in a heterogeneous world and should be portable across multiple hardware and software architectures. Consequently, it must be built on standard and open protocols and interfaces.
- **Anticipatory:** An autonomic application/system should be able to

anticipate to the fullest extent possible, its needs & behaviors and those of its context, and be able to manage itself proactively.

One of the characteristics of an AC system is self-optimization. Self-optimization is defined as the process where terminals and base stations measured performance metrics are used to auto-tune the network, once it is in an operational state [101].

The existing techniques for achieving self-optimization of an AC system were evaluated [102]. Three approaches are listed as follows.

- The first approach is to make use of control theory to achieve self-optimization. Control theory views the self-managing system as a closed control loop that monitors and manages resources. The usage of linear control theory in achieving self-optimization has been demonstrated in [103]. In Lotus Notes application [103], performance is evaluated by monitoring the maximum number of users on the server. The feedback is provided to an external controller that uses statistical models and previous values to generate an output control. This self-optimization technique is evaluated using control theory analysis technique and empirical assessment.
- The second approach is to make use of a learning based self-optimization technique. It has been used in LEO [104]. LEO implements self-optimization in its query optimizer for DB2. The query optimizer updates the query execution plans based on the feedback generated from early executions. The statistics used for mathematical model of the query execution plan are dynamically updated. It also estimates errors during query execution and re-optimize the query.

- The third approach is to make use of active learning approach to achieve self-optimization [105]. This approach relies on building statistical predictive models based on the previous history of computing utilities.

Self-optimizing architectures have been proposed for QoS provisioning [106] and autonomic resource scheduling [107]. In paper [106], a self-optimizing, self-healing architecture was proposed for QoS provisioning in differentiated services. The architecture employs a model of free Reinforcement Learning (RL) approach. Paper [107] presented an implementation of one part of the autonomic communication framework for intelligent radio resource management. Autonomic resource scheduling based on genetic algorithms is embedded in the autonomic element for scheduling optimization in hybrid TDD based wireless networks.

The principle of learning-based self-optimization is attractive. It will be incorporated into the proposed Temporal-Spectrum Block scheduling scheme to achieve self-optimizing sub-carrier allocation and self-optimizing TSB mapping.

6.1.2 Intelligent Spectrum Management

Recently, the traditional approaches for spectrum management have been reconsidered to the actual use of spectrum. FCC's (Federal Communications Commission) Spectrum Policy Task Force has reported plentiful temporal and geographic variations in the usage of allocated spectrum [88]. One way of increasing spectrum utilization is to reuse the spectrum when their primary users are non-active. With this concept as called Opportunistic Spectrum Sharing, secondary users are allowed to access to the radio frequency (RF) bands without

agreement from the primary users. The unlicensed bands play a key role in this wireless system since the deployment of applications in these bands is unencumbered by regulatory delays.

CR is a promising technology for enabling unlicensed devices to efficiently use White Spaces, which are referred to as the parts of the spectrum not in active use by the primary users [89]. It is an intelligent wireless communication system that is aware of its surrounding environment, and uses the method of understanding-by-building to learn from the environment and adapts its internal states to statistical variations in the incoming RF stimulate by making corresponding changes of its parameters in time [90]. The radio dynamically identifies portions of the spectrum that are not in use by primary users, and configure the radio to operate in the appropriate White Spaces. The CR-based wireless system is able to utilize the large amount of unused spectrum in an intelligent way while not interfering with other incumbent devices in the frequency bands already licensed for specific uses. For detailed overview of the dynamic spectrum access schemes for cognitive radio networks, one may refer to new comprehensive surveys in [91, 92].

6.1.3 WiMAX WirelessHUMAN^{TM} Physic Layer

WiMAX systems support 5 different PHY techniques, namely, WirelessMAN-SC PHY specification, WirelessMAN-SCa PHY specification, WirelessMAN-OFDM PHY specification, WirelessMAN-OFDMA PHY specification and WirelessHUMAN^{TM} PHY specification. WiMAX systems take adaptive modulation policy selecting 1 of 3 different modulation schemes at the PHY layer. On the UL, QPSK is mandatory, while 16-QAM and 64-QAM are optional. The

DL supports QPSK and 16-QAM, while 64-QAM is optional. The AMC scheme uses the Receiver Sensitivity together with the non-overlapping regions of SINR thresholds of receivers to select different burst profiles in order to maximize the network throughput while maintaining the Bit Error Rate (BER) under a preset level e.g. $P_{ber} \leq 10^{-5}$.

WirelessHUMAN^{TM} PHY not only complies with the WirelessMAN-SCa PHY, the WirelessMAN-OFDM PHY and the WirelessMAN-OFDMA PHY, but also further complies with the Dynamical Frequency Selection (DFS) protocols. WirelessHUMAN^{TM} PHY specification is implemented for license-exempt frequencies below 11 GHz. The license-exempt nature introduces new mechanisms such as DFS to detect and avoid interference. Hence, it provides the platform whereby Cognitive Radio can perform.

The chapter is based on WirelessHUMAN^{TM} using OFDMA with TDD duplex technique. OFDMA may be considered a hybrid of time- and frequency-division multiple access. In the time domain, data are transmitted in the form of frames. The minimum time-frequency unit that can be allocated to a user is a slot. A slot consists of a sub-channel over one, two, or three OFDM symbols, depending on the sub-channelization scheme that is used. Figure 6.1 [1] shows OFDMA frame structure with TDD in Partial Usage of Sub-channels (PUSC) scenario.

Each UL burst will further be mapped with a number of TSBs according to a mapping algorithm as shown in Figure 6.2.

Figure 6.1 OFDMA frame structure (TDD, PUSC)

Figure 6.2 UL burst mapping structure

6.2 Cross-Layer Cognitive Radio-Based QoS Support Framework and Self-Optimizing Temporal-Spectrum Block Scheduling

6.2.1 Cross-Layer Cognitive Radio-Based QoS Support Framework

As shown in Figure 2.2, IEEE 802.16d left the details of the scheduling algorithm and the mapping algorithm undefined. The existing proposed scheduling schemes operate under the Fixed Spectrum Allocation (FSA) scheme with a pre-defined fixed spectrum and a fixed channel bandwidth without consideration on the mapping constraint. In order to overcome the shortcomings, a Cognitive Radio based Intelligent Spectrum Management Module needs to be embedded in the current QoS architecture in WiMAX WirelessHUMAN^{TM} – OFDMA systems. A new interface of White_Space_Info needs to be designed to provide the status of White Space to the scheduler. Since the performance of the MAC protocol and the functionality of the scheduler and mapping algorithm are influenced by the time-varying wireless channel in WiMAX systems, a new interface of Symbol_Control_Info needs to be designed to link up the AMC scheme, the scheduling scheme and the mapping scheme from the PHY to the MAC layer.

In order to design an efficient cross-layer QoS support architecture and a corresponding scheduling scheme to provide QoS to the heterogeneous traffic in

WiMAX systems, the following factors have been taken into consideration:

- The existing QoS signaling mechanisms and DL/UL_ Map mechanism in WiMAX should be applied;
- Per-connection scheduling overhead should be minimized;
- The QoS requirement parameters of each connection in the system should be guaranteed;
- The impact of AMC on scheduling and mapping scheme should be taken into account as the wireless time-varying channel in which the channel capacity or bandwidth dynamically fluctuates according to the SINR of PHY;
- Maximizing the system revenue should be taken into account. As many connections should be admitted into the system while the QoS requirements of the existing connections are not violated. The system should be able to exploit the unused spectrum portions and spread spectrums over the White Spaces.

Based on above mentioned design factors, a cross-layer CR-based QoS support framework and the corresponding self-optimizing scheduling cum mapping scheme are proposed as shown in Figure 6.3.

The IEEE 802.16d standard gives Information Element (IE) structure and 16 bits CID. Bandwidth is always requested on a CID basis and bandwidth is allocated on a SS basis. In order to distribute the overhead of scheduling computation to the BS and all SSs, the proposed scheduling solution has a two-stage scheduling structure in which divide-and-conquer methodology is adopted.

Figure 6.3 Proposed cross-layer Cognitive Radio-based QoS support framework

The first stage of scheduling scheme exists at the BS only. It distributes the entire bandwidth of the UL sub-frame to each SS by the Max-Min Fair Sharing (MMFS) [69] scheduling algorithm to achieve fairness among all SSs. The second stage of the scheduling scheme is the proposed self-optimizing Temporal-Spectrum Block scheduling scheme, which is a self-optimizing sub-carrier allocation and symbol-duration allocation cum mapping scheme (SOS²M). It exists at the BS as well as all SSs. It optimally performs three functions: sub-carrier allocation, symbol-duration allocation and TSB mapping.

The detailed rationale of the design of the framework is explained as follows.

1. Considering the impact of PHY air interface on MAC layer protocol, a wireless Channel Condition Estimator (CCE) is designed at the PHY layer at

- the BS as well as SS_u , where $u=1,2,3,\dots,U$. It monitors instantaneous channel condition status, SINR, $\gamma(u, BS)$, when the BS receives signal from SS_u or, SINR, $\gamma(BS, u)$, when SS_u receives message from the BS.
2. In order to maximize the system revenue, a spectrum spreading function should be embedded in the system. In order to enable the spectrum spreading functionality of the system under Dynamic Spectrum Allocation policy, a CR-based Intelligent Spectrum Management Module (ISMM) is designed at the PHY layer at the BS as well as at all SSs. It scans White Spaces and monitors the White Space Status (WSS) of $WSS(BS, x)$ or $WSS(u, x)$ which is a list of the spectrums not used by primary users around the BS or SS_u , where x is non-active spectrum index.
  3. PHY Symbol Rate Controller consists of Forward Error Correction (FEC), Symbol Mapper and AMC Controller. Based on the information of the channel's status provided by CCE, the PHY Symbol Rate Controller of the BS or SS selects a feasible modulation and coding scheme $AMC(u, BS, q)$ for each sub-carrier for UL data transmission at a data transmission rate which is determined by modulation and coding level q according to SINR, $\gamma(u, BS)$. The PHY Symbol Rate Controller at the BS forms $Symbol\_Control\_Info_{BS} \{ \gamma(u, BS), AMC(u, BS, q), WSS(BS, x) \}$. It is forwarded to MAC layer via PHY Service Access Point (SAP) for scheduling and symbol mapping. Similarly, the PHY Symbol Rate Controller at SS_u forms $Symbol\_Control\_Info_{SS} \{ \gamma(BS, u), AMC(BS, u, q), WSS(u, x) \}$. It is embedded in a $Polling\_response(u)$ by SS_u . SS_u sends the $Polling\_response(u)$ to the BS when it is polled.
  4. The first stage of the scheduling scheme at the BS distributes all the sub-

- carriers of the UL sub-frame to all SSs by MMFS [69] scheduling algorithm according to their bandwidth requirements. The result of the scheduling is embedded in DL/UL_Map and broadcasted to all SSs.
5. The second stage of the scheduling scheme at every SS makes a scheduling and mapping decision for UL transmission service based on the proposed cross-layer SOS²M scheme described later in detail in Section 6.2.2.
  6. As UGS, rtPS and nrtPS traffic classes have stringent QoS requirements like MSTR, MRTR and ML, the system uses its own spectrum for UGS, rtPS and nrtPS data transmission. Although CR-based radio needs a certain amount of time for channel sensing and radio-frequency hopping and tuning, BE service class can tolerate delay and bandwidth requirements, and its bandwidth requirement mechanism is contention-based, so White Space spectrum x of CR-based radio can be arranged for BE data transmission. The CR-based ISMM at the BS takes in White Space Status $WSS(u, x)$ information from *Polling_response(u)* message. It compares its $WSS(BS, x)$ with every $WSS(u, x)$ of SS_u where $u=1,2,3,...,U$ to find the same value of x , which means that the common non-active spectrum is detected by BS and all SSs. If the BS or a SS transmits a packet belonging to BE service class, the CR-based ISMM will make spectrum x active. The packet is transmitted in White Space x .

Based on the above description and explanation of the proposal, it is clear that the proposed framework has two outstanding modules, namely, CR-based Intelligent Spectrum Management Module (ISMM) and Self-optimization Scheduling cum Mapping Module. With the cross-layer approach, the proposed framework can intelligently explore unused spectrums and spread over them. Thus, the WiMAX system can operate under more than one spectrum. The

system capacity can be improved significantly. The QoS to real-time traffic can be guaranteed in the networks. The proposal can guarantee QoS requirements for different types of service classes and expand the system capacity. It has the following major contributions as follows.

1. The proposed framework is equipped with a CR-based Intelligent Spectrum Management Module at the PHY layer. It exploits unused spectrum portions. When traffic load increases, the system can operate under more than one spectrum. As a result, the system throughput can be improved significantly.
2. The proposed scheduling algorithm has considered the mapping constraint. It has a unique feature, which is the associative cooperation of scheduling and mapping scheme. The efficiency of mapping scheme can influence the scheduling decision, in other words, it can fine-tune the scheduling decision by the suggested self-optimizing autonomic computing framework. The self-optimizing autonomic computing can reduce the complexity of scheduling and mapping significantly. The QoS requirements of real-time traffic can be satisfied while a higher system efficiency and throughput can be achieved.

6.2.2 Cross-Layer Cognitive Radio-Based Self-Optimizing Scheduling cum Mapping Scheme

Further considering the typical features of WiMAX systems – class-based QoS guarantees and per-flow resource assignment – the proposed scheduling scheme needs to provide differentiated QoS to UGS, rtPS, nrtPS and BE service classes. The cross-layer CR-based self-optimizing sub-carrier allocation and symbol-duration scheduling cum mapping scheme (SOS²M) are proposed. Treated as a

primary user, the transmission service to UGS and rtPS and nrtPS should be managed in a QoS guaranteed manner. Treated as a secondary user, the transmission service to BE should be managed in a throughput-maximizing manner over the White Spaces.

The UML (Unified Modeling Language) collaboration diagram, which is to illustrate the interaction of the proposed scheme, is shown in Figure 6.4.

Figure 6.4 UML collaboration diagram of CR-based self-optimizing scheduling cum mapping scheme

The scheduling cum mapping scheme is elaborated based on the UL channel. The scheme can be easily extended to the DL channel.

