國立交通大學

電子工程學系電子研究所碩士班

碩士論文

針對上行頻寬索取與傳輸機制
在行動WiMAX系統

Performance Analysis of Uplink Scheduling And Bandwidth Request Mechanism in Mobile WiMAX

研究生：何玠原

指導教授：黃經堯　博士

中華民國　九十六年十二月

針對上行頻寬索取與傳輸機制在行動WiMAX系統之無線資源管理效能分析

Performance Analysis of Uplink Scheduling And Bandwidth Request Mechanism in Mobile WiMAX

研究生：何玠原　　　　Student: Chieh Yuan Ho

指導教授：黃經堯　　　　Advisor: Ching Yao Huang

國 立 交 通 大 學

電子工程學系電子研究所碩士班

碩 士 論 文

A Thesis
Submitted to Department of Electronics Engineering & Institute of Electronics
College of Electrical and Computer Engineering
National Chiao Tung University
In Partial Fulfillment of the Requirements
For the Degree of
Master
in
Electronics Engineering
December 2007
HsinChu, Taiwan, Republic of China

中華民國 九十六年十二月

針對上行頻寬索取與傳輸機制在行動WiMAX系統之無線資源管理效能分析

國 立 交 通 大 學 電 子 工 程 學 系 電 子 研 究 所 碩 士 班

學生: 何玠原　　　　　　　　　　　　指導教授:黃經堯 博士

摘　　要

在這篇論文中，我們介紹了行動 **WiMAX** 的物理層和媒介接取控制層。而行動 **WiMAX** 引用了 **IEEE 802.16d-2004** 和 **IEEE 802.16e-2005** 兩樣技術標準當參考。我們亦架構了 **Mobile WiMAX** 系統層級的模擬工具，且實現上行頻寬索取機制。 藉此模擬系統，我們完成了一套完整的上行傳輸機制效能分析，並實現了四種最通用的排程演算法。最後，我們更進一步的將不同的排程演算法在單一資料形式與混合資料形式的環境下模擬其效應，以探討不同演算法各自之間在傳輸速率、服務品質控制、以及系統容量等方面的優缺點，並觀察各自最適合的使用時機。這四種排程演算法包含了 **Early Deadline First (EDF)**、**Proportional Fair (PF)**、**Maximum CINR (MaxCINR)**、和 **Round-Robin (RR)**。另外，我們提出了一套時間延遲限制的概念，可應用於非即時傳輸服務(如 **FTP**)的排程演算法。這個概念主要是將最小傳輸速度的限制，針對每一個封包，將其轉化為時間上的限制，不但提高了排程器的設計簡便性，也比單純用傳輸速率來當排程的指標的方法，有更好的效能表現。

Performance Analysis of Uplink Scheduling And Bandwidth Request Mechanism in Mobile WiMAX

Student: Chieh-Yuan Ho **Advisor: Ching-Yao Huang**

Department of Electronic Engineering

Institute of Electronics

National Chiao Tung University

Abstract

In this thesis, we introduce the PHY and MAC of Mobile WiMAX, which applies the IEEE 802.16d-2004 and IEEE 802.16e-2005 standard as reference. Then, we build up a system-level simulator and a signaling control plane for the uplink bandwidth request Mechanism for the purpose of getting a complete uplink performance analysis and investigating the advantages and disadvantages of different scheduling algorithms, including Early Deadline First (EDF), Proportional Fair (PF), Maximum CINR (MaxCINR), and Round-Robin (RR). We further discuss the capacity and QoS issue in terms of different traffic type in both single traffic and mixed traffic environment. Through the simulation result, the uplink MAC throughput in Mobile WiMAX is clearly revealed, and several common scheduling algorithms are implemented to get a complete uplink performance analysis. In addition, a soft delay bound (SDB) method is used in order to properly schedule non-real-time service. It transforms the minimum rate into time domain delay bound. It facilitates the scheduler design and performs better QoS failure rate than just using data rate as the QoS indicator.

致謝

光陰似箭，碩士班的兩年即將畫下了句點。在這一段過程中，認識許多好夥伴，其中更是有許多在我在專業領域研究上的貴人，很慶幸有你們的陪伴與幫助，在此我願意將我小小的成果分享給在這兩年陪伴我的你們。

首先，我要感謝我的指導教授，黃經堯老師，除了在專業上授予我許多關於無線通訊的知識外，也培養我如何去找出問題、分析問題、最後解決問題的能力。也要特別感謝老師在這兩年內，並不反對我往工程以外的領域去發展。兩年來，我參加了許多的創業競賽、並參與一些商業創業相關的活動，老師從沒表示反對過，並支持我去追逐自己的夢想，更常常的提供我一些商業計劃的機會，如IP MALL等，讓我除了電子工程專業之外，有更多機會去磨練培養商業創業的能力。如果沒有黃老師兩年來的支持，我絕對沒有辦法在兩年內完成這麼多的事情，所以在此，我要特別感謝老師兩年來的教誨。

另外，要感謝所有實驗室的學長與同學們兩年來的相處與照顧，平常同儕之間的討論，讓我的專業知識能夠快速的累積。

最後，要感謝一直很支持我的家人，爸爸、媽媽、弟弟等，不論我做了什麼決定，他們都是全力支持。碩一時，為了不讓自己只鑽研於一個狹小的工程領域，並擴展自己的視野，培養自己的競爭力，我決定修習交大科管輔所，並有計畫的參加一系列的創業競賽與商業個案競賽，之後更產生了創業的熱忱。這一切的決定，都需要爸媽與黃老師的大力支持才能支持我不放棄的繼續做下去。畢竟這將會是一條艱辛、坎坷的路，需要的是不斷的努力與堅持。家人在背後的支持，永遠是我動力的泉源。

碩士畢業後，我將繼續攻讀博士，讓自己的專業知識更豐富、更紮實，也希望能在專業上、研究方向、或是其他方面爲Wintech做點貢獻，實驗室越強、越有制度，學生在未來就越能有更好的發展。

Table of Content

Figures and Tables

Chapter 1 Introduction

The WiMAX technology, based on the IEEE 802.16-2004 Air Interface Standard is rapidly proving itself as a technology that will play a key role in fixed broadband wireless metropolitan area networks. In December, 2005 the IEEE ratified the 802.16e amendment to the 802.16 standard. This amendment adds the features and attributes to the standard necessary to support mobility. The WiMAX Forum defined system performance and certification profiles based on the IEEE 802.16e-2005 Amendment standard. Going beyond the air interface, the WiMAX Forum defined the end-to-end network architecture for implementing a Mobile WiMAX network as well.

Mobile WiMAX is a broadband wireless solution that enables convergence of mobile and fixed broadband networks through a common wide area broadband radio access technology and flexible network architecture. It can achieve extremely high data rate to enable many applications and accommodate subscribers' demand nowadays. In Mobile WiMAX, Scalable OFDMA (S-OFDMA) technology is used in order to support scalable channel bandwidth from 1.25MHz to 20MHz. And an OFDMA frame is divided into downlink sub-frame and uplink sub-frame in TDD mode. For data traffic, Mobile WiMAX is designed as a connection-based technology, and each connection has a connection ID (CID) and a service flow ID (SFID) for BS to manage the quality of service. In order to ensure QoS quality or perform high system performance, scheduling algorithms, which determine the order of transmission, are highly required in both downlink and uplink to accommodate various demands in different scenarios.

In this thesis, we focus on the performance analysis of uplink transmission to address the traffic-load capability in Mobile WiMAX. In our simulation platform, we build the architecture of uplink mechanism by implementing several scheduling algorithms and bandwidth request and grant mechanism.

The rest of this thesis is organized as follows: In chapter 2, the overview of PHY and MAC layers of Mobile WiMAX are briefly introduced. In chapter 3, Bandwidth Request Mechanism and Scheduling algorithm of Uplink Transmission is discussed in detail. In chapter 4, the setting of simulation platform is addressed. In chapter 5, the simulation results are shown. Finally, the conclusion and future works are given in chapter 6.

Chapter 2 Overview of PHY and MAC Layer of Mobile WiMAX

2.1 Introduction to PHY Layer of Mobile WiMAX

2.1.1 OFDMA Basis

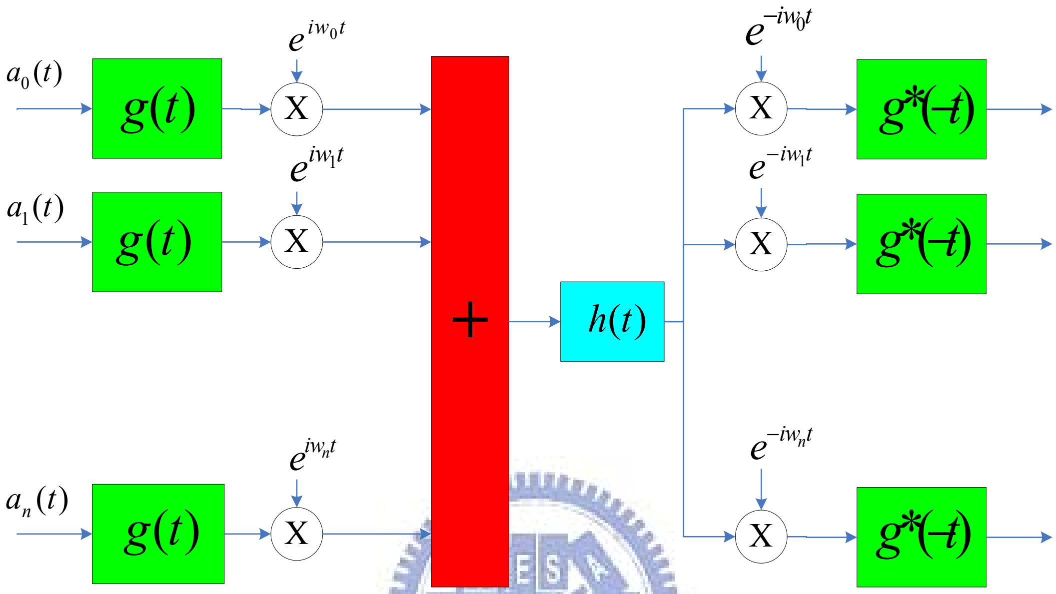

Figure 2-1 Basic Architecture of OFDM system

Orthogonal Frequency Division Multiplexing (OFDM) [6,7] is a multiplexing technique that divides the bandwidth into multiple frequency sub-carriers as shown in Figure 2-1. In an OFDMA system, the input data stream is divided into several parallel sub-streams and each sub-stream is modulated and transmitted on a separate orthogonal sub-carrier. The increased symbol duration improves the robustness of OFDM to delay spread. Furthermore, the cyclic prefix (CP) is being introduced in order to completely eliminate Inter-Symbol Interference (ISI) as long as the CP duration is longer than the channel delay spread. The CP is a repetition of the last part of data portion of the block, and it is appended to the beginning of the data payload as shown in Figure 2-2. The function of Cyclic Prefix is to prevent inter-block interference and makes the channel appear circular and permits low-complexity frequency domain equalization. The obvious drawback of Cyclic Prefix is that it introduces overhead, which effectively reduces bandwidth efficiency. Since OFDM has a very sharp spectrum, a large fraction of the allocated channel bandwidth can be utilized for data transmission, which helps to reduce the loss in bandwidth efficiency caused by Cyclic Prefix.