6.2.2.1 BS UL Channel Inter-SS Scheduling

As the rtPS and nrtPS can be combined together under a general polling

service (PS) umbrella as suggested in [41], UGS, rtPS and nrtPS bandwidth request of individual SS is handled by the first stage of scheduling scheme at the BS. The BE bandwidth request of individual SS is handled by the second stage of the scheduling scheme at all SSs.

The first stage of scheduling scheme firstly allocates a fixed number of sub-carriers to a SS according to its UGS bandwidth request. Then it further allocates the remaining sub-carriers to the SS according to its rtPS and nrtPS bandwidth request. The detail is as follows.

Consider that the total number of sub-carriers (SC) in UL channel is SC_{total} , the total bandwidth request of UGS class from all SSs for UL transmission service is SC_{UGS} .

The remaining sub-carriers for PS service class in the UL channel is SC_{PS} :

$$SC_{PS} = SC_{total} - SC_{UGS} \quad (6.1)$$

The first stage of scheduling scheme further distributes SC_{PS} in the UL channel to each SS according to its PS bandwidth request by the MMFS [69] scheduling algorithm.

Consider a total number of U SSs in the system, $SS_1, SS_2, \dots, SS_u$, each of them has PS bandwidth request $PS_1, PS_2, \dots, PS_u$ where $PS_1 \leq PS_2 \leq \dots \leq PS_u$. The first stage of scheduling scheme initially allocates SC_{PS}/U of the sub-carriers to the SS with demand PS_1 . The residual sub-carriers $(SC_{PS}/U - PS_1)$ are still available as the total sub-carriers exceed the PS bandwidth request of SS_1 . Distribute the remaining sub-carriers to the remaining $(U-1)$ SSs. Each of them gets $(SC_{PS}/U + (SC_{PS}/U - PS_1)/(U-1))$. This process ends when SC_{PS} is fully distributed to all SSs. If a SS's demand is not satisfied, its bandwidth request will be allocated in the next UL sub-frame. The BS informs all SSs about the result of

the sub-carriers allocation via DL/UL_Map.

6.2.2.2 CR-Based Self-Optimizing Sub-Carrier and Symbol-Duration Scheduling cum Mapping Scheme

When CR-based SOS²M scheme schedules a connection from the ACL, it extracts MSTR or MRTR QoS requirement parameter from its AdmittedQoSParameterSet. Since every SS has been allocated an amount of sub-carriers by the first stage of scheduling scheme, its SOS²M scheme can make a scheduling cum mapping decision according to the following comprehensive algorithm:

CR-based SOS²M algorithm

- 1: get Spectrum_Pool () information, get $WSS(x, t)$;
- 2: get the AMC information from $Symbol\_Control\_Info_{SS}()$;
- 3: get the number of sub-carriers $SC_u (UGS)$ allocated to UGS and $SC_u (PS)$ to PS at SS_u from DL/UL_Map;
- 4: get CID and service class priority from connection classifier;
  - ** Priority of UGS=1; priority of rtPS=2; priority of nrtPS=3;
  - **priority of BE=4;
- 5: get MSTR or MRTR from AdmittedQoSParameterSet (CID);

Algorithm 1:

```

Sub_Carrier_Inter_Connection_Allocation ( $SC_u(CID)$ ,
                                           AdmittedQoSParameterSet ( $CID$ ))
switch ( priority of service class)
case 1: **Based on the specification of IEEE 802.16d standard. UGS service
        **class is allocated with the fixed bandwidth. For UGS service class,

```

**the self-optimizing mapping algorithm is merely considered since
 **the fixed number of sub-carriers is already allocated to its
 **connections;

Call Self_ Optimizing ($SC_u(UGS)$, $CID\_priority$, *scheduling*,
mapping);

Update (CID , C , T , B);

Break;

case 2: Call Self_ Optimizing ($SC_u(PS)$, $CID\_priority$, *scheduling*,
mapping);

$SC_u(nrtPS) = SC_u(PS) - SC_u(rtPS)$;

**Sub-carriers, $SC_u(PS)$, firstly is allocated to $rtPS$, the remainder

**is allocated to $nrtPS$.

Update (CID , C , T , B);

Return ($SC_u(nrtPS)$);

Break;

case 3: Get $SC_u(nrtPS)$;

Call Self_ Optimizing ($SC_u(nrtPS)$, $CID\_priority$, *scheduling*,
mapping);

Update (CID , C , T , B);

Break;

case 4: IF ($WSS(x, t)$ exists)

Call Self_ Optimizing (*sub-carrier*, $CID\_priority$, *scheduling*,
mapping);

Update (CID , C , T , B);

ELSE IF ($WSS(x, t)$ does not exist)

```

Wait ();

** for BE service class, the service channel is allocated on WS.

Call ISMM ();

Sensing_ WhiteSpace ();

Break;

```

Algorithm 2:

```

Self_ Optimizing (sub-carrier, CID_priority, scheduling, mapping)

switch ( priority of service class)

    case 1: Call Self _ Optimizing _Mapping (CID, TSB);
            Update (CID, TSB);
            Break;

    case 2: Call Self_ Optimizing_ Sub_ Carrier_ Allocation(rtPS);
            Call Self_ Optimizing_ Mapping (CID, TSB);
            Update (CID, TSB);
            Break;

    case 3 : Call Self _ Optimizing _ Sub_ Carrier_ Allocation(nrtPS);
            Call Self _ Optimizing _Mapping (CID, TSB);
            Update (CID, TSB);
            Break;

    case 4: Call Self_ Optimizing_ Sub_ Carrier_ Allocation(BE);
            Call Self_ Optimizing_ Mapping (CID, TSB);
            Update (CID, TSB);
            Break;

```

Function 2.1:

```

Self_ Optimizing_ Sub_ Carrier_ Allocation ( $SC_u(rtPS)$ )

```


Get *MinLossRate* and *MaxLossRate* of rtPS connections at SS_u ;

Allocate the number of sub-carriers to each rtPS connection at SS_u by

MMFS algorithm;

IF(*MinLossRate*=*MaxLossRate*)

Maintain the sub-carrier allocation for every connection;

** Assume it is optimized.

ELSE IF (*LossRate* of connection $j = \text{MinLossRate}$)

Decrease one sub-carrier for connection j ;

** Self-optimizing sub-carrier allocation by fine-tuning the

**number of sub-carriers allocated.

Call Self_ Optimizing_ Mapping (*CID*, *TSB*);

**Self-optimizing mapping. Sub-carrier allocation scheme

**cooperates with the mapping scheme.

** Re-map after the number of sub-carriers is fine-tuned.

Update (*CID*, *C*, *T*, *B*) ;

Return ();

ELSE IF (*LossRate* of connection $i = \text{MaxLossRate}$)

Increase one sub-carrier for connection i ;

**Self-optimizing sub-carrier allocation by fine-tuning the number

**of sub-carriers allocated.

Call Self_ Optimizing_ Mapping (*CID*, *TSB*);

**Self-optimizing mapping. Sub-carrier allocation scheme

**cooperates with the mapping scheme.

** Re-map after the number of sub-carriers is fine-tuned.

Update (*CID*, *C*, *T*, *B*) ;

Return ();

END IF

Function 2.2:

Self_Optimizing_Sub_Carrier_Allocation ($SC_u(nrtPS)$)

Get *MinQueueLength* and *MaxQueueLength* of nrtPS connections at SS_u ;

Allocate the number of sub-carriers $SC_u(nrtPS)$ to each nrtPS connection at SS_u by MMFS algorithm;

IF (*MinQueueLength*=*MaxQueueLength*)

Maintain the sub-carrier allocation for every nrtPS connection at SS_u ;

**Assume it is optimized.

ELSE IF (*QueueLength* of connection j = *MinQueueLength*)

Decrease one sub-carrier for connection j ;

**Self-optimizing sub-carrier allocation by fine-tuning the number
**of sub-carriers allocated.

Call Self_Optimizing_Mapping (CID, TSB);

**Self-optimizing mapping. Sub-carrier allocation scheme
**cooperates with the mapping scheme.

** Re-map after the number of sub-carriers is fine-tuned.

Update (CID, C, T, B) ;

Return ();

ELSE IF (*QueueLength* of connection i = *MaxQueueLength*)

Increase one sub-carrier for connection i ;

**Self-optimizing sub-carrier allocation by fine-tuning the number
**of sub-carriers allocated.

```

    Call Self _ Optimizing _Mapping (CID, TSB);

    **Self-optimizing mapping. Sub-carrier allocation scheme

    **cooperates with the mapping scheme.

    ** Re-map after the number of sub-carriers is fine-tuned.

    Update (CID, C, T, B);

    Return ();

END IF

```

Function 2.3:

Self _ Optimizing _ Sub _ Carrier _ Allocation (*BE*)

**The procedure is the same as Function 2.2.

Algorithm 3:

Self _ Optimizing _ Mapping (*CID*, *TSB*)

```

    Allocate (CID, symbol_duration);

    Calculate area  $A = length \times width$ ;

    ** The length is the number of symbol-duration, the width of area is the

    **number of sub-carriers;

    Calculate (CID, TSB _ request);

    Sorted _ TSB _ Allocations = Sort (TSB _ request of each connection) in

    descending order;

    Map the first biggest TSB in the UL sub-frame;

    ** Define the mapping efficiency [60] as follows:

    **Given an allocation for a request TSB, Allocated (TSB);

    **Given a two-dimensional area  $A( length, width)$ , and a set of requests

    **  $B = \{TSB_1, TSB_2, \dots, TSB_n\}$ ;

    **A set of allocations  $M(A, B) = \{Allocated(TSB_1), \dots, Allocated(TSB_i)\}$ ;

```

**The efficiency of the mapping M is defined as the ratio between the
 **effective area allocated by M and the size of area A :

$$E(M) = \frac{\sum_{Allocated(TSB_i) \in M} Area(Allocated(TSB_i))}{length \times width} \quad (6.2)$$

FOR each unmapped element in Sorted_ TSB _ Allocations;

Self_ Optimizing _Map (A, B);

**with high mapping efficiency $E(M)$ and fewer number of
 **preambles. The historical maximum mapping efficiency,
 ** $Max\_E(M)$ is recorded.

IF ($E(M)$ less than $Max\_E(M)$)

Call Self_ Optimizing_ Sub_ Carrier_ Allocation ($Service\_Class\_Priority$);

**Self-optimizing mapping. If $E(M)$ is not optimal, the mapping
 **scheme cooperates with sub-carrier allocation scheme. Fine-tune
 **the number of sub-carriers allocated to service class (rtPS, nrtPS or
 **BE). Then, the system autonomously re-map again to achieve the
 **optimal value of $E(M)$.

END FOR;

6.3 Explanation of SOS²M Scheme

Section 6.2.2 presents the SOS²M algorithm. The rationale for the design of the SOS²M algorithm is explained as follows.

6.3.1 SOS²M Scheduling UGS Connections

Based on the specification of the IEEE 802.16d standard, the UGS service class is allocated with a fixed bandwidth. The Self_Optimizing _ Mapping (*CID*, *TSB*) algorithm is considered for a UGS connection when the fixed number of sub-carriers is already allocated to the connection. So the procedure of self-optimizing sub-carrier allocation, symbol-duration allocation and TSB mapping, which is Self_Optimizing ($SC_u(UGS)$, *CID_priority*, *scheduling*, *mapping*), will be called. Three parameters will be passed to the procedure, which are the total number of sub-carriers allocated to UGS at a SS, the CID priority of the connection and the fixed number of sub-carriers allocated to a UGS connection based on its bandwidth request. The self-optimizing mapping algorithm, Self_Optimizing_Mapping (*CID*, *TSB*), will be performed for the mapping. The result of the mapping will be returned. The scheduling procedure then updates the sub-carrier allocation, symbol-duration allocation and TSB mapping information by executing command $Update(CID, C, T, B)$.

6.3.2 SOS²M Scheduling rtPS Connections

The SOS²M scheme performs two functions to schedule rtPS connections. The first function is self-optimizing inter-connection sub-carrier allocation, which is Self_Optimizing_Sub_Carrier_Allocation ($SC_u(rtPS)$). The second function is self-optimizing mapping scheme, which is Self_Optimizing _ Mapping (*CID*, *TSB*). It can perform symbol-duration allocation and TSB mapping functions.

Step 1 Firstly, sub-carriers, $SC_u(PS)$, is allocated to rtPS service class based on its bandwidth request, then the remaining is allocated to nrtPS as $SC_u(nrtPS)=$

$SC_u(PS)$ - $SC_u(rtPS)$. The SOS²M scheme distributes the number of sub-carriers $SC_u(rtPS)$ to each rtPS connection at SS_u by the MMFS algorithm, which is similar to the procedure described in Section 6.2.2.1.

Step 2 Two parameters, one of which is the minimum packet loss rate of connections at SS_u , $MinLossRate$, and the other is the maximum packet loss rate, $MaxLossRate$, of rtPS connections at SS_u , are statistically recorded and compared with each other. If the condition $MinLossRate = MaxLossRate$ is true, the result of the sub-carrier allocation for every rtPS connection at SS_u by the MMFS described in Step 1 is maintained, no further adjustment of the sub-carrier allocation for rtPS connections at SS_u will be performed since the result is optimal.

The reason why $MinLossRate$ is chosen as an indicator to show whether the sub-carrier allocation scheme is optimal is based on the following rationale.

In a wireless network environment, common channel errors due to multipath fading, shadowing and attenuation may cause bit errors and packet loss, which are quite different from the packet loss caused by congestions in the wired networks. Therefore, in wireless system, packet loss can mistakenly lead to dramatic performance degradation. The information of packet loss rate in wireless network not only reflects the condition of the PHY wireless channel but also serves as an index of effective rate adjustment.

A packet is successfully transmitted to the BS from connection i at SS_u at time t only if:

- it is not dropped from the queue with probability $[1 - P_d(i, u, t)]$ and
- it is correctly transmitted through the wireless channel with probability $[1 - PER(i, u, t)]$.

where $PER(i, u, t)$ is the packet error rate of connection i at SS_u at time t . Hence, the average packet loss rate for connection i at SS_u at time t is:

$$\overline{\varepsilon(i, u, t)} = 1 - [1 - \overline{P_d(i, u, t)}][1 - PER(i, u, t)] \quad (6.3)$$

$\overline{P_d(i, u, t)}$ denotes the average packet dropping (due to queue being full, packet delay exceeding its Maximum Latency or waiting longer than the TCP re-transmission threshold) probability of the queue for connection i at SS_u at time t .

$$\overline{P_d(i, u, t)} = \frac{E_{i,u}\{D\}}{E_{i,u}\{A\}} \quad (6.4)$$

$\overline{P_d(i, u, t)}$ is the ratio of the average number of dropped packets $E_{i,u}\{D\}$ over the average number of arrived packets $E_{i,u}\{A\}$ during the sliding window or scheduling period.

So packet loss rate is used as a key indicator in the proposed self-optimizing scheduling scheme for rtPS connection. If $MinLossRate = MaxLossRate$, it means that the outcome of the sub-carrier allocation scheme has already achieved minimum packet loss rate. Thus, the optimization is achieved.

Step 3 On the other hand, if the condition $MinLossRate = MaxLossRate$ is not true, it means that the outcome of the sub-carrier allocation is not optimal. The number of sub-carriers allocated to rtPS connections will be fine-tuned by the self-optimizing function of the SOS²M scheme. The connection with the minimum packet loss rate will be given fewer sub-carriers while the connection with the maximum packet loss rate will be given more sub-carriers. The step size to increase or to decrease the number of sub-carriers is set to one. Therefore, if the packet loss rate of connection j is equal to $MinLossRate$, the number of sub-carriers allocated to connection j will be decreased by one. The sub-carrier, which

is released from connection j , will be given to connection i if the packet loss rate of connection i is equal to $MaxLossRate$.

The self_optimizing_sub_carrier_allocation algorithm fine-tunes the number of sub-carriers allocated to rtPS connections at SS_u in order to achieve the optimal criterion, which is the minimum packet loss rate of rtPS connections at SS_u .

Step 4 As the sub-carrier allocation scheme cooperates with the TSB mapping scheme, the procedure Self_Optimizing_Mapping (CID , TSB) is called to re-map TSB into a given area every time after the number of sub-carriers is allocated or fine-tuned.

The scheduling procedure then updates sub-carrier allocation, symbol-duration allocation and TSB mapping information by executing command $Update(CID, C, T, B)$ and returns the results.

6.3.3 SOS²M Scheduling nrtPS Connections

SOS²M follows the similar steps described in Section 6.3.2 to perform sub-carrier allocation, symbol-duration allocation and TSB mapping for nrtPS connections. The procedures of Self_Optimizing ($SC_u(nrtPS)$, $CID\_priority$, $scheduling$, $mapping$) is called and executed.

However, the criterion to determine whether the result of the scheduling is optimal is different. Firstly, nrtPS connection does not have a maximum latency requirement in its *AdmittedQoSParameterSet*, in other words, it can tolerate some delay. Secondly, some nrtPS applications, for example, file-transfer, cannot tolerate packet loss. The packet loss rate is not suitable as an indicator for scheduling nrtPS connections. Thus, the minimum queue length is used as an indicator instead of *MinLossRate* in this scenario. The objective of SOS²M for

nrtPS connections is to achieve minimum queue length of nrtPS connections at SS_u .

6.3.4 SOS²M Scheduling BE Connections

As mentioned earlier in Section 6.2, White Space spectrum x is arranged for BE data transmission. If the BS or a SS transmits a packet belonging to BE service class, the CR-based ISMM will make spectrum x active, so the packet will be transmitted in White Space x . The procedure of SOS²M to perform sub-carrier allocation, symbol-duration and mapping scheduling for BE connections is similar to the procedure of SOS²M scheduling nrtPS connections.

6.3.5 Self_Optimizing_Mapping (CID, TSB) Algorithm

Wireless access networks have unique characteristics – time-varying channel conditions and multi-user diversity. Although each connection will be allocated a certain number of sub-carriers in the WiMAX OFDMA system based on the Self_Optimizing_Sub_Carrier_Allocation function, the bandwidth or channel capacity may not fulfill the requirement of a connection as SINR of a sub-carrier at PHY fluctuates. The number of bits transmitted during a symbol-duration or time can be different depending on the modulation and coding scheme applied to each sub-carrier. By using the AMC scheme based on the SINR of a sub-carrier and a pre-set target Bit Error Rate (BER), the maximum number of bits per symbol per Hz that sub-carrier i at SS_u can be transmitted to the BS, denoted by $E(i, u, BS)$, can be expressed as [86]:

$$E(i, u, BS) = \left\lfloor \log_2 \left(1 + \frac{-1.5}{\ln(5P_{ber})} \gamma(i, u, BS) \right) \right\rfloor \quad (6.5)$$

where $\gamma(i, u, BS)$ is the BS received instantaneous SINR of sub-carrier i at source SS_u at a symbol time and P_{ber} is the target BER.

For the wireless fading channel under study, the channel quality can be captured by a parameter, SINR. Since the transmission channel varies from frame to frame, the general Nakagami- m model [84] is adopted to describe SINR statistically. For a Nakagami- m fading channel, the received SINR $\gamma(i, u, BS)$ per frame is thus a random variable with a Gamma probability density function:

$$p[\gamma(i, u, BS)] = \frac{m^m \gamma(i, u, BS)^{m-1}}{\overline{\gamma(i, u, BS)}^m \Gamma(m)} \exp\left(-\frac{m\gamma(i, u, BS)}{\overline{\gamma(i, u, BS)}}\right) \quad (6.6)$$

where $\overline{\gamma(i, u, BS)}$ is the average BS received SINR of subcarrier i at source SS_u .

$\Gamma(m) := \int_0^\infty t^{m-1} e^{-t} dt$ is the Gamma function and m is the Nakagami fading parameter ($m \geq 1/2$).

The AMC scheme adjusts the modulation and coding scheme, hence the data transmission rate on a frame-by-frame basis. The modulation and coding scheme controlled by AMC is independent from SSs and sub-carriers, and depends on the SINR of a sub-carrier and a pre-set target BER. The data transmission rate is an adaptive discrete rate (ADR) [86] with different modulation and coding schemes. There are Q possible modulation and coding schemes available, numbered from $AMC(i, u, BS, 1)$ to $AMC(i, u, BS, Q)$. Let $E(i, u, BS, q)$ be the efficiency of modulation and coding scheme which represents the number of bits per Hz in a PHY symbol when modulation and coding scheme $AMC(i, u, BS, q)$ is chosen in

sub-carrier i from source SS_u to the BS. The SINR can be partitioned into $Q + 1$ consecutive non-overlapping regions with boundary points denoted by $\gamma(i, u, BS, q)$ where $q = 0, 1, \dots, Q$.

Suppose a target BER, P_{ber} , is set. By using equation (6.5), the region boundaries or switching thresholds to achieve the BER can be obtained as follows.

$$\gamma(i, u, BS, q) = \frac{(2^{E(i, u, BS, q)} - 1) \ln(5P_{ber})}{-1.5} \quad (6.7)$$

When $\gamma(i, u, BS, q) \leq \gamma(i, u, BS) \leq \gamma(i, u, BS, q+1)$, the modulation and coding scheme $AMC(i, u, BS, q)$ is applied to sub-carrier i at source SS_u . As a result, the modulation and coding efficiency $E(i, u, BS, q)$ can be achieved. Obviously, in order to avoid possible transmission loss, no packet will be transmitted if the modulation and coding scheme is $AMC(i, u, BS, 0)$.

The probabilities $p(i, u, BS, q)$ to choose a modulation and coding scheme $AMC(i, u, BS, q)$, hence $E(i, u, BS, q)$ for sub-carrier i of source SS_u in the region $[\gamma(i, u, BS, q), \gamma(i, u, BS, q+1))$, can be expressed as:

$$p(i, u, BS, q) = \int_{\gamma(i, u, BS, q)}^{\gamma(i, u, BS, q+1)} p[\gamma(i, u, BS)] d\gamma(i, u, BS) \quad (6.8)$$

In an OFDM symbol time, T_s , the number of bits $r(i, u, BS, q)$ that can be transmitted for sub-carrier i of source SS_u to the BS is generally defined as a function of the modulation and coding scheme efficiency $E(i, u, BS, q)$:

$$r(i, u, BS, q) = T_s \times \Delta f \times E(i, u, BS, q) \quad (6.9)$$

Since $T_s = \frac{1}{\Delta f}$ in an OFDM system, $r(i, u, BS, q)$ is expressed as:

$$r(i, u, BS, q) = E(i, u, BS, q) \quad (6.10)$$

For source SS_u , a set of possible bit transmission rates during an interval of an OFDM symbol in sub-carrier i can be obtained as follows:

$$r(i, u, BS) = \{E(i, u, BS, 0) \times p(i, u, BS, 0), E(i, u, BS, 1) \times p(i, u, BS, 1), \dots, E(i, u, BS, Q) \times p(i, u, BS, Q)\} \quad (6.11)$$

The average bit transmission rate $\overline{r(i, u, BS)}$ that can be transmitted in sub-carrier i from source SS_u to the BS in the interval of an OFDM symbol, T_s , is:

$$\overline{r(i, u, BS)} = \sum_{q=1}^Q E(i, u, BS, q) \times p(i, u, BS, q) \quad (6.12)$$

Let N be the symbol-duration in the unit of T_s , the average number of bits that can be transmitted in the number of sub-carrier, SC , during N symbol-duration is:

$$R_N(sc, u, BS) = N \times \sum_{i=1}^{SC} \sum_{q=1}^Q E(i, u, BS, q) \times p(i, u, BS, q) \quad (6.13)$$

There could be up to 2048 sub-carriers supported by the physical layer of the WiMAX OFDMA system. However, up to 65536 connections can also be simultaneously supported by the 16-bit CID in the WiMAX system [1]. It implies that it is not possible just simply to assign one sub-carrier to one connection. A modified version of TDMA multiplexing technique is introduced to tackle the issue. It is termed as symbol-duration multiplexing by allocating different numbers of symbol-duration to different connections. This is done by function *Allocate (CID, symbol_duration)* in the Self_Optimizing_Mapping (*CID, TSB*) algorithm.

Self_Optimizing_Mapping (*CID, TSB*) algorithm can perform two functions. The first function is symbol-duration allocation, as mentioned above. The objective of the symbol-duration allocation scheme is to determine the value of N in an UL sub-frame to a connection according to its bandwidth requirement.

The MMFS algorithm is applied again to allocate symbol-duration to connections according to its bandwidth request. As a result, an Area A , where its width is the number of sub-carriers and its length is the number of symbol-duration, is allocated to a connection.

The second function of Self_Optimizing _ Mapping (CID , TSB) algorithm is to fill up Area A with TSBs with the optimal criteria, which are high mapping efficiency $E(M)$ and fewer number of preambles in an autonomic approach.

Self_Optimizing _ Mapping (CID , TSB) algorithm is to fill up Area A with TSBs based on the following steps.

Step 1: Sorted_ TSB _ Allocations = Sort (TSB _ request of each connection) in descending order. Then map the first biggest TSB in Area A .

Step 2: For each unmapped element in Sorted_ TSB _ Allocations, Self_Optimizing _ Map (A , B) with high mapping efficiency $E(M)$ and fewer number of preambles will be executed.

Step 3: If the current mapping scheme has a current optimal mapping efficiency $E(M)$ that is less than the historical maximum mapping efficiency, $Max\_E(M)$, Self_Optimizing_ Sub_Carrier_Allocation () is called to fine-tune the number of sub-carriers and the symbol-duration for the connection. Consequently, the length and width of Area A will be changed.

Step 4: After the area of A is adjusted, go to Step 1 to conduct self-optimizing mapping again.

Sub-carrier allocation scheme and symbol-duration allocation scheme cooperates with the mapping scheme to achieve a high mapping efficiency $E(M)$ with fewer number of preambles.

Although the algorithm used in the proposal in the thesis is that the TSBs are

mapped with the biggest TSB first, the proposal is different from the proposals in [59-62]. The proposals in [59-62] are to insert symbols into the fixed width and length of a “rectangle area” which means that the number of sub-carriers and the number of symbol-duration are pre-fixed by a scheduler. The proposed solution is to fill up a “rectangle area” with TSBs where the width and length of the “rectangle” can be dynamically fine-tuned. The self-optimizing function is to fine-tune the width and length of the “rectangle” in order to maximize mapping efficiency $E(M)$.

6.3.6 Discussion on SOS²M Algorithm

Based on the above description and rationale of the design of the framework and corresponding algorithm, the proposal has the following advantages:

1. As shown in Figure 6.4, the scheduling scheme has a two-stage structure.

The overhead of the scheduling and mapping scheme can be shared by the BS and all SSs. Scheduling efficiency can be achieved. The detailed discussion on the computational complexity is explained below.

The first stage of the scheduling scheme is the BS UL inter-SS scheduling scheme. The MMFS scheduling scheme is adopted. The complexity of MMFS algorithm is $O(U)$, where U is the number of SSs.

The second scheduling scheme consists of three functions, namely, Self_Optimizing_Sub_Carrier_Connection_Allocation(), symbol-duration allocation function, $Allocate(CID, symbol\_duration)$ and Self_Optimizing_Mapping(CID, TSB) algorithm.

The SS only monitors the packet loss rates of rtPS connections for the function of Self_Optimizing_Sub_Carrier_Allocation($rtPS$). The

function at the SS only sorts the packet loss rates of rtPS connections in descending order. The scope of the sorting list is limited to the number of rtPS connections at the SS. It selects the parameters of *MaxLossRate* and *MinLossRate* to fine-tune the sub-carrier allocation in order to achieve a lower packet loss rate of rtPS connections optimally. The complexity of the algorithm is $O(N\log N)$, where N is the number of rtPS connections at the SS.

Similarly, the SS only monitors the queue lengths of nrtPS and BE connections for the function of Self_Optimizing_Sub_Carrier_Allocation (*nrtPS*) and Self_Optimizing_Sub_Carrier_Allocation (*BE*), respectively. The complexity of both algorithms is $O(N\log N)$, where N is the number of nrtPS or BE connections at the SS. The Self_Optimizing_Mapping (*CID, TSB*) firstly performs symbol-duration allocation by using the MMFS scheduling scheme. The complexity of MMFS algorithm is $O(N)$, where N is the number of rtPS connections at the SS. Then it sorts the TSBs according to their sizes in descending order. The complexity of the step of the algorithm is $O(K\log K)$ where K is the number of TSBs of UGS, rtPS, nrtPS or BE burst to be filled in the particular Area A. In the next step, the Self_Optimizing_Mapping (*CID, TSB*) will fill TSBs in Area A optimally. The biggest TSB will be allocated in Area A first. The Self_Optimizing_Mapping (*CID, TSB*) will select a suitable TSB from the sorted TSBs to fill up the gap with fewer preambles and less unused space after the biggest TSB is allocated. Then, the second biggest (or the third biggest TSB, if the second biggest TSB is already allocated) will be allocated. The Self_Optimizing_Mapping (*CID, TSB*) will select a

suitable TSB from the sorted TSBs to fill up the gap with fewer preambles and less unused space after the second biggest (or third biggest TSB) is allocated. This procedure iterates until all TSBs are filled in Area A. Therefore, the complexity of this step of the algorithm is $O(K \log K)$ where K is the number of TSBs of UGS, rtPS, nrtPS or BE burst to be filled in the particular Area A.