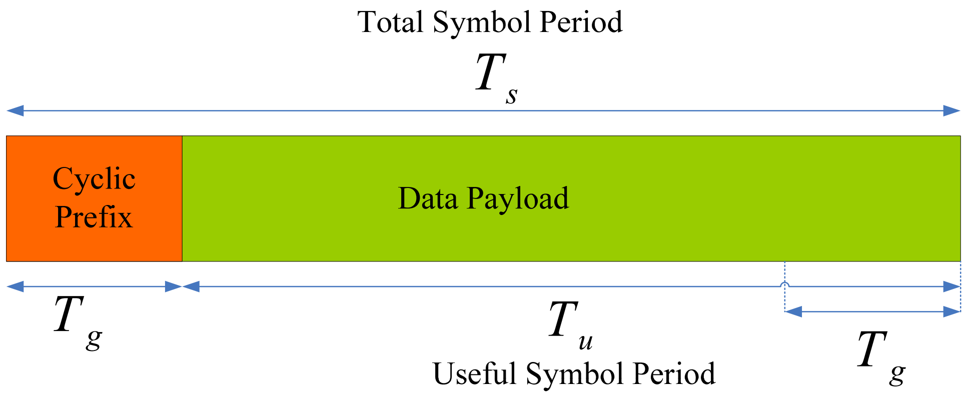

Figure 2-2 Insertion of Cyclic Prefix (CP)

2.1.2 OFDMA Symbol Structure and Sub-Channelization

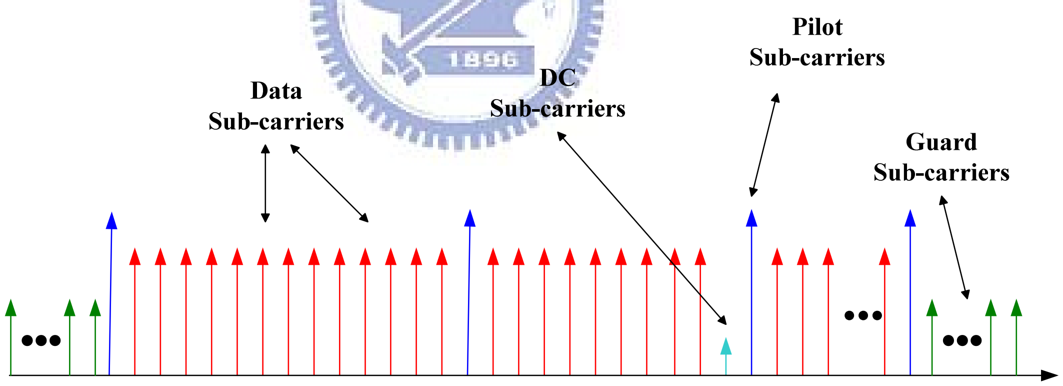

Figure 2-3 OFDMA Sub-Carriers structure

The OFDMA symbol structure consists of three types of sub-carriers as shown in Figure 2-3:

- Data sub-carriers for data transmission
- Pilot sub-carriers for estimation and synchronization purposes
- Null sub-carriers for no transmission; used for guard bands and DC carriers

Data and pilot sub-carriers are grouped into subsets of sub-carriers, called sub-channels.

The WiMAX OFDMA PHY [3] supports sub-channelization in both Downlink and Uplink. The minimum frequency-time resource unit of sub-channelization is one slot, which is equal to 48 data tones.

Scalable OFDMA

The IEEE 802.16e-2005 OFDMA mode is based on the concept of scalable OFDMA supporting various bandwidths to flexibly address the need for various spectrum allocation and application requirements.

The scalability is supported by adjusting the FFT size while fixing the sub-carrier frequency spacing at 10.94 kHz. Since the resource unit sub-carrier bandwidth and symbol duration is fixed, the impact to higher layers is minimal when scaling the bandwidth. The Scalable OFDMA parameters are listed in Table 2-1, 2-2.

Table 2-1 Scalable OFDMA parameters

Parameters	Values			
System Channel Bandwidth (MHz)	1.25	5	10	20
Sampling Frequency (F_p in MHz)	1.4	5.6	11.2	22.4
FFT Size (N_{FFT})	128	512	1024	2048
Number of Sub-Channels	2	8	16	32
Sub-Carrier Frequency Spacing	10.94 kHz			
Useful Symbol Time(T_b = 1/f)	91.4 microseconds			
Guard Time ($T_g = T_b/8$)	11.4 microseconds			
OFDMA Symbol Duration ($T_s = T_b + T_g$)	11.4 microseconds			
OFDMA Symbol Duration ($T_s = T_b + T_g$)	48			

Table 2-2 Downlink and Uplink sub-carriers setting

Parameter	Downlink	Uplink	Downlink	Uplink
System Bandwidth	5 MHz		10 MHz	
FFT Size	512		1024	
Null Sub-Carriers	92	104	184	184
Pilot Sub-Carriers	60	136	120	280
Data Sub-Carriers	360	272	720	560

Sub-Channels	15	17	30	35
Symbol Period, T_s	102.9 microseconds			
Frame Duration	5 milliseconds			
OFDM Symbols/Frame	48			
Data OFDM Symbols	44			

2.1.3 TDD Frame Structure of OFDMA

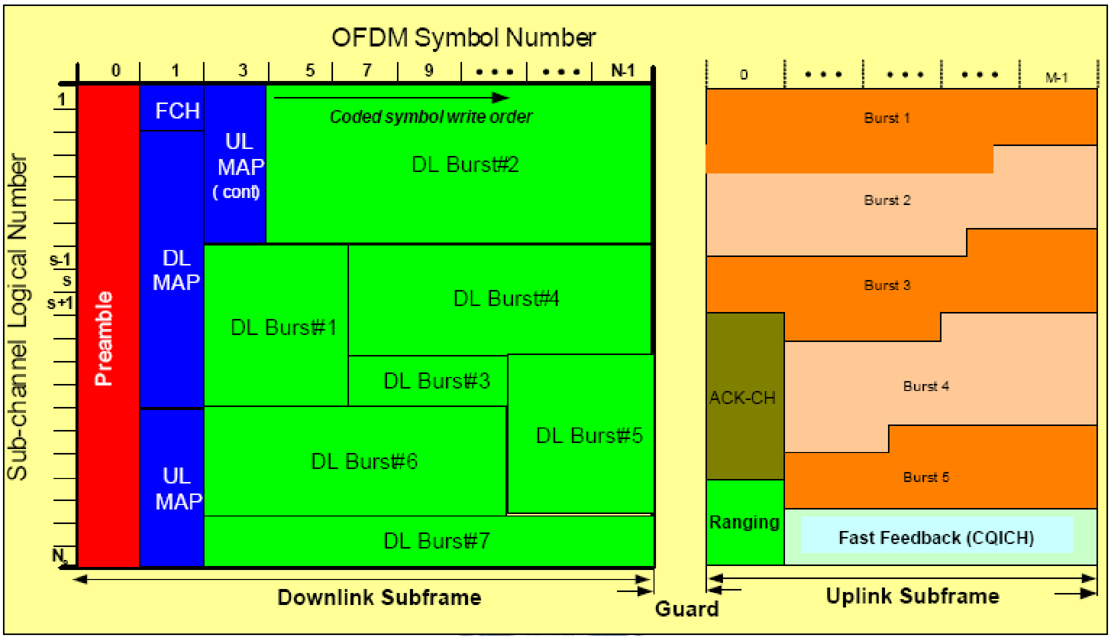

Figure 2-4 WiMAX OFDMA Frame Structure

The PHY Layer in the IEEE 802.16e-2005 standard [3] supports TDD and Full and Half-Duplex FDD operation, but TDD mode is much more common than FDD mode which is usually applied for some specific reasons; TDD mode enables adjustment of the downlink/uplink ratio to efficiently support asymmetric downlink/uplink traffic load. Unlike FDD, which requires a pair of channels, TDD only requires a single channel for both downlink and uplink providing greater flexibility for adaptation to varied global spectrum allocations. In addition, Transceiver designs for TDD mode implementations are less complex and therefore less expensive.

Figure 2-4 shows the OFDMA frame structure for Time Division Duplex (TDD) mode of Mobile WiMAX. Each frame is divided into DL and UL sub-frames separated by Transmit/Receive and Receive/Transmit Transition Gaps (TTG and RTG, respectively) to prevent DL and UL transmission collisions. In the downlink sub-frame, the Preamble is allocated in the beginning in order to execute synchronization which is a critical issue for TDD mode operation. In order to completely address the synchronization issue, system-wide synchronization is highly required. The following introduces control information in the DL/UL sub-frame, which is used to ensure optimal system operation:

- **Preamble:** The preamble, used for synchronization, is the first OFDMA symbol of the frame.
- **Frame Control Header (FCH):** The FCH follows the preamble. It provides the frame configuration information such as MAP message length and coding scheme and usable sub-channels.
- **DL-MAP and UL-MAP:** The DL-MAP and UL-MAP provide sub-channel allocation and other control information for the DL and UL sub-frames respectively.
- **UL Ranging:** The UL ranging sub-channel is allocated for mobile stations (MS) to perform closed-loop time, frequency, and power adjustment as well as bandwidth requests.
- **UL CQICH:** The UL CQICH channel is allocated for the MS to feedback channel state information.
- **UL ACK:** The UL ACK is allocated for the MS to feedback DL HARQ acknowledge.