The Self_Optimizing_Mapping (CID, TSB) has the complexity of $O(N) + O(2K \log K)$.

The historical maximum mapping efficiency, $Max\_E(M)$, is recorded to evaluate the mapping efficiency of each mapping. If the current optimal mapping efficiency $E(M)$ is less than $Max\_E(M)$, Self_Optimizing_Sub_Carrier_Allocation () is called to adjust the length and width of Area A. The number of adjustment is set to one. The computational complexity will increase dramatically if the number of times of adjustment increases. The trade off is to set the number of adjustment to one. Although the system will take a longer time to converge at the optimality, the computational complexity is limited so that the system can handle the sub-carrier allocation, symbol-duration allocation and TSB mapping timely.

As the SOS²M algorithm has a hierarchical structure, the complexity of the SOS²M algorithm scheduling for rtPS, nrtPS or BE service class is equal to $O(N \log N) + O(N) + O(2K \log K)$.

The maximum number of connections supported by the 16-bit CID in the WiMAX system is 65536 [1]. Assume that the number of SS is 8 and the 65536 CID is equally distributed to the four service classes at every SS.

There are 2048 rtPS (or UGS, nrtPS, BE) connections at every SS. The complexity of the SOS²M algorithm is dominated by $O(N\log N)$ which is equal to 6781.

The instruction cycle time varies from instruction to instruction. Assume that the instruction execution time for standard CPU instructions is two clock cycles. Assume that the delay cycle is one clock cycle. A 3G CPU will take $3/3000=1\mu\text{s}$ to complete an instruction. The computing time of 6781 instructions is 6.78 ms. A frame time, e.g. 10 ms, is sufficient for the SOS²M scheme to make a scheduling and mapping decision.

In summary, the overhead of scheduling is divided into two stages and it is distributed and shared by all SSs. A frame time is sufficient for the SOS²M scheme to make a scheduling and mapping decision. The SOS²M scheme can timely schedule the connections for the UL transmission service. The outcome of the scheme will be loaded in DL/UL_Map. The packets in the connections will be transmitted according to the scheduling result in the next frame.

2. SOS²M algorithm schedules traffic of the UGS, rtPS and nrtPS service classes in its own primary spectrum. It schedules traffic of BE service class in White Spaces. This ensures the QoS provisioning to the UGS, rtPS and nrtPS service classes. The resource allocation and mapping scheme can maximize the throughput.
3. According to the network overall performance parameters, like packet loss rate of rtPS service class for real-time traffic, queue length of nrtPS and BE service class, and mapping efficiency $E(M)$, self-optimizing computing can autonomously adapt to the real network environment. The

scheduling and mapping computation complexity can be reduced while higher spectrum efficiency can be achieved.

6.4 Queueing Analysis for WiMAX OFDMA System

To analyze a packet-level QoS in real-time and non-real-time data transmission, the network performance parameters like average packet delay and effective transmission rate need to be investigated. A queueing analytical model can be used off-line to obtain the network performance parameters. In this sub-section, the queueing model analysis on the proposed solution is carried out in the WiMAX OFDMA system. Based on the queueing model analysis, the average packet delay of UGS, rtPS, nrtPS and BE service classes are obtained. The effective transmission rates for UGS, rtPS, nrtPS and BE service classes are derived.

The communication between SSs and BS has two directions – UL and DL. The queueing model analysis is based on the UL channel. Packets that arrive at each SS belong to one of the four priority classes: UGS (priority-1, high priority), rtPS (priority-2), nrtPS (priority-3) or BE (priority-4, lowest priority). At each station, packet arrival is a Poisson arrival process, so that λ_j is the average arrival rate of class j messages, $j = 1, 2, 3, 4$. A packet from rtPS, nrtPS or BE service classes is transmitted in the number of sub-carriers allocated to each one of the four corresponding service class. The time taken to transmit the k^{th} packet belonging to class j is represented as s_k^j .

Based on the proposed framework and the proposed sub-carrier allocation

scheme described above, the UL of the WiMAX OFDMA system can be modeled as four parallel M/G/1 models as shown in Figure 6.5. Each one of the four service classes is served by one M/G/1 queueing model with different number of sub-carriers.

Figure 6.5 WiMAX OFDMA system UL channel queueing model: Four M/G/1 queueing system

Based on equation (6.12), in each OFDM symbol, the capacity $C(SC_u, BS)$ (in the unit of bits/sec) of the number of sub-carriers, SC_u , that can transmit data from source SS_u to the BS can be easily obtained by scaling with the total number of sub-carriers, SC_u , over an OFDM symbol time T_s :

$$C(SC_u, BS) = \left[\sum_{i=1}^{SC_u} \sum_{q=1}^Q E(i, u, BS, q) \times p(i, u, BS, q) \right] / T_s \quad (6.14)$$

Assume that all sub-carriers have the same SINR during an interval of an OFDM symbol from source SS_u , so the same $AMC(u, BS, q)$ is chosen based on the BS received SINR from source SS_u . Hence, $C(SC_u, BS)$ can be expressed as:

$$C(SC_u, BS) = \frac{SC_u}{T_s} \sum_{q=1}^Q E(u, BS, q) \times p(u, BS, q) \quad (6.15)$$

where $p(u, BS, q)$ is the probabilities for choosing $AMC(u, BS, q)$, hence $E(u, BS, q)$, for source SS_u in the BS received SINR region $[\gamma(u, BS, q), \gamma(u, BS, q+1))$.

Based on Section 6.2.2, the BS allocates the number of sub-carriers $SC_u(UGS)$, $SC_u(rtPS)$, $SC_u(nrtPS)$, $SC_u(BE)$ to UGS, rtPS, nrtPS and BE at SS_u respectively.

The average data transmission rate for UGS connections at SS_u is:

$$C(SC_u(UGS), BS) = \frac{SC_u(UGS)}{T_s} \sum_{q=1}^Q E(u, BS, q) \times p(u, BS, q) \quad (6.16)$$

Similarly, the average data transmission rate for rtPS connections at SS_u is:

$$C(SC_u(rtPS), BS) = \frac{SC_u(rtPS)}{T_s} \sum_{q=1}^Q E(u, BS, q) \times p(u, BS, q) \quad (6.17)$$

The average data transmission rate for nrtPS connections at SS_u is:

$$C(SC_u(nrtPS), BS) = \frac{SC_u(nrtPS)}{T_s} \sum_{q=1}^Q E(u, BS, q) \times p(u, BS, q) \quad (6.18)$$

The average data transmission rate for BE connections at SS_u is:

$$C(SC_u(BE), BS) = \frac{SC_u(BE)}{T_s} \sum_{q=1}^Q E(u, BS, q) \times p(u, BS, q) \quad (6.19)$$

The delay of a packet is defined as the total time spent by the packet to get through the system. Denoted by D_k^j , the delay of the k^{th} packet of priority class j , is expressed as:

$$D_k^j = W_k^j + s_k^j + P_k^j + F_L \quad (6.20)$$

where W_k^j denotes the packet waiting-time, s_k^j represents the total time needed to

transmit a priority- j k^{th} packet. F_L denotes the packet frame latency. P_k^j denotes the packet propagation delay. Then the average packet delay of j class $\overline{D_j}$ is given by:

$$\overline{D_j} = E[s_j] + \overline{W_j} + P^j + F_L \quad (6.21)$$

where $E[s_j]$ is the average time needed to transmit a priority- j packet, $\overline{W_j}$ is the average waiting-time of a priority- j packet. $\overline{W_j}$ in the M/G/1 model is given by [94][108]:

$$\overline{W_j} = \frac{\lambda_j E[s_j^2]}{2(1 - \lambda_j E[s_j])} \quad (6.22)$$

6.5 Performance Analysis

A series of simulation experiments have been carried out to evaluate the performance of the proposed cross-layer Cognitive Radio-based QoS support framework and self-optimizing Temporal-Spectrum Block scheduling scheme in the UL channel of the WiMAX WirelessHUMAN^{TM} OFDMA system. The performance of the proposed scheduling scheme is compared with that of the JPCAS proposal in [57]. The average packet delays of service classes obtained by quantitative analysis of the queueing model are further compared and validated with the simulation results. The simulation results show that the proposed scheduling scheme not only effectively supports rtPS traffic with a very low packet delay and loss rate, but also significantly improves the system throughput, as well as greatly enhancing the spectral efficiency.

6.5.1 Simulation Design

Following the signaling mechanism specified by the IEEE 802.16d standard, the WiMAX simulation model has been developed by using ExtendSim [71] software. The scheduling and mapping algorithms have been developed by using C language. The simulation model for the WiMAX system has been designed in PMP topology with one BS and eight SSs. The WiMAX system under study operates as WirelessHUMAN^{TM} OFDMA in TDD duplex mode. The positions of SSs are assumed to be independent and distributed randomly. The traffic parameters are selected from the supported range of values which fully cover the required values for multimedia services defined in Recommendation ITU-R M.1225 [85]. The inter-arrival time of UGS packets is constantly set to 10.1 ms. Packet length is fixed to 200 bytes. The total traffic load of UGS from the eight SSs is fixed to 1.37 Mbps. The rtPS and nrtPS traffic are burst traffic flows. The packet arrival of rtPS and nrtPS flows is a Poisson process with packet inter-arrival time exponentially distributed. The packet arrival rates of rtPS and nrtPS flows are the same and they change according to the different percentage of traffic loads. The mean packet length is set to 500 bytes for both rtPS and nrtPS flows. The ML of rtPS is set to 80 ms. In order to evaluate the function of CR-based scheduling solution to support BE service class, BE traffic load is further introduced. The packet arrival of BE flows is a Poisson process with packet inter-arrival time exponentially distributed. The maximum BE traffic load is set to 20% of the UL channel bandwidth. The mean packet length is set to 500 bytes. The tuning delay [109] of the cognitive radio for BE is set to 10 ms.

The frame size is set to 10 ms, the duration of UL sub-frame is set to 5 ms. The

parameters for OFDMA PHY layer for simulation experiments are outlined in Table 6.1.

TABLE 6.1 SIMULATION PARAMETERS BASED ON IEEE 802.16D
WIRELESSHUMAN^{TM}-OFDMA SYSTEM

<i>Parameters</i>	<i>Value</i>
Bandwidth C per spectrum	20 MHz
Sampling rate F_s	22.86 MHz
Number of sub-carrier N_{FFT}	2048
OFDMA symbol time ($T_s = T_b + T_g$)	100.8 μ s
UL frame sizes (OFDM symbols)	48
Adaptive Modulation Scheme	Squared M -QAM modulations ($M = 4, 16, 64$)
Target BER P_{ber}	$\leq 10^{-5}$
Wireless channel model	Nakagami- m fading channel ($m=1$)

6.5.2 Performance Comparison between Simulation Results and Theoretical Results

According to the configuration of the WiMAX system mentioned above, the numerical results of the average packet delay of UGS, rtPS, nrtPS and BE classes are shown in Table 6.2. It is evaluated and compared with the simulation results.

As shown in Figure 6.6, 6.7, 6.8, 6.9, 6.10, 6.11 and 6.12, the performance of the two scheduling schemes are evaluated with the traffic loads changing from

10% to 90% of the total UL bandwidth accordingly. The performance metrics are average packet delay of the UGS connections, average packet delay, packet loss rate and UL throughput of the rtPS connections, UL throughput of the nrtPS connections, entire UL throughput and average packet delay of the BE connections.

TABLE 6.2 NUMERICAL RESULTS OF AVERAGE PACKET DELAY

<i>Load</i>	<i>UGS</i> (<i>Mbps</i>)	<i>rtPS</i> (<i>Mbps</i>)	<i>nrtPS</i> (<i>Mbps</i>)	<i>Delay</i> (<i>UGS</i>)	<i>Delay</i> (<i>rtPS</i>)	<i>Delay</i> (<i>nrtPS</i>)	<i>Delay</i> (<i>BE</i>)
0.1	1.37	0.24	0.24	12.03	7.22	7.22	17.07
0.2	1.37	1.16	1.16	12.03	7.38	7.38	17.20
0.3	1.37	2.08	2.08	12.03	7.57	7.57	17.36
0.4	1.37	3.01	3.01	12.03	7.79	7.79	17.55
0.5	1.37	3.93	3.93	12.03	8.05	8.05	17.76
0.6	1.37	4.85	4.85	12.03	8.36	8.36	18.02
0.7	1.37	5.77	5.77	12.03	8.75	8.75	18.33
0.8	1.37	6.70	6.70	12.03	9.24	9.24	19.54
0.9	1.37	7.62	7.62	12.03	9.88	9.88	20.28

Figure 6.6 shows the relationship between the average packet delays of the UGS connections with the different traffic loads of the two scheduling schemes and the numerical calculation results of the queueing model. As the UGS connections are allocated with a fixed number of sub-carriers according to their bandwidth requirements based on the QoS framework specified by the IEEE 802.16d standard, it is observed that average packet delays of the UGS connections obtained by the queueing model are similar to the simulation results

of the two scheduling schemes. The average packet delays of the proposed scheduling scheme are lower than the numerical results. The reason is that AMC scheme is taken into account in the proposed scheduling and mapping scheme to improve the spectral efficiency in the system. When the feedback of the AMC information is forwarded to the scheduling and mapping scheme with the cross-layer approach, the SOS²M scheduling and mapping scheme can make an accurate and feasible decision, as a result, the average packet delays of the UGS connections can be reduced.

Figure 6.6 UGS packet delay vs. traffic load

Figure 6.7 shows the relationship between the average packet delays of the rtPS connections of two scheduling schemes and the numerical calculation results of the queueing model. The average packet delays of the rtPS connections in the SOS²M scheme are lower than that of the JPCAS algorithm. The reasons are as

follows. The function of Self_Optimizing_Sub_Carrier_Allocation($SC_u(rtPS)$) in the SOS²M scheduling scheme can optimally allocate a number of sub-carriers to rtPS connections based on the optimizing criteria which is to achieve lower average loss rate of the rtPS connections. The function of Self_Optimizing_Mapping (CID, TSB) can map the rtPS burst into PHY with a high mapping efficiency $E(M)$. This self-optimizing computing can autonomously adapt to the real network environment so as to achieve lower average packet delays of the rtPS service class for real-time traffic.

Figure 6.7 rtPS packet delay vs. traffic load

It is also observed that the average packet delays of the rtPS connections of the queueing model and the SOS²M scheme are very close to each other. Based on Figure 6.6 and 6.7, the simulation model has been validated by the queueing

model.

Figure 6.8 shows the relationship between the packet loss rates of rtPS traffic of the two scheduling schemes with various traffic loads.

Figure 6.8 Loss rate vs. traffic load for rtPS traffic

It is clear that the packet loss rates of the rtPS traffic of the SOS²M scheme are lower than that of the JPCAS scheme when the traffic intensity increases. The reason is that the proposed framework is equipped with a CR-based Intelligent Spectrum Management Module at the PHY layer. It exploits unused spectrum portions. When traffic load increases, the system can operate under more than one spectrum. The function of Self_Optimizing_Sub_Carrier_Allocation($SC_u(rtPS)$) in the SOS²M scheduling scheme can optimally allocate a number of sub-carriers to rtPS connections based on the optimizing criteria which

is to achieve a lower average loss rate of the rtPS connections. Furthermore, the self-optimizing computing can autonomously adapt to the real network environment so as to achieve lower average packet delays of the rtPS service class for real-time traffic. Therefore, it can achieve a much lower packet loss rate for the rtPS traffic connections.

Figure 6.9, 6.10 and 6.11 depict the relationship between the UL throughputs of the rtPS connections, nrtPS connections and entire UL channel with different traffic loads of the two scheduling schemes, respectively.

Figure 6.9 rtPS UL throughput vs. traffic load

Figure 6.10 nrtPS UL throughput vs. traffic load

Figure 6.11 System UL throughput vs. traffic load

It is observed that the SOS²M scheduling scheme can achieve a higher rtPS UL throughput, a higher nrtPS UL throughput and a higher UL system throughput. With the CR-based Intelligent Spectrum Management Module embedded in the system, the proposed SOS²M scheduling scheme can enable the system to operate under more than one spectrum. Furthermore, the proposed solution can perform sub-carrier allocation, symbol-duration scheduling and mapping function cooperatively with a high mapping efficiency $E(M)$, so higher rtPS, nrtPS UL throughput and UL channel throughput are achieved.

Figure 6.12 shows the relationship between the average packet delays of the BE connections of two scheduling schemes.

Figure 6.12 BE packet delay vs. traffic load

The average packet delays of the BE connections of the proposed SOS²M scheduling scheme are lower than that of the JPCAS scheme. The transmission

channel for the BE service class is the same as the UGS, rtPS and nrtPS in the JPCAS scheme. When the traffic load is heavy, for example, when the load is great than 0.7, more transmission opportunities will be given to UGS, rtPS and nrtPS service classes. Fewer transmission opportunities will be given to the BE service class, as priority of UGS > priority of rtPS > priority of nrtPS > priority of BE. The average packet delays of the BE connections of JPCAS are increased sharply. However, treated as a secondary user, the transmission service for BE is managed in a throughput maximizing manner over the White Spaces in the proposed Cognitive Radio-based scheduling framework. When the spectrums around the BS and SSs are in low-usage, there is a White Space spectrum x available for BE data transmission. Furthermore, the function of Self_Optimizing_Sub_Carrier_Allocation($SC_u(BE)$) in the SOS²M scheduling scheme can optimally allocate a number of sub-carriers to BE connections based on the optimization criterion, which is to achieve a lower average queue length of the BE connections. The function of Self_Optimizing_Mapping(CID, TSB) can map the BE burst into PHY with high mapping efficiency $E(M)$. Although a delay of 10 ms is introduced as the tuning delay [109] for the frequency switching in cognitive radio, lower average packet delays of BE are achieved in the SOS²M scheduling scheme when the traffic load is heavy.

6.6 Summary

In this chapter, the novel Cognitive Radio-based QoS support framework and Cognitive Radio-based self-optimizing Temporal-Spectrum Block (TSB) scheduling are proposed. The proposed scheduling scheme is a joint sub-carrier allocation and symbol-duration scheduling cum mapping scheme (SOS²M) in WiMAX WirelessHUMAN^{TM}-OFDMA PMP systems.

The main objective of the Cognitive Radio-based QoS support framework is to expand the capacity of the WiMAX system when the spectrums around the BS and SSs are in low-usage and to enhance the QoS provisioning to UGS and rtPS traffic. The Cognitive Radio-based self-optimizing Temporal-Spectrum Block (TSB) scheduling has two unique functions:

- Firstly, the proposed SOS²M has a unique policy to arrange BE transmission service in White Spaces based on the new Cognitive Radio technique. The UGS, rtPS and nrtPS are treated as primary users so that the QoS provisioning to the three service classes are ensured.
- Secondly, sub-carrier allocation scheme, symbol-duration scheduling scheme and mapping strategy operate collaboratively. The complexity of scheduling and mapping can be easily managed by the self-optimizing function of the Autonomic Computing technique.

With the novel Cognitive Radio-based Intelligent Spectrum Management Module embedded, the proposed framework can enable flexible and efficient spectrum usage by adapting the operating parameters of radio to the real-time conditions of the environment. It can utilize the large amount of unused spectrum

in an intelligent way.

With the cross-layer approach by applying AMC to the sub-carriers in wireless channel, the proposed SOS²M scheme can improve the performance of the WiMAX system significantly. Based on QoS requirement parameters like MSTR and packet loss rate of rtPS service class for real-time traffic as well as network traffic condition like queue length of nrtPS and BE service class, the self-optimizing function can autonomously adapt to the real network environment. The average packet delay of the rtPS service class can be reduced. The average packet loss rate of the rtPS service class can be lowered down.

In order to analyze the performance of the WiMAX OFDMA system under study and to validate the simulation model, M/G/1 queueing modeling of the proposed solution is presented. The average data transmission rates for UGS, rtPS and nrtPS and BE service classes are derived. The average packet delays of each service class are obtained.

Extensive simulation has been carried out. By comparing the simulation results with the JPCAS scheduling scheme and the theoretical calculation results, the proposed solution has shown its superior performance in terms of packet delay, packet loss rate and throughput. The proposed solution can expand the capability of the WiMAX system while providing the QoS provisioning to the heterogeneous traffic.

Chapter 7 Conclusion and Recommendations

7.1 Conclusion

The next generation wireless access is going to be ubiquitous. The wireless networks are expected to support a wide variety of multimedia services such as video on demand, stock market information, digital TV broadcasting, and so on. The characteristics of wireless links make significant challenges to provide QoS to the heterogeneous traffic in terms of end-to-end delay, packet loss rate, bandwidth, or jitter. Those QoS parameters vary significantly over a wide range for different classes of traffic. Efficient and effective QoS provisioning techniques are very important to make such wireless access networks successful.

This dissertation is dedicated to tackle the design and performance analysis of Medium Access Control protocols for QoS provisioning in broadband wireless access networks (WiMAX). The QoS support MAC protocol, architecture and QoS support mechanisms have been studied. Scheduling algorithm, Adaptive Power Control (APC), Connection Admission Control (CAC) and mapping scheme have been proposed to provide QoS provisioning for real-time and non-

real-time applications. The proposed solutions can lower packet delay and loss rate. They can improve the system throughput. They can enhance spectral efficiency and capacity in WiMAX systems. The performances of the proposed solutions also have been evaluated by the simulation results and validated by the queueing analytical model.

The proposed solutions of this work complete the missing QoS modules in WiMAX systems. Their specific contributions have been presented in Chapter 3, 4, 5 and 6. They are summarized as follows.

In Chapter 3, a three-tier with DRR QoS framework and scheduling schemes have been proposed to enhance QoS in WiMAX WirelessMAN-SC PMP systems. Associated with the framework, the proposed DRR and scheduling algorithms with a hierarchical structure can effectively and efficiently provide QoS to the integrated traffic flows, which consist of UGS, PS and BE service flows with or without time constraints. A queueing model has developed to evaluate the performance of the proposal. The simulation and analytical studies show that the proposed solution can not only provide QoS provisioning to the UGS service class, improve the performance of QoS for the PS service class in terms of packet loss rate, packet delay and throughput, but also overcome the starvation problem of the BE service class.

In Chapter 4, firstly, a cross-layer MAC protocol and QoS support framework has been proposed to enhance QoS provisioning in the WiMAX WirelessMAN-SC PMP system. The proposed cross-layer MAC protocol and QoS support framework takes the impact of the air interface on the MAC layer protocol into account to enhance the system efficiency.

Secondly, associated with the proposed MAC protocol and QoS support

framework, the HOS algorithm has been proposed. The proposed HOS has features of channel-awareness, queue-awareness and traffic QoS-awareness. It determines the dynamic Scheduling Priority of each packet by its four key scheduling parameters, namely dynamical priority index, channel specification index, normalized time delay satisfaction index and normalized predictive starvation index. The proposed HOS is a type of opportunistic scheduling algorithm in view of communication over spatiotemporally varying wireless link whereby the multi-user diversity is exploited to maximize bandwidth efficiency and system throughput. With the MAC-PHY cross-layer approach, the proposed HOS algorithm with a two-stage structure can enhance QoS provisioning to rtPS service class with less scheduling complexity and scheduling overhead.

Thirdly, associated with the HOS scheduling algorithm, the adaptive power control scheme has been proposed. The APC scheme can not only perform transmission power optimization to enhance the system power efficiency, but also control the transmission power adaption for rtPS traffic with QoS-oriented consideration.

In Chapter 5, two novel CAC schemes to provide connection-level and packet-level QoS for WiMAX systems have been proposed and their performance have been analyzed.

In the first part of Chapter 5, a cross-layer MAC protocol and QoS support framework has been proposed to enhance QoS provisioning in the WiMAX WirelessMAN-SC PMP system.

The proposed cross-layer MAC protocol and QoS support framework takes the impact of the air interface on the MAC layer protocol into account to enhance the

system efficiency. As system using the Adaptive Discrete Rate AMC with possible $AMC(q)$ based on different channel SINR regions, the concept of the effective UL channel bandwidth has been proposed. It has been proved as stated as Lemma 5.1. The proposed QoS support frame reflects the importance of the combination of scheduling policy with an efficient CAC scheme to guarantee QoS in WiMAX networks.

Associated with the proposed MAC protocol and QoS support framework, the connection admission control scheme termed as ECAC scheme has been proposed. The proposed ECAC can provide the connection-level and packet-level QoS (i.e., connection blocking probability, MSTR and ML) to the heterogeneous traffic. The criteria used by the ECAC are $Effective\_UL\_BW_{remain}$, $\overline{D}_{ugs}$ and $\overline{D}_{rtPS}$, it works cooperatively with a scheduling scheme. It has an ability to catch the PHY condition like AMC and real-time network performance like average packet delay and effective bandwidth by using the AMC information from the PHY layer. The proposed ECAC can improve the accuracy of a CAC decision. A better accuracy results in that:

- higher network utilization ratio can be achieved.
- network overloading can be prevented.

The proposed ECAC sets the bandwidth-borrowing algorithm so that a higher service class can borrow bandwidth from a lower service class when the lower service class consumes the bandwidth more than its threshold. With this function, the system can admit more real-time traffic to yield higher revenue.

In the second part of Chapter 5, one novel cross-layer Cognitive Radio-based QoS support framework and Cognitive Radio-based CAC scheme, termed as

CRCAC, has been proposed in WiMAX PMP OFDM systems.

This QoS support frame is inspired by Dynamic Spectrum Allocation (DSA) function and motivated by the technology of Cognitive Radios (CR). The proposed QoS framework is equipped with intelligent Bandwidth Spreading Module (BSM). The BSM links to CR-based Intelligent Spectrum Management Module at the PHY layer. It can exploit unused spectrum portions. The system can operate under more than one spectrum. As a result, the system capacity can be expanded. The system revenue can be maximized.

Based on the proposed Cognitive Radio-based QoS framework, the queueing model analysis for the WiMAX system using OFDM with TDD duplex technique is carried out. The network performance parameters such as channel's data transmission rate (effective channel bandwidth) and average packet delays of each service class are investigated and calculated.

With the cross-layer approach by applying AMC in wireless channel, the proposed CRCAC can improve the accuracy of a CAC decision. The proposed CRCAC consists of three modules, namely, Bandwidth Allocation Estimation Module (BAEM), QoS Control Module (QoS_CM) and Bandwidth Spreading Module (BSM). It not only checks up the MSTR requirement of a new flow, but also examines the maximum latency requirement of the new flow. The CRCAC can guarantee the whole set of ActiveQoSParameterSet of a connection.

With the novel Cognitive Radio-based ISMM embedded, the CRCAC can enable flexible and efficient spectrum usage by adapting the operating characteristics of radio to the real-time conditions of the environment. It can utilize a large amount of unused spectrum in an intelligent way. The simulation results show that the proposed solution can expand the channel capability of

WiMAX OFDM systems up to two times while providing connection-level and packet-level QoS to the heterogeneous traffic.

Finally, in Chapter 6, a novel Cognitive Radio-based QoS support framework and Cognitive Radio-based self-optimizing Temporal-Spectrum Block (TSB) scheduling (SOS²M) has been proposed in WiMAX WirelessHUMAN^{TM} - OFDMA PMP systems. The SOS²M is a joint sub-carrier allocation and symbol-duration scheduling cum mapping scheme.

The main objective of the Cognitive Radio-based QoS support framework is to expand the WiMAX system capacity when the spectrums around the BS and SSs are in low-usage and to enhance QoS provisioning to real-time traffic. The Cognitive Radio-based self-optimizing Temporal-Spectrum Block (TSB) scheduling has two unique functions:

- Firstly, the proposed SOS²M has a unique policy to arrange transmission service for BE in White Spaces based on the new Cognitive Radio technique. The UGS, rtPS and nrtPS are treated as primary users so that the QoS provisioning to the three service classes are ensured.
- Secondly, sub-carrier allocation scheme, symbol-duration scheduling scheme and mapping strategy operate collaboratively based on the new Autonomic Computing technique. The complexity of scheduling and mapping can be easily managed by its self-optimizing function.