2.1.4 Resource Allocation of Downlink and Uplink Sub-Frame

The OFDMA slot is a minimum unit for data transmissions. One OFDMA slot occupies one sub-channel and several OFDMA symbols depending on different slot structures. For downlink Full Usage of Sub-carriers (FUSC) using the distributed sub-carrier permutation, one slot is one sub-channel by one OFDMA symbol. For downlink Partial Usage of Sub-carriers (PUSC) using the distributed sub-carrier permutation, one slot is one sub-channel by two OFDMA symbols. For uplink PUSC using either of the distributed sub-carrier permutations, one slot is one sub-channel by three OFDMA symbols. For uplink and downlink Band Adaptive modulation and coding (Band AMC) using the adjacent sub-carrier permutation. One slot is one sub-channel by one, two, three, or six OFDMA symbols.

A Data Region is a two-dimensional allocation which contents a group of contiguous sub-channels and OFDMA symbols. All the allocation refers to logical sub-channels. The minimum unit of data mapping is an OFDMA slot.

Based on the standard, how many and which resource units would be assigned to a transmission is decided by BS, and the mechanism is different in downlink and uplink transmission. In downlink resource allocation, system will consider the data size and try to fulfill the resource units into sub-channels of frequency domain first. After the frequency domain is fulfilled in that particular user data burst, then it goes to time domain to fulfill resource units until data traffic for that frame is done. In uplink resource allocation, in a particular user data burst, resource units are fulfilled in time domain first, after the time domain resource units are full, then go to the next frequency domain until data traffic in that frame is done.

Figure 2-5 and Figure 2-6 show downlink resource allocations and uplink resource allocation mechanism.

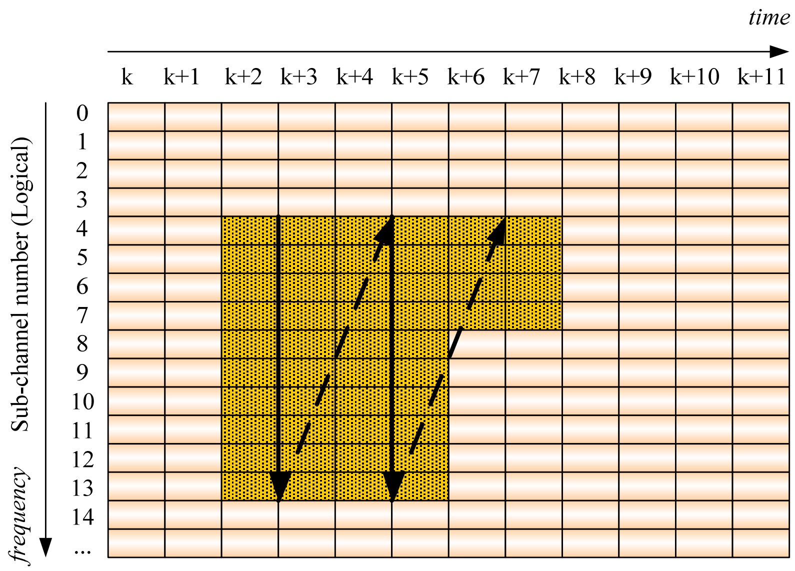

Figure 2-5 DL Resource Allocation

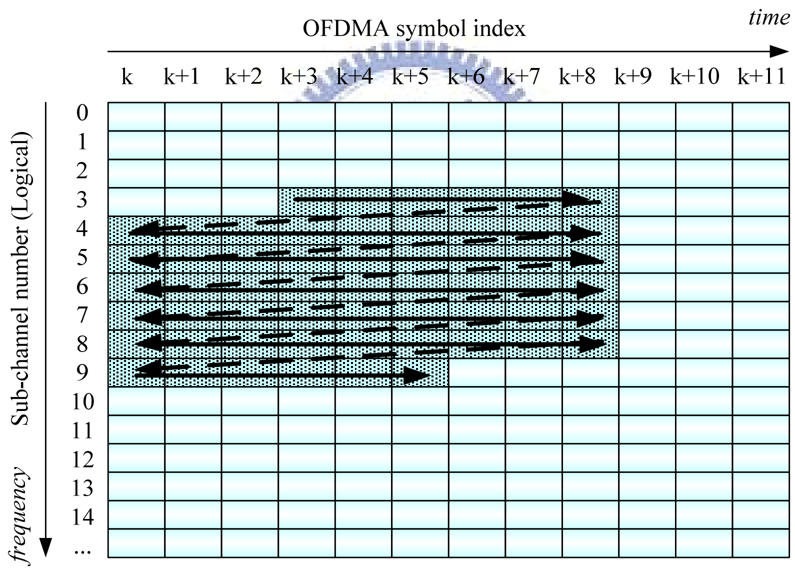

Figure 2-6 UL Resource Allocation

2.1.5 Sub-carrier Permutation

Sub-carrier permutation is a method to assign frequency sub-carriers into sub-channels. The allocation of sub-carriers to sub-channels is accomplished via permutation rule.

There are two categories of permutation modes: distributed sub-carrier permutation and adjacent sub-carrier permutation. Distributed permutation means the sub-carriers belonging to a sub-channel are selected pseudo-randomly from all sub-carriers. It can significantly reduce interference from other sectors or cells, and avoid fading effect, such as frequency selective fading. The adjacent sub-carrier permutation will form the sub-channel whose sub-carriers coming from adjacent sub-carriers. System using this permutation mode can take advantage of frequency select fading and get multi-user diversity on the frequency domain at the same time.

In order to facilitate a wide range of usage and applications under various requirements and geographic constraints, Mobile WiMAX supports different sub-carriers permutation mode for grouping sub-carriers into sub-channels:

- Full Usage of Sub-channels (FUSC)

  This method is used in downlink only and can use all sub-carriers to do permutation for one sub-channel. It can achieve the best frequency diversity by spreading sub-carriers over entire band. It will use distributed permutation mode.

- Partial Usage of Sub-channels (PUSC)

  This method can be used both in downlink and uplink. It's the most common permutation method used to implement Mobile WiMAX or 802.16e-2005. Sub-carriers are grouped into several clusters first. And sub-carriers are chosen one by one from each sub-carrier cluster to form a sub-channel. There are two different implementations in Downlink PUSC and Uplink PUSC. The following is the summary of DL-PUSC used 2048 FFT point as an example.

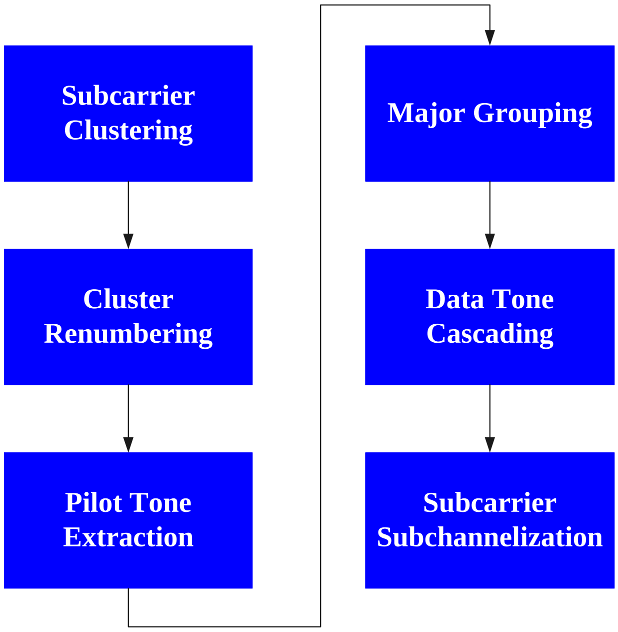

Figure 2-7 Flow chart of OFDMA PUSC mode

In 802.16e-2005, 2048 carriers include 1681 used tones (data plus pilot tones). Figure 2-7 shows the flow chart of OFDMA PUSC. The first step of DL-PUSC is to divide these 1680 tones into 120 physical clusters of 14 tones. And then Re-number these divided 120 physical clusters with logical indices for the purpose of interleaving the clusters. After that, 2 tones in each of 120 clusters will be taken out as the pilot tones. Then, major grouping is done by dividing 120 clusters into 6 major groups. The even groups have 24 clusters, and the odd ones have 16 clusters. Fourth, all data tones of each of the major groups are concatenated for sub-channel selection. Finally, Data tones are selected from each major group as a sub-channel according to a specific sequence.

➢ Band Adaptive Modulation and Coding (BandAMC)

This method also can be applied in downlink and uplink. It is one of the adjacent permutation methods. The total bandwidth is divided into several sub-bands and tries to utilize the frequency select fading to enhance system performance. The following figure shows the basic rule of Band AMC sub-carrier allocation.

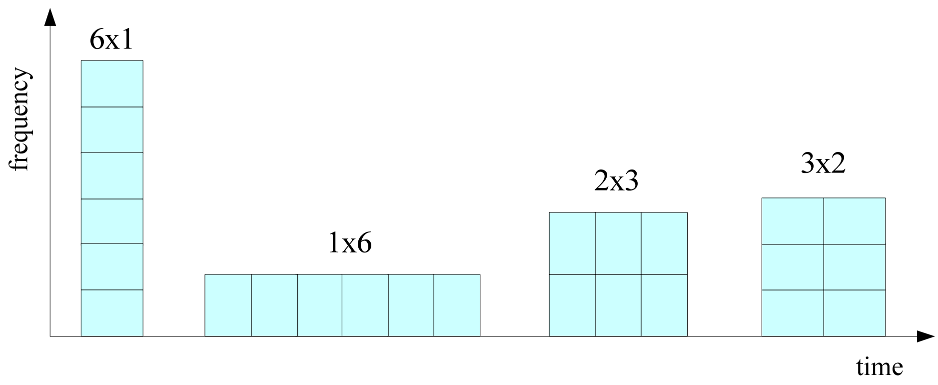

Figure 2-8 Band AMC slot structure

Take 2048 FFT size as an example, there're 1729 used tones and 319 guard tones. A band-AMC sub-channel consists of six 9-carrier bins. That is 54 carriers per sub-channel 192 bins in total. And there are four categories of six bins grid according to different OFDMA symbol duration in Band-AMC mode, as shown in Figure 2-8.

- 1-OFDMA-symbol duration, 32 sub-channels
- 2-OFDMA-symbol duration, 62 sub-channels.
- 3-OFDMA-symbol duration, 96 sub-channels.
- 6-OFDMA-symbol duration, 192 sub-channels.

2.1.6 Adaptive Modulation and Coding (AMC)

Adaptive modulation and coding (AMC), Hybrid Automatic Repeat Request (HARQ) and Fast Channel Feedback (CQICH) were introduced with Mobile WiMAX to enhance coverage and capacity for WiMAX in mobile applications. Support for QPSK, 16QAM and 64QAM are mandatory in the DL with Mobile WiMAX. In the uplink transmission, 64QAM is optional. Both Convolutional Code (CC) and Convolutional Turbo Code (CTC) with variable code rate and repetition coding are supported. Block Turbo Code and Low Density Parity Check Code (LDPC) are supported as optional features. Table 2-3 summarizes the coding and modulation schemes supported in the Mobile WiMAX profile the optional UL codes and modulation are shown in italics.

Table 2-3 Supported Code and Modulations

		DL	UL
Modulation		QPSK, 16QAM, 64QAM	QPSK, 16QAM, 64QAM
Code Rate	CC	1/2, 2/3, 3/4, 5/6	1/2, 2/3, 5/6
	CTC	1/2, 2/3, 3/4, 5/6	1/2, 2/3, 5/6
	Repetition	x2, x4, x6	x2, x4, x6

2.1.7 Hybrid Auto Repeat Request (HARQ)

Hybrid Auto Repeat Request (HARQ) is also supported by Mobile WiMAX. HARQ is enabled using *N* channel "Stop and Wait" protocol which provides fast response to packet errors and improves cell edge coverage. Chase Combining and Incremental Redundancy are supported to further improve the reliability of the retransmission. A dedicated ACK channel is also provided in the uplink for HARQ ACK/NACK signaling.

Multi-channel HARQ operation is supported. Multi-channel stop-and-wait ARQ with a small number of channels is an efficient, simple protocol that minimizes the memory required for HARQ and stalling [8]. HARQ combined together with CQICH and AMC provides robust link adaptation in mobile environments at vehicular speeds in excess of 120 km/hr.

2.1.8 Frequency Reuse Factor

Cellular deployment scenarios specify the pattern of radio frequency channel (or carrier) usage in terms of a “frequency reuse factor”. Sub-carriers in the radio frequency band are assigned to different cells or sectors and this allocation is repeated across adjacent sites (cells or sectors) or adjacent cluster of sites throughout the wireless infrastructure. The resulting frequency reuse planning can be indicated as the triplet (*c, s, n*) where *c* is the number of BS sites per cluster, *s* is the number of sectors per BS site and *n* is the number of unique frequency channels needed for reuse. Typical examples of reuse (1, 3, 1) and (1, 3, 3) are shown in Figure 2-9 and 2-10.

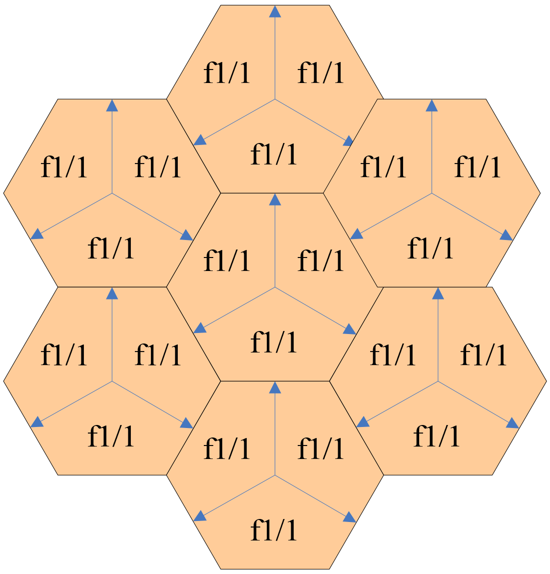

Figure 2-9 frequency reuse (1, 3, 1)

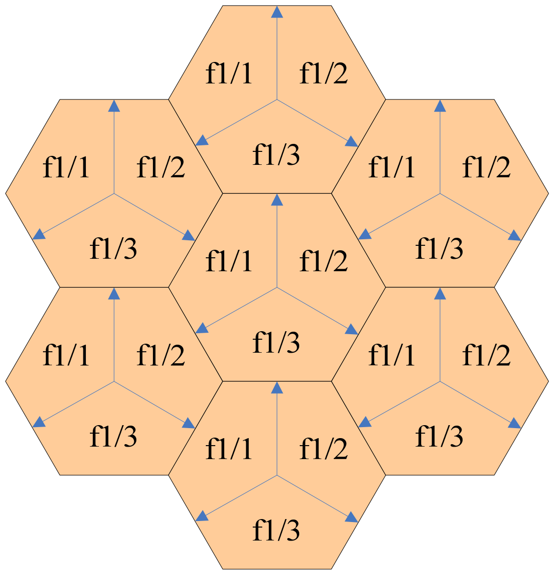

Figure 2-10 frequency reuse (1, 3, 3)

➢ Frequency reuse (3, 3, 3)

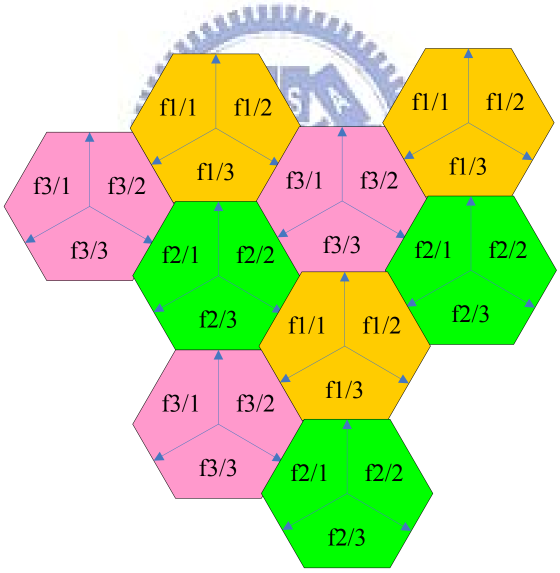

Figure 2-11 frequency reuse (3, 3, 3)

As shown in Figure 2-11, frequency reuse (3, 3, 3) represents for three BS sites per cluster. The total RF channel is divided into three fractions, and each of BS in a cluster operates on a different fraction of the RF channel. Also, there are three unique frequency channels per BS site needed for reuse.

➢ Fractional frequency reuse

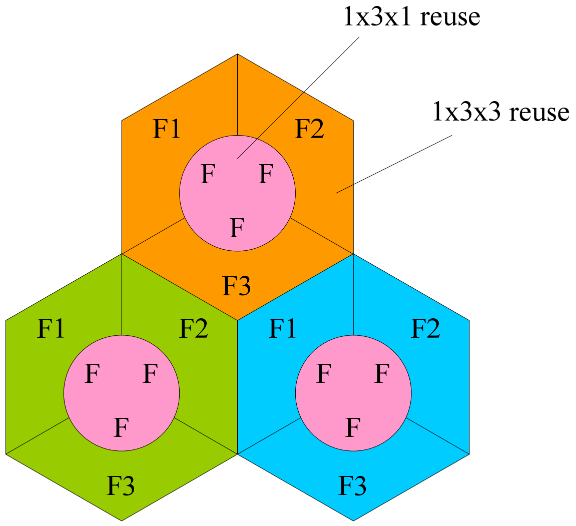

Figure 2-12 fractional frequency reuse inner (1, 3, 1) & outer (1, 3, 3)

In addition to the frequency reuse factors mentioned above, factional frequency reuse can be a more efficient way to reuse the bandwidth without much increasing the interference MSs gets. The sub-channel reuse pattern can be configured so that users close to the base station operate on the zone with all sub-channels available. While for the edge users, each cell or sector operates on the zone with a fraction of all sub-channels available, which is frequency reuse (1, 3, 3), and F1, F2, and F3 represent different sets of sub-channels in the same frequency channel, as shown in Figure 2-12. With this configuration, while fractional frequency reuse is implemented for edge users to assure edge-user QoS quality and throughput the full load frequency reuse (1, 3, 1) is used for center users to fully utilize the air resource and maximize spectral efficiency. Based on network load, the frequency reuse planning can be dynamically adjusted and optimized across sectors or cells and channel interference conditions on a frame by frame basis.

2.1.9 Ranging

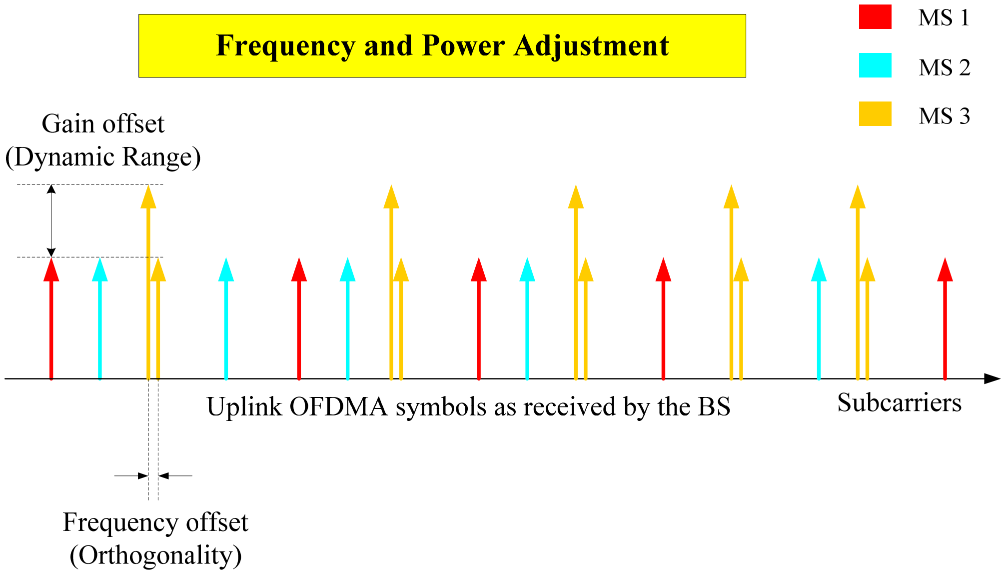

Figure 2-13 Ranging – frequency and power adjustment

The ranging process in which the MS acquires frequency, time and power adjustments, after which all MS transmissions are aligned with the UL sub-frame received by the BS. Ranging process is based on MS transmitting a signal and BS responding with required adjustments (RNG-REQ/RSP).