With the cross-layer approach by applying AMC to the carriers in wireless channels, the proposed SOS²M can improve the performance of the WiMAX system significantly.

Based on QoS requirement parameters like MSTR and packet loss rate of rtPS service class for real-time traffic, and network traffic condition like queue length

of nrtPS and BE service classes, self-optimizing function of the SOS²M can adapt to the real network environment. It can reduce the average packet delay of rtPS service class as well as to lower the average packet loss rate of rtPS service class.

In order to analyze the performance of the WiMAX OFDMA system under study and to validate the simulation model, M/G/1 queueing modeling on the proposed solution has been presented. Extensive simulation has been carried out. By comparing the simulation results with the JPCAS scheduling scheme and the theoretical calculation results, the proposed solution has shown its superior performance in terms of packet delay, packet loss rate and throughput. The proposed solution can expand the capability of the WiMAX system while providing QoS to the heterogeneous traffic.

7.2 Recommendations for Future Research

7.2.1 Further Improvement on Current Results

In this thesis, the QoS support MAC protocol, architectures and QoS support mechanisms have been studied. Scheduling algorithm, Adaptive Power Control (APC), Connection Admission Control (CAC) and mapping scheme have been proposed to provide QoS for real-time and non-real-time applications. The proposed solutions can lower the packet delay and loss rate. They can further enhance spectral efficiency and capacity in WiMAX WirelessMAN-SC, WirelessHUMAN^{TM}-OFDM or OFDMA systems.

Surely, there are some rooms, on current solutions and results, which can be further enhanced. The detailed discussion is as follows.

1. As part of the future work, the design of efficient opportunistic scheduling algorithms employed by MAC protocols with a cross-layer approach in WiMAX systems is still a valuable direction. Although many opportunistic scheduling schemes address packet scheduling and sub-carrier allocation issues have been proposed, the consideration of the special characteristic of WiMAX systems is not adequate. As presented in Chapter 4, the two-stage opportunistic scheduling algorithm termed as Holistic Opportunistic Scheduling (HOS) has features of channel-awareness, queue-awareness and traffic QoS-awareness. With the MAC-PHY cross-layer approach, the proposed HOS algorithm with a two-stage structure can enhance QoS provisioning to rtPS service class with less scheduling complexity and scheduling overhead. However, scheduling to achieve fairness among SSs and fairness among traffic flows belonging to the same class has still not figured out clearly. Therefore, more efforts are still needed on how to develop new opportunistic scheduling algorithms to enhance QoS provisioning with high fairness index in WiMAX systems.
2. The proposed adaptive power control scheme presented in Chapter 4 not only can perform transmission power optimization to enhance the system power efficiency but also control the transmission power adaption for rtPS traffic with QoS-oriented consideration. When the transmission power of a SS increases for rtPS traffic, it might introduce interference with surrounding wireless networks. The proposed APC does not consider multiple-input multiple-output (MIMO) and smart antenna factors. Therefore, new APC algorithms that consider interferences and support MIMO and SMART antenna are still valuable to be designed and developed.

3. As presented in Chapter 2, different mechanisms can address different issues regarding QoS support in WiMAX systems. Scheduling algorithm provides mechanisms for bandwidth allocation and multiplexing at the packet level. Admission control policy is dependent on the specific scheduling policy used. For the uplink traffic, the scheduling algorithm has to work in tandem with Call Admission Control (CAC) to satisfy the QoS requirements of traffic. The CRCAC scheme has advanced features of cross-layer approach and intelligent spectrum spreading ability to recognize and use one of unused spectrum portions, hence multiplying the system capacity without the fixed spectrum bandwidth constraint in WiMAX WirelessHUMAN^{TM}-OFDM PMP systems. The effective UL channel bandwidth is derived with an assumption that UL channel bandwidth is equally distributed to all U SSs by inter-SS schedule algorithm like Round Robin scheme. The effective UL channel bandwidth is influenced by inter-SS scheduling algorithm. How to propose a suitable traffic scheduling scheme in tandem with the CRCAC in a cross-layer structure to enhance QoS in WiMAX WirelessHUMAN^{TM}-OFMD systems can be further addressed.
4. Moreover, the Cognitive Radio-based self-optimizing TSB scheduling which is a joint sub-carrier allocation and symbol-duration scheduling cum mapping scheme (SOS²M) in WiMAX PMP systems has been proposed in Chapter 6. The proposed SOS²M is based on Cognitive Radio and self-optimizing Autonomic Computing techniques. How fast the optimizing result can be converged by Autonomic Computing based on its self-optimizing function is still not clear. Further research on self-optimizing scheduling can be focused on.

5. In addition, the SOS²M has such an arrangement to transmit BE service class in White Space. This scheduling policy is suitable for BE as BE service class does not have a stringent QoS requirement like bandwidth or maximum latency constraints, even though regarding Cognitive Radio needs an amount of time to sense the spectrums, to allocate the WS and to tune its radio parameter according to the WS. However, how do the unique features of Cognitive Radio like spectrum hopping and radio parameter tuning affect the QoS provisioning performance if such a WS is allocated to UGS and rtPS traffic? Therefore, it is necessary to put research efforts on Cognitive Radio-based scheduling algorithm.

7.2.2 Potential Research Directions

The potential research directions can be in MAC perspective, PHY perspective or MAC-PHY cross-layer perspective to enhance the performance of WiMAX systems. Those potential research directions are briefly discussed as follows.

1. The first potential research direction is how to enhance IP multimedia services over WiMAX systems. Recently, there is a rapid growth in the number of multimedia applications. Multimedia application contains a variety of media: data, graphics, images, audio and video. The transmission of the multimedia application is a kind of real-time and stream-oriented communication. The Quality of Service required of a stream communication includes guaranteed bandwidth (throughput), delay and delay variation (jitter). Different multimedia applications require various classes of transmission service including the transmission of data, audio, and various types of video and image on wireless network. The IP Multimedia Subsystem (IMS)

[110][111] framework represents an integrated traffic of data, voice & video over the network. With the typical features of WiMAX systems: class-based QoS guarantees and per-flow resource assignment as presented in Chapter 2, WiMAX systems hold great promise as the potential solution to IP multimedia for corporations. However, the Quality of Service of different kinds of media varies. On one hand, hard real-time traffic like voice and video requires stringent time-delay and delay variance, but tolerates a small percentage of packet loss. On the other hand, soft or non real-time traffic like images, graphics, text, and data requires no packet loss, but tolerates time delay. Therefore, to enhance IP multimedia services over WiMAX systems to support different types of multimedia applications is an interesting research topic.

2. The second potential research direction is to enhance QoS provisioning to real-time traffic like multimedia streaming and to boost the system performance by spatial multiplexing (SM) [112] via multiple-input multiple-output (MIMO) [113] and smart antenna techniques. By encoding the data over both the temporal and spatial domains, space-time block codes (STBC) provide spatial diversity and robustness against fading. However, since redundant information is transmitted in each of the antennas, this diversity comes at the expense of peak data rate. Spatial multiplexing (SM), also known as MIMO, is a powerful technique for multiple-antenna systems that, in principle, increases the data rate in proportion to the number of transmit antennas since each transmit antenna carries a unique stream of data symbols. Hence, if the number of transmit antennas is M and the data rate per stream is R , it is straightforward to see that the transmit data rate is $M \times R$ under Spatial

Multiplexing [114]. Popular receiver structures for SM include linear receivers, such as Zero-Forcing (ZF) or Minimum Mean Square Error (MMSE), nonlinear receivers such as the optimum Maximum Likelihood Detector (MLD), and spatial interference canceling receivers such as BLAST. One restriction for all these receivers is that the number of receive antennas should not be lesser than the number of transmitted data streams, or the MIMO channel will be ill conditioned and the data cannot be decoded correctly. Linear receivers are easy to implement in a practical system due to their low computational complexity, but are subject to severe noise enhancement in an interference-limited cellular system. To improve the performance of the IEEE 802.16d standards in terms of increasing data rate and throughput, the emerging technology of Spatial Multiplexing can be one of the interesting research trends.

3. The third potential research direction is to design a Cognitive Radio-based WiMAX system to coexist with Wireless Regional Area Networks (WRAN) [115] to support digital TV broadcasting.

IEEE 802.22 defines a new air-interface standard for Wireless Regional Area Networks mainly aimed at extending broadband access [116] in low population density rural areas by using the VHF/UHF TV broadcast bands for their better signal propagation characteristics. The Cognitive Radio capabilities of IEEE 802.22 are the ability of RF sensing to detect the presence of broadcast incumbents on co-channel and adjacent channels, necessary to avoid interference. They are detailed as:

- Four blind sensing schemes proposed for generic signal detection
- Ten sensing techniques for ATSC DTV signal detection

- One sensing technique for DVB-T DTV signal detection
- One sensing technique for NTSC TV signal detection
- Two sensing techniques for wireless microphone detection
- One wireless beacon standard (802.22.1) and one detection technique to acquire the wireless microphone beacon to better protect legitimate wireless microphone operation.

IEEE 802.16h, an amendment of IEEE 802.16, has mechanisms covering both the coexistence between IEEE 802.16 systems, and between IEEE 802.16 and non-IEEE 802.16 systems. For this reason, the IEEE 802.16h group has proposed coordinated and uncoordinated coexistence mechanisms to meet the different regulatory rules and deployment scenarios.

IEEE 802.16h (fixed/mobile) can provide WiMAX systems with the capability to implement collaborative networks by using a co-existence control channel and share co-channel spectrum. It is able to identify and avoid interference generated by other co-channel systems (non-WiMAX systems). It is also able to identify and control interference generated by similar WiMAX systems.

Digital television (DTV) supports many different picture formats defined by the combination of size, aspect ratio (height to width ratio) and interlacing. The range of formats can be coarsely divided into two categories: high definition television (HDTV) and standard definition television (SDTV). DVB technology has become an integral part of global broadcasting, setting the standard for satellite, cable, terrestrial and IP-based services. With the advancement of Wireless Region Area Network (WRAN) driven by IEEE 802.22 standard, WiMAX networks can coexist with WRAN to support

digital TV broadcasting service. The end-to-end QoS support mechanisms by cognitive spectrum sensing, intelligent opportunistic spectrum and sub-carrier allocation or further Temporal/Spatial Partitioning of a TDD co-channel is another new emerging trend.

4. The fourth potential research direction is how to evolve the MAC protocol of WiMAX system to position itself in Open Wireless Architecture (OWA) platform [117]. The future wireless service provision will be characterized by global mobile access, high Quality of Services, and easy and simple access to multimedia services for voice, data, message, video, world-wide web, GPS, etc. This vision from the user perspective can be implemented by integration of these different evolving and emerging wireless access technologies in a common flexible and expandable platform to provide a multiplicity of possibilities for current and future services and applications to users in a single terminal. Systems of fourth generation mobile will mainly be characterized by a horizontal communication model, where different access technologies such as cellular, cordless, WLAN systems, short-range wireless connectivity, broadband wireless access systems and wired systems will be combined on a common platform. They can complement each other in an optimum way for different service requirements and radio environments. This technology is called “Open Wireless Architecture” invented by Delson (R&D) Group including USCWC, 4G R&D Center, Sieneon Technologies, Obama Lab and CRCWC. OWA defines the open and extensible interfaces in wireless networks and systems, including base-band signal processing parts, RF parts, networking parts, and OS and application parts, so that the system can support different industrial standards and integrate the various wireless

networks into an open broadband platform. For comparison, Software Defined Radio (SDR) [118] is only a radio in which the operating parameters including inter alia frequency range, modulation type, and/or output power limitations can be set or altered by software. Therefore, SDR is just one of the implemental modules of the OWA system. OWA will eventually become the global industry leading solution to integrate various wireless air-interfaces into one wireless open terminal where the same end equipment can flexibly work in the wireless access domains as well as in the mobile cellular networks [119]. This single equipment with a single number and multiple air-interfaces (powered by OWA) will definitely dominate the wireless communication industries. To fuse the WiMAX system into OWA platform by strengthening its MAC protocol, QoS support architecture and PHY air-interface will be an interesting research direction.

Author's Publications

1. Papers (Accepted):

Book Chapters:

- [1] Maode Ma and Jinchang Lu, "QoS Provision Architecture and Mechanisms in WiMAX Networks," *Current Technology Developments of WiMAX Systems*, edited by Dr. Maode Ma, Publisher, Springer, 2008.
- [2] Maode Ma and Jinchang Lu, "QoS Support Mechanisms in WiMAX," *Wireless Network Traffic and Quality of Service Support: Trends and Standards*, edited by Dr. Thomas Lagkas, Publisher, IGI Global, 2009, accepted for publication.

Conference Papers:

- [3] Maode Ma, Jinchang Lu, Sanjay Kumar Bose and Boon Chong Ng. "A Three-Tier Framework and scheduling to Support QoS Service in WiMAX," *Proceedings of IEEE 6th International Conference Information, Communications & Signal*, Dec. 2007, pp.1-5.
- [4] Jinchang Lu and Maode Ma, "A Cross-Layer MAC Protocol and An Opportunistic Scheduling with APC for QoS Support in WiMAX," *Proceedings of IEEE 7th International Conference Information, Communications & Signal*, 2009, accepted for publication.
- [5] Jinchang Lu and Maode Ma, "Self-Optimizing Temporal-Spectrum Block Scheduling for QoS in WiMAX Networks," *IEEE Globecom 2010*, accepted for publication.

Journal Papers:

- [6] Maode Ma, Jinchang Lu and Cheng Peng Fu “A Hierarchical Scheduling Framework for QoS Service in WiMAX PMP Systems,” *IET Communication*, 2009, accepted for publication.
- [7] Jinchang Lu and Maode Ma, “A Cross-Layer Elastic CAC and Holistic Opportunistic Scheduling for QoS Support in WiMAX,” *Computer Networks*, Vol. 54 (2010), pp. 1155-1168, DOI information: 10.1016/j.comnet.2009.10.010.
- [8] Jinchang Lu and Maode Ma, “A Cross-Layer MAC Protocol and Opportunistic Scheduling With APC for QoS in WiMAX,” *The International Journal of Wireless Personal Communications*, accepted for publication.