As shown in Figure 2-13, assume MS #3 is going to enter the ranging procedure. Before ranging, there are a power offset and frequency offset which is unaligned with the uplink sub-frame with the serving BS. Those unaligned offsets in frequency and power domain cause data loss, high packet error rate, high interference, and incapability to decode data burst due to the unsynchronized uplink sub-frame.

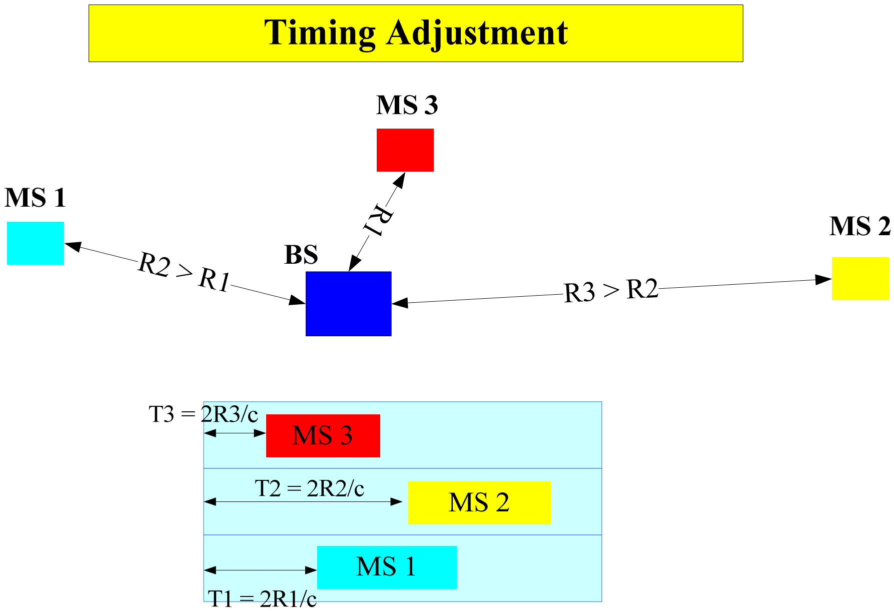

Figure 2-14 Ranging – timing adjustment

In addition to power and frequency adjustment, timing adjustment is also necessary in ranging process. The distances between MSs and the serving BS are R1, R2, and R3 for MS #1, MS #2, and MS #3 respectively. For the purpose of adjusting the timing of MS uplink transmission for the alignment of uplink sub-frame, different transmission timing, T1, T2, and T3 for MS #1, MS #2, and MS #3 respectively are calculated based on the distance between MS and serving BS, so that uplink transmissions of all MSs in a serving BS are synchronized in terms of timing.

After frequency, power, and time adjustment as shown in Figure 2-14, the MS is aligned with the serving BS on the uplink sub-frame in terms of those three items, so that the reception of its data burst is able to be done successfully.

2.1.10 Smart Antenna Technology

Smart antenna technologies typically involve complex vector or matrix operations on signals due to multiple antennas. OFDMA allows smart antenna operations to be performed on sub-carriers. Complex equalizers are not required to compensate for frequency selective fading. OFDMA therefore, is very well-suited to support smart antenna technologies. In fact, MIMO-OFDM/OFDMA is envisioned as the corner-stone for next generation broadband communication systems [11,12]. Mobile WiMAX supports a full range of smart antenna technologies to enhance system performance. The supporting smart antenna technologies include:

- **Beamforming:** With beamforming [13], the system uses multiple-antennas to transmit weighted signals to improve coverage and capacity of the system and reduce outage probability.
- **Space-Time Code (STC):** Transmit diversity is supported to provide spatial diversity and reduce fade margin.
- **Spatial Multiplexing (SM):** Spatial multiplexing [16,17] is supported to take advantage of higher peak rates and increased throughput. With spatial multiplexing, multiple streams are transmitted over multiple antennas. If the receiver also has multiple antennas, it can separate the different streams to achieve higher throughput in comparison to single antenna systems. With 2x2 MIMO, SM increases the peak data rate two times by transmitting two data streams. In UL, each user has only one transmit antenna, two users can transmit collaboratively at the same time slot as if two streams are spatially multiplexed from two antennas of the same user. This is called UL collaborative SM.

2.2 Introduction to MAC Layer of Mobile WiMAX

2.2.1 Layer Structure

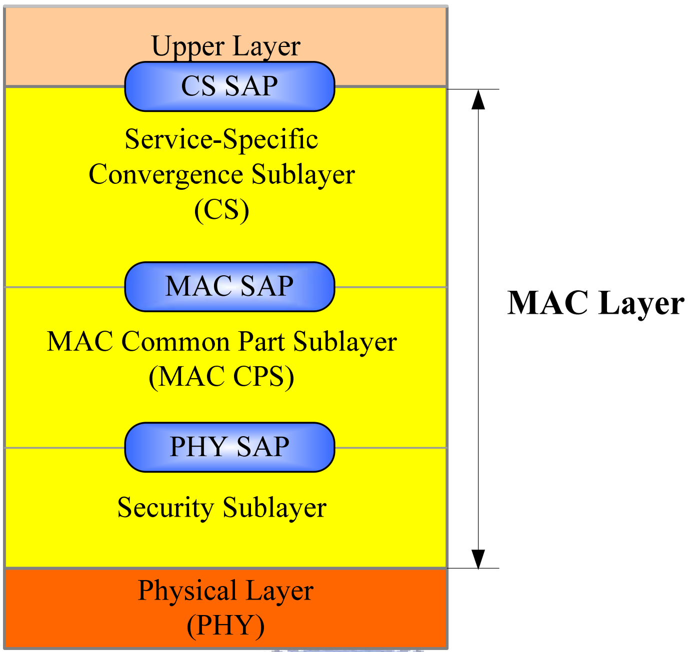

Figure 2-15 PHY-MAC Structure in Mobile WiMAX

As shown in Figure 2-15, MAC Layer includes three sublayers which are Service-Specific Convergence Sublayer (CS), MAC Common Part Sublayer (MAC CPS), and Security Sublayer respectively. And in between each of two sublayers pair, there is a service access point, so-called SAP, as a interface to access the data from the upper layer or downer layer. CS SAP is the interface between MAC layer and Network layer, and PHY SAP is the interface between MAC layer and Physical layer. While the CS and Security sub-layer are in charge of connection management and security management respectively, most of the resource managements are done in the MAC CPS, including PDU generation, handoff, ARQ, network entry, etc. The details about the three sub-layers are introduced in the following.

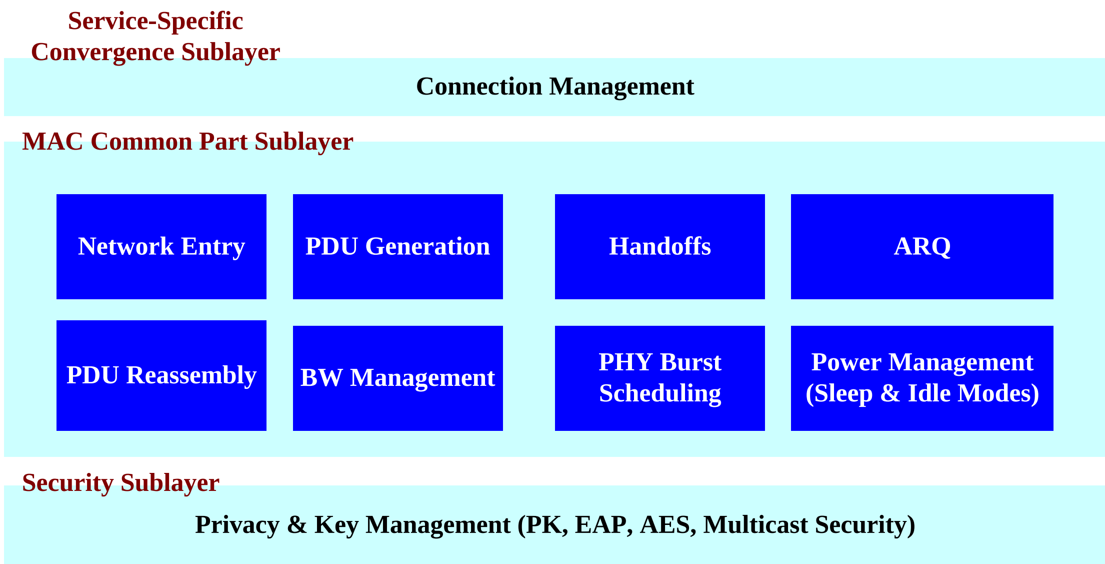

Figure 2-16 MAC Structure in Mobile WiMAX

In Figure 2-16, the three sub-layers in MAC layer are introduced.

- **Service-Specific Convergence Sublayer (CS)**

There are two general service-specific convergence sublayers in 802.16e-2005 that mapping services to and from MAC connections. The ATM convergence sublayer and the packet convergence sublayer are for ATM services and packet-based services such as IP, Ethernet and VLAN respectively. The main purpose of SSCS is to classify service data units (SDUs) into proper MAC connections, preserve or enable QoS, and enable bandwidth allocation. The mapping forms may vary due to the type of service. Furthermore, more complicated functions are also provided by the convergence sublayers such as payload header suppression and reconstruction to enhance air link efficiency. [Eklund02]

- **MAC Common Part Sublayer (MAC CPS)**

The 802.16e-2005 MAC Layer supports point-to-multipoint (PMP) architecture with a central base station (BS) dealing with multiple sectors simultaneously.

The MAC Common Part Sublayer is connection-oriented. Each of services is being mapped to one connection even if it is a connectionless service inherently. It enables the service flow to have the capabilities of bandwidth request, QoS association, data transmission and routing, and all other related actions. Connections are identified by 16-bit connection identifiers (CIDs) and may require continuously granted bandwidth or bandwidth on demand.

There is a standard 48-bit MAC address in each MS, but this serves mainly as an equipment identifier, since the primary addresses used during operation are the CIDs. While entering the network, the SS is assigned three management connections per direction. Three different QoS requirements are needed due to these three connections may have different management levels. The first one is the basic connection, which is used for the transfer of short, time-critical MAC and radio link control (RLC) messages. The primary management connection is used to transfer longer, more delay-tolerant messages such as authentication and connection setup. The secondary management connection is used for the transfer of standards-based management messages such as Dynamic Host Configuration Protocol (DHCP), Trivial File Transfer Protocol (TFTP), and Simple Network Management Protocol (SNMP). Except these management connections, transport connections are allocated for the contracted services by MSs. Transport connections are unidirectional for being accommodated to different uplink and downlink QoS and traffic parameters which are usually assigned to services in pairs.

There are some additional connections reserved for contention-based initial access, downlink broadcast transmissions, signaling broadcast contention-based polling, and multicast contention-based polling. MSs may be ordered to join multicast polling groups associated with them. [Eklund02]

- **Security Sublayer**

Two main protocols work in this security sublayer, one is for encrypting packet data across the fixed BWA i.e. encapsulation protocol, and the other one is for secure distribution of keying data from BS to MS called Privacy and Key Management

Protocol (PKM). The PKM protocol applies, RSA public-key algorithm, X.509 digital certificates, and strong encryption algorithm to perform secure key exchanges between MSs and BSs. This Privacy protocol is based on the PKM protocol and has been enhanced to fit seamlessly into the 802.16e MAC Layerand to accommodate to stronger cryptographic methods such as AES.

2.2.2 MAC PDU Format

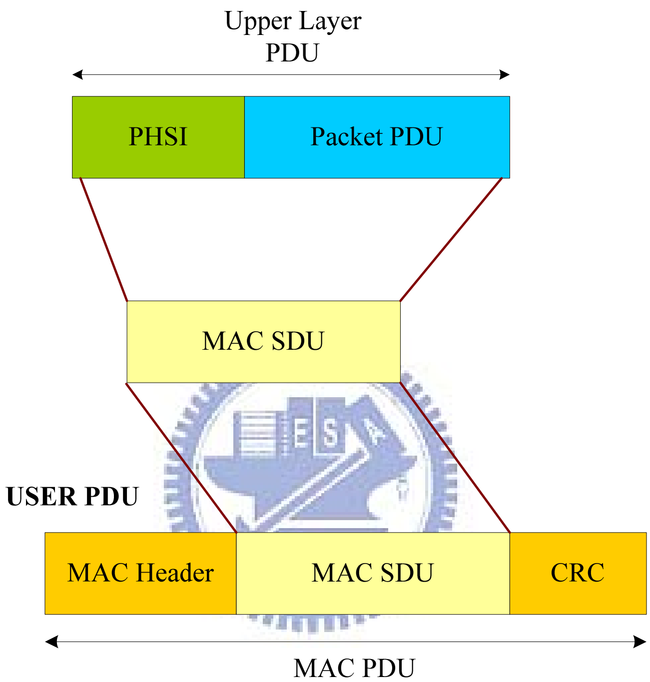

Figure 2-17 MAC PDU Format

The MAC PDU is a data unit between the peer MAC. A MAC PDU consists of a 48bit MAC header, a variable length data payload, and an optional 32 bits Cyclic Redundancy Check (CRC) as shown in Figure 2-17. Sometimes some MAC PDU will not include payload and CRC bits. These kinds of PDUs are used only in the uplink to transmit control message. Those MAC signaling headers include bandwidth request, uplink transmit power report, CINR report, CQICH allocation request, PHY channel report, uplink sleep control, SN report, and feedback functionalities. MAC PDUs also include some subheaders. Those sub-headers are inserted in MAC PDUs following the generic MAC header. Those sub-headers enable the system to perform grant management, packing, ARQ feedback, and so on.

2.2.3 Fragmentation and Packing

MAC SDUs from SSCS will be formatted according to the MAC PDU format in the CPS, possibly with fragmentation and packing due to the precious radio resources and efficient utilization of the resources.

In fragmentation process, a SDU is divided into different PDUs payload areas due to the constraint of the MAC PDU size with the maximum 2048 bytes in the IEEE 802.16e standard. In addition, for preventing high Packet error rate due to too large PDU size, dividing SDU properly according to the channel condition is necessary. Figure 2-18 illustrates fragmentation process.

In packing process, several SDUs are packed into a single PDU payload for saving the overhead, Generic MAC Header, CRC, and etc. Figure 2-19 illustrates packing process.

Both processes may be enabled by either a BS for a downlink connection or a MS for an uplink connection.