2. Papers (Under Review)

Journal Papers:

- [9] Jinchang Lu and Maode Ma, “Cross-Layer QoS Support Framework and Holistic Opportunistic Scheduling for QoS Provisioning in WiMAX systems,” *Journal of Network and Computer Applications*, under review.
- [10] Jinchang Lu, Maode Ma, “A Cross-Layer Cognitive Radio-Based Framework and CAC Scheme for QoS Support in WiMAX Systems,” *IEEE Systems Journal*, under review.
- [11] Jinchang Lu and Maode Ma, “Cross-Layer Framework and Self-Optimizing Temporal-Spectrum Block Scheduling for QoS Provisioning in WiMAX,” *IEEE Transactions on Parallel and Distributed Systems*, under review.
- [12] Jinchang Lu and Maode Ma, “Cognitive Radio-Based Framework and Self-

Optimizing Temporal-Spectrum Block Scheduling for QoS Provisioning in WiMAX,” *Computer Networks*, under review.

Bibliography

- [1] “IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems,” IEEE Std 802.16-2004 (Revision of IEEE Std 802.16-2001), pp. 1-857, 2004.
- [2] J. P. Conti, “The long road to WiMAX [wireless MAN standard],” *IEE Review*, Oct. 2005, vol. 51, no. 10, pp. 38- 42.
- [3] A. R. Shankar and R. Hegde, “WiMAX on the road to future,” *IET Int. Conf. Wireless, Mobile and Multimedia Networks*, Jan. 2008, no. 11-12, pp. 275-278.
- [4] S. J. Golestani, “A Self-clocked Fair Queueing Scheme for Broadband Applications,” *Proc. IEEE 13th Conf. Networking for Global Communications, INFOCOM '94*, 1994, vol. 2, pp. 636-646.
- [5] M. Katevenis, S. Sidiropoulos, and C. Courcoubetis, “Weighted Round-robin Cell Multiplexing in a General-purpose ATM switch Chip,” *IEEE Journal on Selected Areas in Communications*, Oct. 1991, vol. 9, no. 8, pp. 1265-1279.
- [6] H. Chetto and M. Chetto, “Some Results of the Earliest Deadline Scheduling Algorithm,” *IEEE Transactions on Software Engineering*, vol. 15, no. 10, Oct. 1989, pp. 1261-1269.
- [7] C. Chang, S. Shao, M. R. Perati, and J. Wu, “Performance Study of Various Packet Scheduling Algorithms for Variable-packet-length Feedback Type WDM Optical Packet Switches,” *Workshop on High Performance Switching and Routing*, Jun. 2006, pp. 5
- [8] H. S. Alavi, M. Mojdeh, N. Yazdani, “A Quality of Service Architecture

- for IEEE 802.16 Standards,” *Asia-Pacific Conference on Communications*, Oct. 2005, pp. 249- 253.
- [9] Yi-Ting Mai, Chun-Chuan Yang and Yu-Hsuan Lin, “Cross-Layer QoS Framework in the IEEE 802.16 Network,” *The 9th International Conference on Advanced Communication Technology*, Feb. 2007, vol. 3, pp. 2090-2095.
- [10] T. Kwon, H. Lee, S. Choi, J. Kim, D. Cho, S. Cho, S. Yun, W. Park and K. Kim, “Design and implementation of a simulator based on a cross-layer protocol between MAC and PHY layers in a WiBro compatible IEEE 802.16e OFDMA system,” *IEEE Communication Magazine*, 2005, vol. 43, no. 12, pp. 136-146.
- [11] Kamal GAKHAR, Mounir ACHIR and Annie GRAVEY, “Dynamic Resource Reservation in IEEE 802.16 Broadband Wireless Networks,” *IWQoS 2006*, Jun. 2006, pp. 140-148.
- [12] HF. Hou, P. H. Ho, X. Shen, “WLC17-1: Performance Analysis of a Reservation Based Connection Admission Scheme in 802.16 Networks,” *IEEE GLOBECOM'06*, Nov. 2006, pp. 1-5.
- [13] H. Wang, W. Li, D. P. Agrawal, “Dynamic admission control and QoS for 802.16 wireless MAN,” *Wireless Telecommunications Symposium*, Apr. 2005, pp. 60 - 66.
- [14] Chi-Hong Jiang and Tzu-Chieh Tsai, “Token bucket based CAC and packet scheduling for IEEE 802.16 broadband wireless access networks,” *Consumer Communications and Networking Conference, CCNC 2006*, 2006, vol. 1, pp. 183-187.
- [15] H. Wang, B. He and D. P. Agrawal, “Admission control and bandwidth allocation above packet level for IEEE 802.16 wireless MAN,” *Parallel*

and Distributed Systems, ICPADS 2006, Jul. 2006, pp. 6.

- [16] Bo Rong, Yi Qian, Kejie Lu, Hsiao-Hwa Chen and M. Guizani, "Call Admission Control Optimization in WiMAX Networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, Jul. 2008, pp. 2509-2522.
- [17] D. Niyato and E. Hossain, "A Queueing-Theoretic and Optimization-Based Model for Radio Resource Management in IEEE 802.16 Broadband Wireless Networks," *IEEE Transactions on Computers*, vol. 55, no. 11, Nov. 2006, pp. 1473 - 1488.
- [18] Yin Ge, Geng-Sheng Kuo, "An Efficient Admission Control Scheme for Adaptive Multimedia Services in IEEE 802.16e Networks," *Proceedings of IEEE 64th Vehicular Technology Conference, VTC 2006-Fall*, Sept. 2006, pp. 1-5.
- [19] D. Niyato and E. Hossain, "Radio Resource Management Games in Wireless Networks: An Approach to Bandwidth Allocation and Admission Control for Polling Service in IEEE 802.16," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, Feb. 2007, pp. 27-35.
- [20] Liping Wang, Fuqiang Liu, Yusheng Ji, Nararat Ruangchaijaturon, "Admission Control for Non-preprovisioned Service Flow in Wireless Metropolitan Area Networks," *Proceedings of IEEE Fourth European Conference on Universal Multiservice Networks, ECUMN '07*, Feb. 2007, pp. 243-249.
- [21] K. Wongthavarawat and A. Ganz, "Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems," *International Journal of Communication Systems*, vol. 16, no. 1, Feb. 2003, pp. 81-96.

- [22] E. Kwon, K. Jung, J. Lee and S. Ryu, "A Performance Model for Admission Control in IEEE 802.16," *Proceedings of WWIC 2005*, vol. 3510, May 2005, pp. 159-168.
- [23] Xinbing Wang, Do Young Eun and Wenye Wang, "A Dynamic TCP-Aware Call Admission Control Scheme for Generic Next Generation Packet-Switched Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 9, Sept. 2007, pp. 3344-3352.
- [24] N. Zorba and A. I. Perez-Neira, "CAC for Multibeam Opportunistic Schemes in Heterogeneous WiMax Systems under QoS Constraints," *Proceedings of Global Telecommunications Conference, GLOBECOM 07*, Nov. 2007, pp. 4296-4300.
- [25] J. Chen, W. Jiao and Q. Guo, "An integrated QoS control architecture for IEEE 802.16 broadband wireless access systems," in *Proc. Global Telecommunication Conf., (Globecom) 2005*, 2005, vol. 6, pp. 3330-3335.
- [26] W. T. Chen, S. B. Shih, and J. L. Chiang, "A Two-Stage Packet Classification Algorithm," *Advanced Information Networking and Applications, 2003. AINA 2003*, Mar. 2003, pp. 762-767.
- [27] J. B. Seo, H. W. Lee and C. H. Cho, "Performance of IEEE 802.16 Random Access Protocol - Transient Queueing Analysis," *IEEE Global Telecommunications Conference, 2006, GLOBECOM'06*, Nov. 2006, pp. 1-6.
- [28] J. B. Seo, S. J. Kim, H. W. Lee and C. H. Cho, "An Efficient Capacity Allocation Scheme of Periodic Polling Services for a Multimedia Traffic in an IEEE 802.16 System," *Mobile Adhoc and Sensor Systems (MASS) 2006*, Oct. 2006, pp. 11-20.

- [29] Luís Felipe, M. de Moraes and Paulo Ditarso Maciel Jr, “An Alternative QoS Architecture for the IEEE 802.16 Standard,” www.ravel.ufrj.br/arquivosPublicacoes/conext06.pdf.
- [30] Y. Chen and L. Li, “A Random Early Expiration Detection Based Buffer Management Algorithm for Real-time Traffic over Wireless Networks,” *Computer and Information Technology, 2005, CIT 2005*, Sept. 2005, pp. 507-511.
- [31] J. Lakkakorpi, A. Sayenko, J. Karhula, O. Alanen and J. Moilanen, “Active Queue Management for Reducing Downlink Delays in WiMAX,” *IEEE 66th Vehicular Technology Conference, 2007, VTC-2007*, pp. 326-330.
- [32] Marc Engels, Peter Coenen (IMEC), Christian Hoymann, “Congestion control mechanism for interworking between WLAN and WMAN,” *IEEE C802.16d-03/83. Congestion Control*, http://www.wirelessman.org/tgd/contrib/C80216d-03_83.pdf.
- [33] H. Hsiao, A. Chindapol, J. A. Ritcey, Y. C. Chen; J. N. Hwang, “A new multimedia packet loss classification algorithm for congestion control over wired/wireless channels,” *Acoustics, Speech, and Signal Processing, 2005. ICASSP '05*, Mar. 2005, vol. 2, pp. ii/1105 - ii/1108.
- [34] C. W. Huang, J. N. Hwang and D.C.-W. Chang, “Congestion and error control for layered scalable video multicast over WiMAX,” *IEEE Mobile WiMAX Symposium*, 2007, Mar. 2007, pp. 114-119.
- [35] T. Wiegand, “Joint Final Committee Draft,” *Doc. JVT-E146d37ncm, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6)*, Nov. 2002.
- [36] V. Srivastava, M. Motani, “Cross-layer design: a survey and the road

- ahead,” *IEEE Communications Magazine*, vol. 43, no. 12, pp. 112-119, 2005.
- [37] S. A. Xergias, N. Passas, L. Merakos, “Flexible resource allocation in IEEE 802.16 wireless metropolitan area networks,” *Local and Metropolitan Area Networks, 2005, LANMAN 2005*, Sept. 2005, pp. 6.
- [38] W. K. Wong, H. Tang, S. Guo and Leung, “Scheduling algorithm in a point-to-multipoint broadband wireless access network,” *Vehicular Technology Conference, 2003. VTC 2003*, vol. 3, pp. 1593-1597.
- [39] S. Ryu, B. Ryu; H. Seo and M. Shi, “Urgency and efficiency based wireless downlink packet scheduling algorithm in OFDMA system,” *IEEE 61st Vehicular Technology Conference, 2005*, May 2005, vol. 3, pp. 1456-1462.
- [40] A. Shejwal and A. Parhar, “Service Criticality Based Scheduling for IEEE 802.16 WirelessMAN,” *The 2nd International Conference on Wireless Broadband and Ultra Wideband Communications (AusWireless 2007)*, Aug. 2007, pp. 12-12.
- [41] Maode Ma and Boon Chong Ng, “Supporting Differentiated Services in Wireless Access Networks,” *ICCS 2006*, pp. 1-5.
- [42] Jianfeng Chen, Wenhua Jiao, Hongxi Wang, “A service flow management strategy for IEEE 802.16 broadband wireless access systems in TDD mode,” *IEEE International Conference, ICC 2005*, vol. 5, pp. 3422- 3426.
- [43] H. Safa, H. Artail, M. Karam, R. Soudan and S. Khayat, “New Scheduling Architecture for IEEE 802.16 Wireless Metropolitan Area Network,” *IEEE/ACS International Conference on Computer Systems and Applications, 2007. AICCSA07*, May 2007, pp. 203-210.
- [44] Naian Liu, Xiaohui Li, Changxing Pei and Bo Yang, “Delay Character of a

- Novel Architecture for IEEE 802.16 Systems,” *Proceedings of the Sixth International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT’05)*, Dec. 2005, pp. 293-296.
- [45] Xin Liu, E. K. P. Chong and N. B. Shroff, “Optimal Opportunistic Scheduling in Wireless Networks,” *Proceedings of IEEE 58th Vehicular Technology Conference*, 2003, vol. 3, pp. 1417-1421.
- [46] Vegard Hassel, “Tutorial: Opportunistic Multiuser Scheduling for Wireless Networks,” Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU), Jun. 2006, www.unik.no/personer/porten/MoPSAR/MOPSARTutorial_Hassel.pdf.
- [47] T. Bonald, “Flow-level performance analysis of some opportunistic scheduling algorithms,” *European Transactions on Telecommunications*, vol. 16, 2005, pp. 65- 75.
- [48] A. Jalali, R. Padovani and R. Pankaj, “Data throughput of CDMAHDR: a high efficiency-high data rate personal communication wireless system,” *Proceedings of IEEE VTC 2000*, May 2000, pp. 1854-1858.
- [49] Jinri Huang, Zhisheng Niu, “A Cross-Layer Proportional Fair Scheduling Algorithm with Packet Length Constraint in Multiuser OFDM Networks,” *Proceedings of IEEE International Conference on Global Telecommunications Conference, GLOBECOM ’07*, Nov. 2007, pp. 3489-3493.
- [50] H. J. Bang, T. Ekman and D. Gesbert, “Channel predictive proportional fair scheduling,” *IEEE Transactions on Wireless Communications*, vol. 7, no. 2, Feb. 2008, pp. 482-487.

- [51] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar and P. Whiting, "CDMA data QoS scheduling on the forward link with variable channel conditions," *Bell Labs Technical report preprint*, Apr. 2000.
- [52] M. Hu, J. Zhang and J. Sadowsky, "Traffic aided opportunistic scheduling for wireless networks: algorithms and performance bounds," *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 46, no. 4, 2004, pp. 505-518.
- [53] S. Shakkottai and A. Stolyar, "Scheduling algorithms for a mixture of real-time and non-real-time data in HDR," *17th Int. Teletraffic Congress (ITC-17)*, Sept. 2001.
- [54] A. Lera, A. Molinaro and S. pizzi, "Channel-Aware Scheduling for QoS and Fairness Provisioning in IEEE 802.16/WiMAX Broadband Wireless Access Systems," *IEEE Network*, Sept. 2007, vol. 21, no. 5, pp. 34-41.
- [55] J. Sang, D. Jeong and W. Jeon, "Cross-layer Design of Packet Scheduling and Resource Allocation in OFDMA Wireless Multimedia Networks," in *Proc. 63rd Vehicular Technology Conf. (VTC) 2006*, pp. 309-313.
- [56] Q. Liu, X. Wang and G. B. Giannakis, "A Cross-Layer Scheduling Algorithm With QoS Support in Wireless Networks," *IEEE Transactions on Vehicular Technology*, May 2006, vol. 55, no. 3, pp. 839-847.
- [57] Lihua Wan, Wenchao Ma and Zihua Guo, "A Cross-layer Packet Scheduling and Subchannel Allocation Scheme in 802.16e OFDMA System," *Wireless Communications and Networking Conference, 2007, WCNC 2007*, Mar. 2007, pp. 1865 -1870.
- [58] Bo Rong, Yi Qian and Hsiao-Hwa Chen, "Adaptive power allocation and call admission control in multiservice WiMAX access networks," *IEEE*

- Wireless Communications*, Feb. 2007, vol. 14, no. 1, pp. 14-19.
- [59] T. Ohseki, M. Morita and T. Inoue, "Burst Construction and Packet Mapping Scheme for OFDMA Downlinks in IEEE 802.16 Systems," *GLOBECOM 2007*, pp. 4307-4311.
  - [60] Y. Ben-Shimol, I. Kitroser and Y. Dinitz, "Two-dimensional mapping for wireless OFDMA systems," *IEEE Transactions on Broadcasting*, vol. 52, no. 3, 2006, pp. 388-396.
  - [61] J. Kim, "Some Practical Guidelines for Modeling Mobile WiMAX MAC Features," *WiMAX Forum*, 2007.
  - [62] A. Bacioccola, C. Cicconetti, L. Lenzini, E. Mingozzi and A. Ertu, "A downlink data region allocation algorithm for IEEE 802.16e OFDMA," *Proc. Int. Conf. on Information, Communication & Signal Processing*, 2007, pp. 1-5.
  - [63] C. Desset, E.B. De Lima Filho, G. Lenoir, "WiMAX Downlink OFDMA Burst Placement for Optimized Receiver Duty-Cycling," *IEEE International Conference on Communications, ICC'07*, pp. 5149-5154.
  - [64] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, 1979, pp. 1-340.
  - [65] A. Lodi, S. Martello and M. Monaci, "Two-dimensional packing problems: A survey," *European Journal of Operational Research*, vol. 141, no. 2, 2002, pp. 241-252.
  - [66] F. Clautiaux, J. Carlier, and A. Moukrim, "A new exact method for the two-dimensional orthogonal packing problem," *European Journal of Operational Research*, vol. 127, no. 3, 2007, pp. 1196-1121.
  - [67] E. Hopper and B. C. H. Turton, B.C.H., "A Review of the Application of

- Meta-Heuristic Algorithms to 2D Strip Packing Problems,” *Artif. Intell. Rev. Journal*, vol. 16, no. 4, 2001, pp. 257-300.
- [68] L. Kleinrock, *Queueing Systems*, John Wiley & Sons, New York, 1975.
- [69] S. Keshav, *An Engineering Approach to Computer Networking*, Addison-Wesley, Reading, MA, 1997.
- [70] L. Moraes and I. Rubin, “Message Delays for a TDMA Scheme Under a Nonpreemptive Priority Discipline,” *IEEE Transactions on Communications*, May 1984, vol. 32, no. 5, pp. 583-588.
- [71] David Krah, “EXTEND: AN INTERACTIVE SIMULATION TOOL,” *Proceedings of the 35th conference on Winter simulation: driving innovation*, 2003, pp. 188 -196.
- [72] M. Ma, J. C. Lu, S. K. Bose and C. Ng, “A three-tier framework and scheduling to support QoS service in WiMAX,” *Proc. IEEE 6th Int. Conf. Information, Communications & Signal Processing, ICICS 2007*, Dec. 2007, pp. 1-5.
- [73] Hung-Hui Juan, Hsiang-Chun Huang, ChingYao Huang and Tihao Chiang, “Cross-layer System Designs for Scalable Video Streaming over Mobile WiMAX,” *Proceedings of IEEE Wireless Communications and Networking Conference, WCNC 2007*, pp. 1860-1864.
- [74] D. Triantafyllopoulou, N. Passas, A.K. Salkintzis and A. Kaloxylos, “A Heuristic Cross-Layer Mechanism for Real-Time Traffic in IEEE 802.16 Networks,” *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2007, pp. 1-5.
- [75] Chao-Lieh Chen, Jeng-Wei Lee, Chi-Yuan Wu and Yau-Hwang Kuo, “Fairness and QoS Guarantees of WiMAX OFDMA Scheduling with

- Fuzzy Controls,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, pp. 1-14, 2009.
- [76] Hanwu Wang and Weijia Jia, “An optimized scheduling scheme in OFDMA WiMax networks,” *International Journal of Communication Systems*, vol. 23, no. 1, pp. 23-39, Jan. 2010.
- [77] Charan Langton, *Intuitive Guide to Principles of Communications*, 2002, www.complextoreal.com/chapters/fm.pdf.
- [78] MathWorks - MATLAB and Simulink for Technical Computing. <http://www.mathworks.com/>.
- [79] L. Zhang, “Virtual clock: A new traffic control algorithm for packet switching networks,” *Proceedings of ACM SIGCOMM’90*, Philadelphia, PA, Sept. 1990, vol. 20, no. 4, pp. 19-29.
- [80] G. Song and Y. Li, “Cross-layer optimization for OFDM wireless networks-part II: algorithm development,” *IEEE Transactions on Wireless Communication*, Mar. 2005, vol. 4, no. 2, pp. 625-634.
- [81] Q. Liu, S. Zhou and G. B. Giannakis, “Cross-layer modeling of adaptive wireless links for QoS support in multimedia networks,” *Proceedings of IEEE First International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*, 2004, pp. 68-75.
- [82] Lingjie Li, Octavian Sarca, “FEC Performance with ARQ and Adaptive Burst Profile Selection,” 80216ab-01_01r2 working document, http://wirelessman.org/tg3_4/contrib/80216abc-01_60.pdf.
- [83] Theodore Rappaport, *Wireless Communications: Principles and Practices*, Pearson Education Inc., 2nd Edition, 2002.
- [84] M. Nakagami, “The m-distribution - A General Formula of Intensity

- Distribution of Rapid Fading,” In W. G. Hoffman, *Statistical Methods in Radio Wave Propagation: Proceedings of a Symposium held at the University of California*, 1960, pp. 3-36.
- [85] ITU RADIO COMMUNICATION STUDY GROUPS. Document 8F/1359-E, Aug. 2007, http://members.wimaxforum.org/documents/WiMAX_IMT_2000/itu/8F_1359.doc.
- [86] Alouini M-S and Goldsmith A. J, “Adaptive Modulation over Nakagami Fading Channels,” *Wireless Personal Communications*, vol. 13, no. 1-2, May 2000, pp. 119-143.
- [87] H. Hassanein, A. Taha and Bd-Elhamid M, “Cognitive Radio Networks Technologies Enabling Dynamic Spectrum Allocations,” 2009, http://www.ieeeegcc.org/Program/Tutorials/T1COGNITIVE_RADIO_NETWORKS_2009.pdf.
- [88] Federal Communications Commission Spectrum Policy Task Force. Report of the Spectrum Efficiency Working Group. (Tech. Rep. No. 02-135). Federal Communications Commission, 2002, http://www.fcc.gov/sptf/files/SEWGFinalReport_1.pdf.
- [89] Y. Yuan, P. Bahl, R. Chandra, T. Moscibroda and Y. Wu, “Allocating dynamic time-spectrum blocks in cognitive radio networks,” *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*, 2007, pp. 130-139.
- [90] S. Haykin, “Cognitive Radio: Brain-Empowered Wireless Communications,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, 2005, pp. 201-220.
- [91] I. F. Akyildiz, W. Lee, M. C. Vuran and S. Mohanty, “Next

- generation/dynamic spectrum access/cognitive radio wireless networks: A survey,” *Computer Networks Journal*, (Elsevier), Sept. 2006, vol. 50, no. 13, pp. 2127-2159.
- [92] Q. Zhao and B. M. Sadler, B. M, “A survey of dynamic spectrum access: Signal processing, networking, and regulatory policy,” *IEEE Signal Processing Magazine*, vol. 24, no. 3, 2007, pp. 79-89.
- [93] A. Czylik, “Adaptive OFDM for wideband radio channels,” *Proceedings of IEEE Global Telecommunications Conference, GLOBECOM'96*, 1996, pp. 713-718.
- [94] Y. J. Chang, F. C. Chien and C. C. Jay Kuo, “Cross-layer QoS Analysis of Opportunistic OFDM-TDMA and OFDMA Networks,” *IEEE Journal on Selected Areas in Communications*, 2007, vol. 25, no. 4, pp. 657-666.
- [95] M.R. Nami, M. Reza, Sharifi and Mohsen, “Autonomic Computing: A New Approach,” *First Asia International Conference on Modelling & Simulation, AMS '07*, 2007, pp. 352-357.
- [96] S. Hariri and M. Parashar, “Autonomic Computing: An overview,” *Lecture Notes in Computer Science*, Springer, 2005, pp. 247-259.
- [97] M.R. Nami and K. Bertels, “A Survey of Autonomic Computing Systems,” *Third International Conference on Autonomic and Autonomous Systems (ICAS07)*, 2007, pp. 26-26.
- [98] J. O. Kephart and D. M. Chess, “The vision of autonomic computing,” *IEEE Computer*, vol. 36, pp. 41-50, Jan. 2003.
- [99] IBM, “An Architectural Blueprint for Autonomic Computing,” April 2003, <http://www03.ibm.com/autonomic/pdfs/AC%20Blueprint%20White%20Paper%20V7.pdf>.

- [100] S. Davy, K. Barrett, S. Balasubramaniam, S. Meer, B. Jennings, J. Strassner, J, "Policy-based architecture to enable autonomic communications - a position paper," *3rd IEEE Consumer Communications and Networking Conference, CCNC 2006*, vol. 1, pp. 590-594.
- [101] D. Soldani, G. Alford, F. Parodi, M. Kylvaja, "An Autonomic Framework for Self-Optimizing Next Generation Mobile Networks," *IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks, WoWMoM 2007*, pp. 1-6.
- [102] A. Khalid, M.A. Haye, M.J. Khan, S. Shamail, "Survey of Frameworks, Architectures and Techniques in Autonomic Computing," *Fifth International Conference on Autonomic and Autonomous Systems, ICAS '09*, pp. 220 -225.
- [103] N. Ganghi, J. L. Hellerstein, S. Parekh and D. M. Tilbury. "Managing the performance of lotus notes: A control theoretic approach," 2005, http://www.research.ibm.com/PM/cmg01_CT.pdf.
- [104] V. Markl, G. M. Lohman, and V. Raman, "Leo: An autonomic query optimizer for db2," *IBM Systems Journal*, vol. 42, no. 1, pp. 98-106, 2003.
- [105] P. Shivam, S. Babu, and J. S. Chase, "Learning application models for utility resource planning," *Proceedings of the International Conference on Automatic Computing*, 2006, pp. 255-264.
- [106] D.Yagan, C.-K. Tham, "Self-Optimizing Architecture for QoS Provisioning in Differentiated Services," *Proceedings of the Second International Conference on Autonomic Computing (ICAC' 05)*, Seattle, WA, USA, June 2005.
- [107] Chong Shen, D. Pesch, J. Irvine, "Autonomic TDD link optimizing using

- hybrid wireless network and genetic algorithms,” *IEEE 62nd Vehicular Technology Conference, VTC-2005-Fall*, vol. 1, pp. 262-266.
- [108] I. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. 1991, Prentice Hall.
- [109] Kiam Cheng How, Maode Ma; Yang Qin, “A cognitive Power-Controlled Rate-Adaptive MAC protocol to support differentiated service in wireless mesh networks,” *7th International Conference on Information, Communications and Signal Processing, ICICS 2009*, pp. 1-5.
- [110] C. Balakrishna and K. Al-Begain, “Towards a User-Centric and Quality-Aware Multimedia Service Delivery Implementation on IP Multimedia Subsystem,” *Proceedings of Next Generation Mobile Applications, Services and Technologies, 2007. NGMAST '07*, Sept. 2007, pp. 36-42.
- [111] D. Vingarzan and P. Weik, “End-to-end performance of the IP multimedia subsystem over various wireless networks,” *Proceedings of IEEE Wireless Communications and Networking Conference, WCNC 2006*, vol. 1, Apr. 2006, pp. 183-188.
- [112] G. Wetzker, “Definition of spatial multiplexing gain,” *Electronics Letters*, vol. 41, no. 11, May 2005, pp. 656-657.
- [113] A. Phasouliotis and D. K. So, “Performance Analysis and Comparison of Downlink MIMO MC-CDMA and MIMO OFDMA Systems,” *Proceedings of IEEE 69th Vehicular Technology Conference, VTC Spring 2009*, Apr. 2009, pp. 1-6.
- [114] A. Ghosh, D. R. Wolter, J. G. Andrews and R. Chen, “Broadband Wireless Access with WiMax/802.16: Current Performance Benchmarks and Future Potential,” *IEEE Communications Magazine*, vol. 43, no. 2, Feb. 2005, pp. 129-136.

- [115] C. Stevenson, G. Chouinard, Zhongding Lei, Wendong Hu, S. Shellhammer and W. Caldwell, "IEEE 802.22: The first cognitive radio wireless regional area network standard," *IEEE Communications Magazine*, vol. 47, no. 1, Jan. 2009, pp. 130 - 138.
- [116] Hongjie Liu, Shufang Li and Min Yi, "Coexistence between Cognitive Radio Network and Digital TV," *Proceedings of Communication Software and Networks, ICCSN '09*. Feb. 2009, pp. 239-242.
- [117] Janny Hu and W.W. Lu, "Open wireless architecture - the core to 4G mobile communications," *Proceedings of Communication Technology, ICCT 2003*, vol. 2, Apr. 2003, pp. 1337-1342.
- [118] J. Mitola, "Cognitive radio: making software radio more personal," *IEEE Pers. Comm.*, vol. 6, no. 4, pp. 48-52, August 1999.
- [119] W.W. Lu, "An open baseband processing architecture for future mobile terminal design," *IEEE Wireless Communications*, vol. 15, no. 2, Apr. 2008, pp. 110-119.