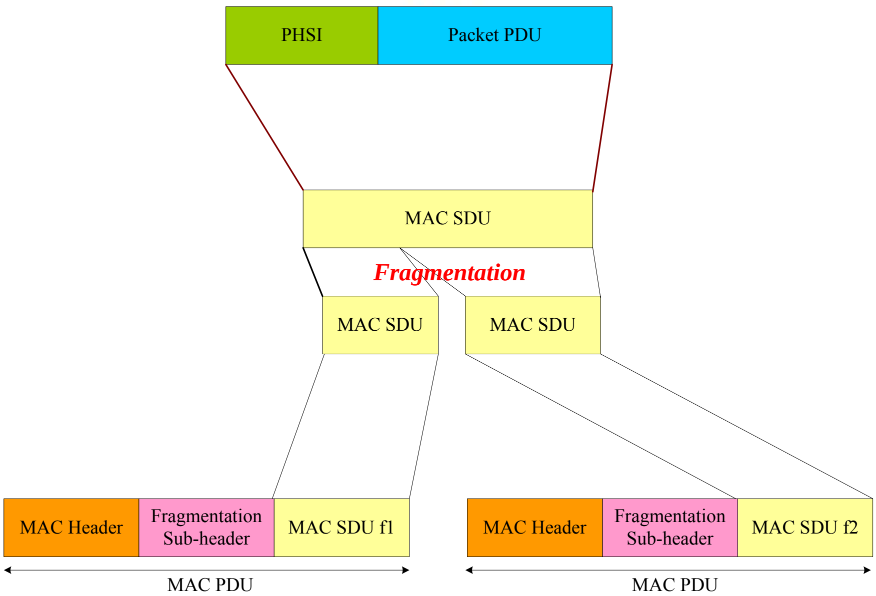

Figure 2-18 Fragmentation

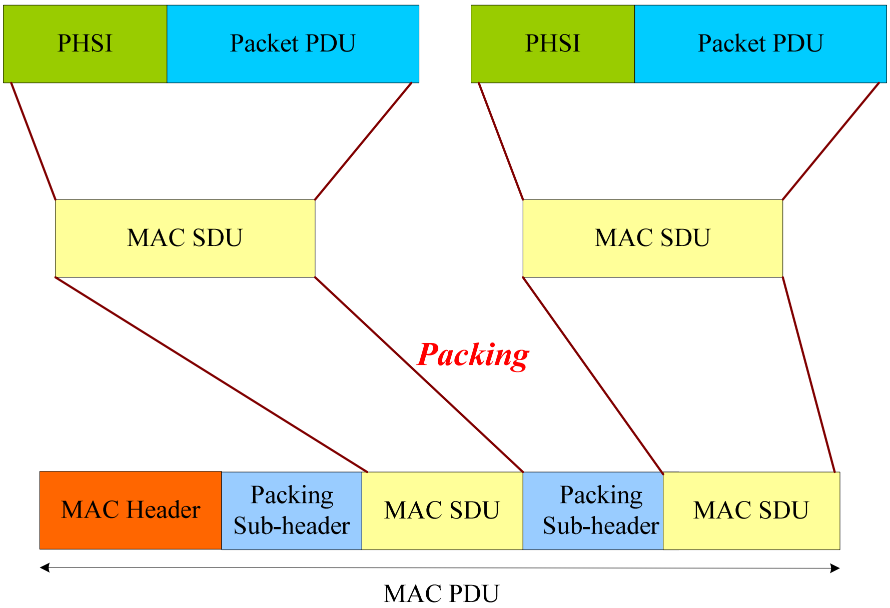

Figure 2-19 Packing

2.2.4 QoS based service classes

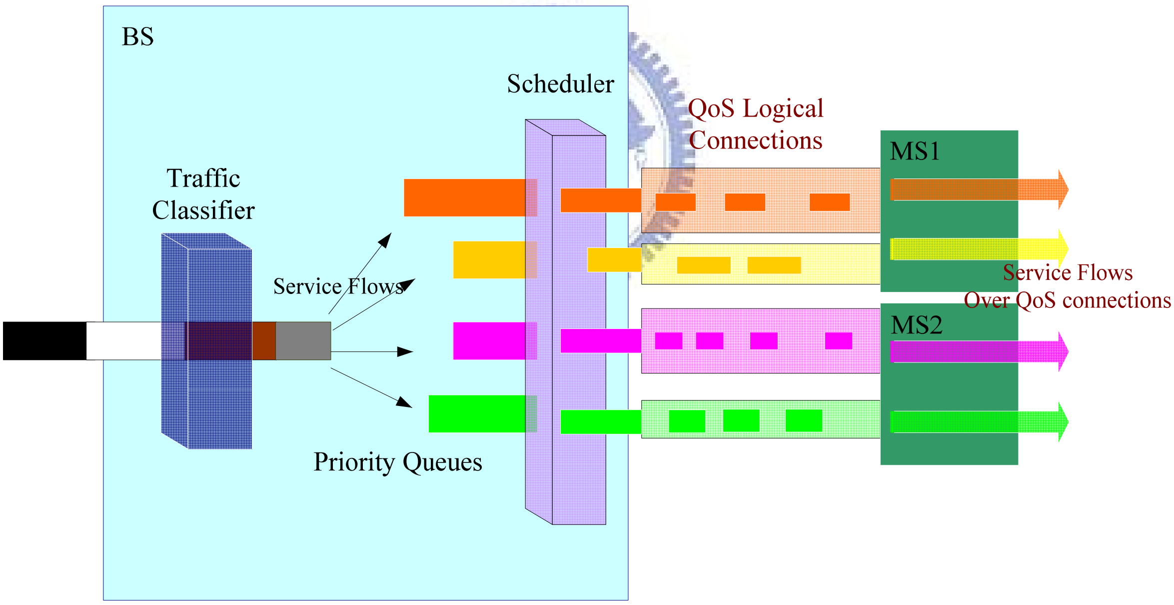

Figure 2-20 QoS Scheduler

In the Mobile WiMAX MAC layer, QoS is provided via service flows as illustrated in Figure 2-20. A service flow of packets is provided with a particular set of QoS parameters. Before identifying a certain type of data service, the base station and user-terminal first establish a logical link between the peer MACs called a connection, and a connection ID (CID) is given to each connections. Each MS can have more than one connection. The scheduler of MAC Layer then associates packets traversing the MAC interface into a service flow to be delivered over the connection. Based on scheduling algorithms, the QoS parameters associated with the service flow is the key for the determination of the transmission ordering and scheduling on the air interface. Therefore, the connection-oriented QoS can provide accurate control over the air interface based on their QoS parameters. Since the air interface is usually the bottleneck, the connection-oriented QoS can effectively enable the end-to-end QoS control. The service flow parameters can be dynamically managed through MAC messages to accommodate the dynamic service demand. The service flow based QoS mechanism applies to both DL and UL to provide improved QoS in both directions. Mobile WiMAX supports a wide range of data services and applications with varied QoS requirements as described in Table 2-4.

Table 2-4 Mobile WiMAX Applications and Quality of Service

QoS Class	*Applications*	*QoS Specifications*
UGS UnSolicited Grant Service	**VoIP**	**Maximum sustained rate Maximum latency tolerance Jitter tolerance**
rtPS Real-Time Polling Service	**Streaming Audio, Video**	**Minimum Reserved Rate Maximum Sustained Rate Maximum Latency Tolerance Traffic Priority**
ErtPS Extend Real-Time Polling Service	**Voice with Activity Detection, Video**	**Minimum Reserved Rate Maximum Sustained Rate Maximum Latency Tolerance Jitter Tolerance Traffic Priority**
nrtPS Non-Real-Time Polling Service	**FTP, HTTP**	**Minimum Reserved Rate Maximum Sustained Rate Traffic Priority**
BE Best Effort Service	**Data Transfer, Web Browsing**	**Maximum Sustained Rate Traffic Priority**

2.2.5 Network Entry

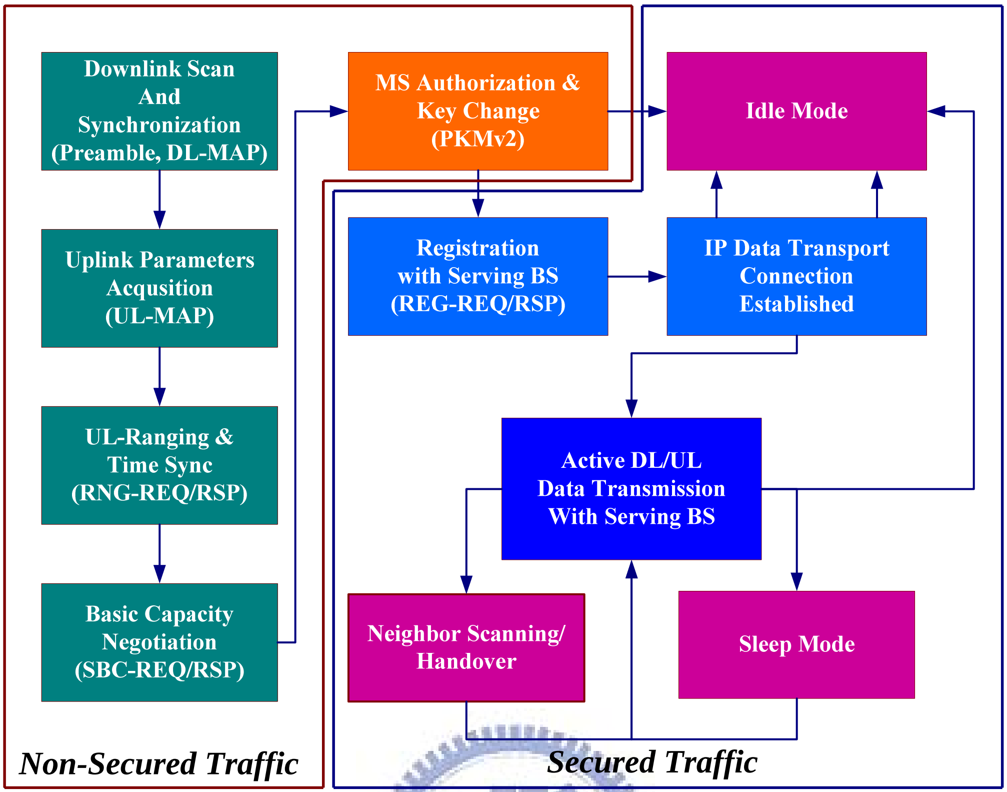

Figure 2-21 Flow Chart of Network Entry

Network entry process is executed while a MS enters the network, or a handoff event occurs. It's being divided into two main parts which are non-secured traffic and secured traffic. As shown in Figure 2-21, downlink synchronization is executed first to synchronize with downlink sub-frame for acquiring Frame Control Header (FCH) and DL-MAP. While DL-MAP is acquired, the location information of UL-MAP is available in DL-MAP. And uplink parameters are available in UL-MAP. After that, UL-Ranging for time, frequency, and power adjustment is executed through RNG-REQ/RSP message exchanging. Basic capacity negotiation follows with ranging through SBC-REQ/RSP message exchanging. After that, MS authorization and key exchange for security is executed to secure user's traffic on the air link. Registration to the serving BS is executed while authorization process is complete. The last step of network entry is to establish IP transport connection for the upper layer communication with other networks.

2.2.6 Channel condition feedback

Due to the mobility, channel condition feedback is critical in Mobile WiMAX. The channel condition may change rapidly in moving situation. Therefore, channel condition information of a MS is necessary for determine the suitable modulation and coding scheme to sustain QoS quality and utilize RF resources efficiently.

CINR is used as an indicator for channel quality. BS and MS may measure the CINR to get the channel condition information and send it back. In distributed sub-carrier permutation mode, to get downlink channel quality, MS is responsible to measure the preamble or a permutation zone and get CINR value. Then MS shall send REP-RSP message to BS to report the measured CINR. Based on the feedback information, BS can determine the suitable AMC in the next frame. REP-RSP message might be sent in response to REP-REQ message from BS or in an unsolicited fashion. The REP-RSP message could be sent through CQICH. The location of CQICH is different from MSs and is assigned by BS in uplink sub-frame. As to Band-AMC mode, MS will also measure CINR on different band and report the message through CQICH. While the information of uplink channel condition is needed by serving BS, MS might use UL-Sounding Zone to transmit data for BS to execute CINR estimation.

2.2.7 Handoff

There are three handoff methods supported within the 802.16e standard – Hard Handoff (HHO), Fast Base Station Switching (FBSS) and Macro Diversity Handover (MDHO). Among these, the HHO is mandatory while FBSS and MDHO are two optional modes. For meeting the QoS requirement of delay-sensitive and high data rate applications, such as video streaming, techniques for optimizing hard handoff within the framework of the 802.16e standard are needed to be developed to achieve the goal of keeping Layer 2 handoff delays to less than 50 milliseconds. The following is the brief introduction to three different handoff method supported in the 802.16e standard.

- **Hard Handoff**

During hard handover the MS communicates with only just one BS in each time. Connection with the old BS is broken before the new connection is established. Handover occurs while the difference between the signal strength measured from neighbor's cell and the signal strength measured from the current cell is exceeding a certain threshold value for a threshold time duration. As shown in Figure 2-22, in brief, the black thick line at the boarder of the cells presents the place where the hard handover is realized.

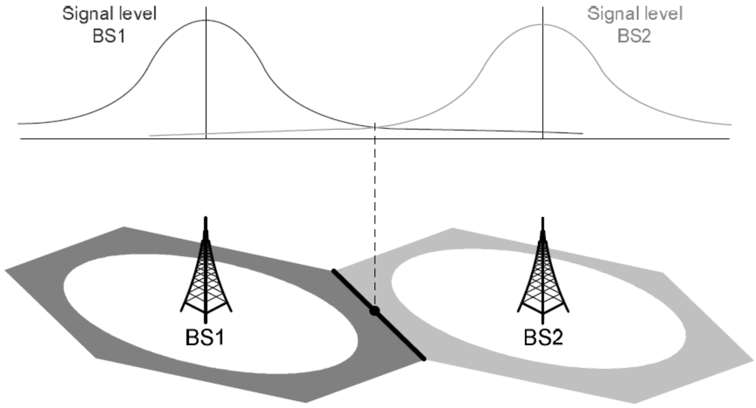

Figure 2-22 Hard Handoff

- **Fast Base Station Switching (FBSS)**

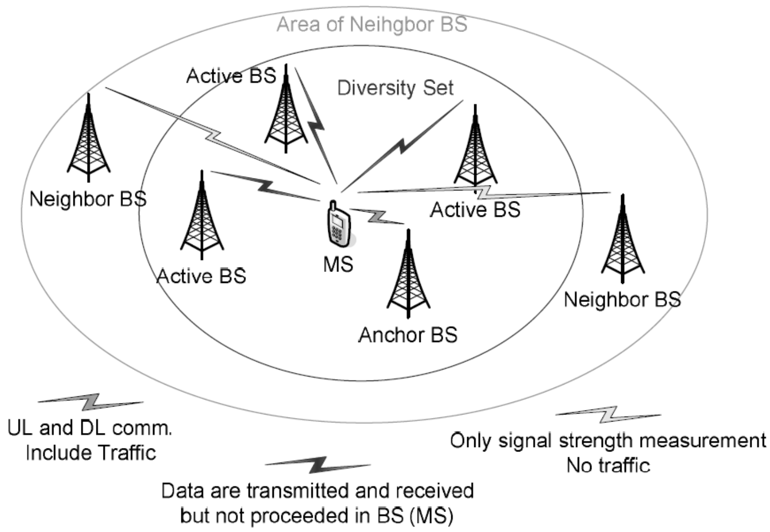

Figure 2-23 Fast Base Station Switching

In Fast Base Station Switching (FBSS) shown in Figure 2-23, the MS and BS maintain a list of BSs that are involved with the MS. This set is called a diversity set. Among the BSs in the Diversity set, an Anchor BS is defined, and it plays the role as the serving BS in Hard Handoff. In FBSS, the MS continuously monitors the signal strength of the BSs that are in the diversity set and selects one BS from the set to be the Anchor BS. When operating in FBSS, the MS only communicates with the Anchor BS for uplink and downlink messages including management and traffic connections. Anchor update and diversity set update are two main procedures while operating in FBSS. Compared to Hard Handoff, anchor update procedure performs the significant reduction of HO time which represents the transition time from one Anchor BS to another. It is done by backhaul communication between the current anchor BS and the candidate BS in the diversity set without MSs' real transmission of explicit HO signaling messages, so the less-than 50 millisecond HO time is acheived. In diversity set update procedure, the MS scans the neighbor BSs and selects those that are suitable to be included in the diversity set in terms of signal strength, Traffic condition, and etc. The MS reports the selected BSs, and the diversity set update procedure is performed by the BS and MS. An important requirement of FBSS is that the data is simultaneously transmitted to all members of a diversity set of BSs that are able to serve the MS.

➢ **Macro Diversity Handover (MDHO)**

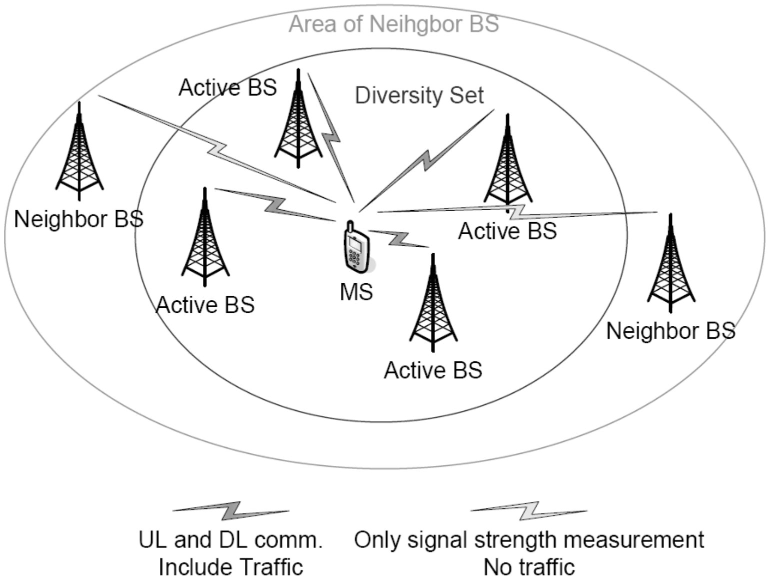

Figure 2-24 Macro Diversity Handover (MDHO)

The basic concept and advantages of Macro Diversity Handover is the same as Soft Handoff mechanism in CDMA2000 or UMTS. From a system-wise point of view, it utilizes more air-link resources from more than one BS to considerably reduce HO time and perform the diversity gain to increase the performance. In Macro Diversity Handover, same as FBSS, a diversity set of BSs and an anchor BS in the diversity set are defined as well. When operating in MDHO, the MS communicates with all BSs in the active set of uplink and downlink unicast messages and traffic meanwhile as shown in Figure 2-24. An anchor BS in the active set not only transmits data traffic, but also takes charge of transmitting control information and DL broadcast message (DL-MAP, UL-MAP, etc) to the MS. For downlink MDHO, two or more BSs provide synchronized transmission of MS downlink data such that diversity combining is performed at the MS. For uplink MDHO, the transmitted data from a MS is received by multiple BSs where selection diversity of the information received is performed. The frame synchronization and same frequency assignment to all members in the diversity set are also the requirement to enable MDHO.

Chapter 3 Bandwidth Request Mechanism and Scheduling algorithm of Uplink Transmission in Mobile WiMAX

3.1 Bandwidth Request and Grant Mechanism

Request Mechanism

Requests refer to the mechanism that MSs use to indicate to the BS that they need UL bandwidth allocation. A Request may come as a BR header or it may come as a PiggyBack Request. The capability of Piggyback Request is optional. Because the UL burst profile can change dynamically, all requests for bandwidth shall be made in terms of the number of bytes needed to carry the MAC PDU excluding PHY overhead. The BR message may be transmitted during any UL allocation, except during any initial ranging interval. BRs may be incremental or aggregate. When the BS receives an incremental BR, it shall add the quantity of bandwidth requested to its current perception of the bandwidth needs of the connection. When the BS receives an aggregate BR, it shall replace its perception of the bandwidth needs of the connection with the quantity of bandwidth requested. The message flow of request and grant is shown in Figure 3-1.

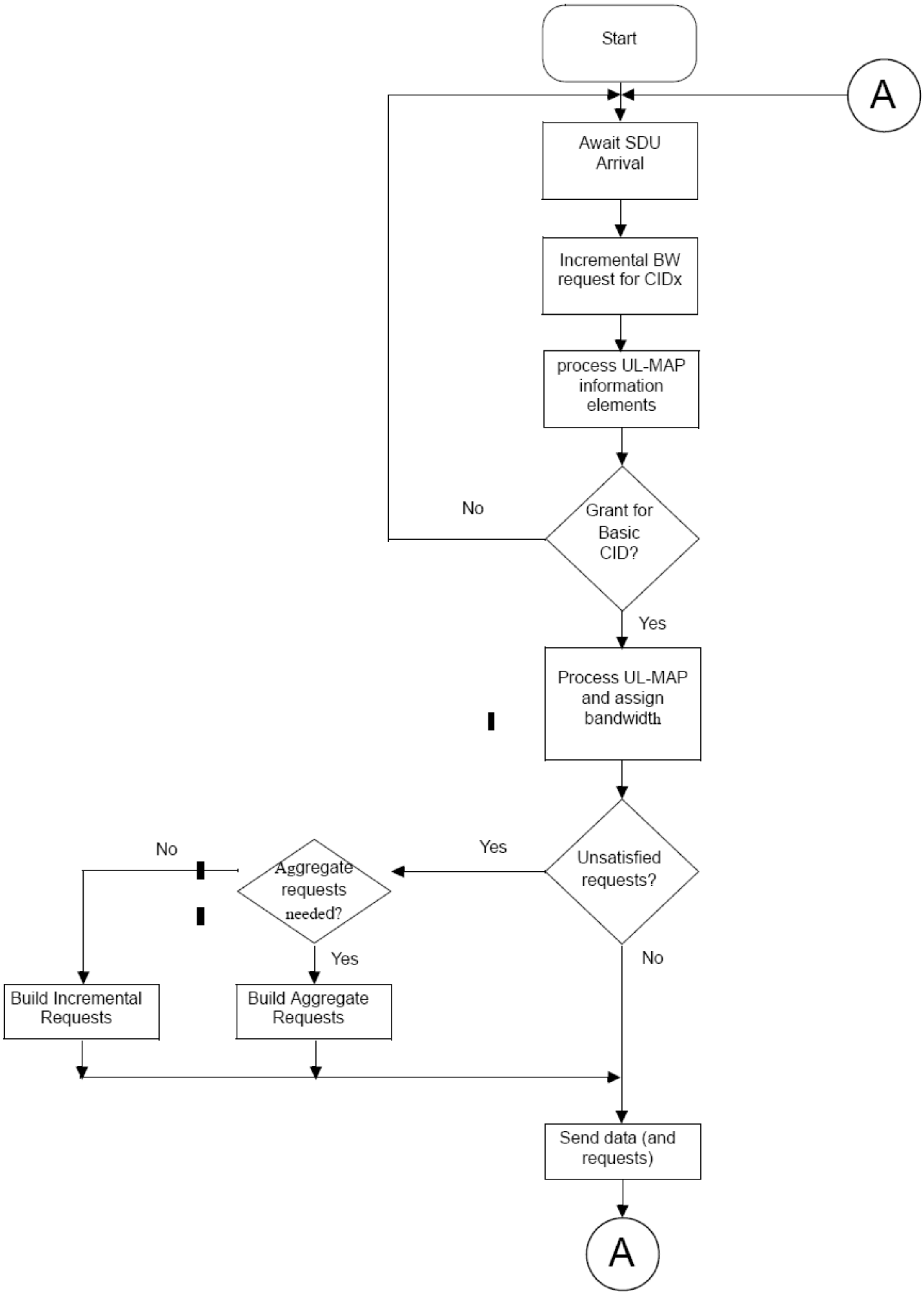

Figure 3-1 Request and Grant flow chart

Polling Mechanism

Polling is the process by which the BS allocates bandwidth to the MSs specifically for the purpose of making BRs. These allocations may be to individual MSs or to groups of MSs. Allocations to groups of connections and/or MSs actually define BR Contention IEs. The allocations are contained as a series of IEs within the UL-MAP. Note that polling is done on MS basis. Bandwidth is always requested on a CID basis and bandwidth is allocated on an MS basis.

Unicast Polling

When a MS is polled individually, no explicit message is transmitted to poll the SS. Instead, the MS is allocated, in the UL-MAP, bandwidth which is sufficient to respond with a BR. MSs that have an active UGS connection of sufficient bandwidth shall not be polled individually unless they set the PM bit in the header of a packet on the UGS connection. The message flow of unicast polling is shown in Figure 3-2.

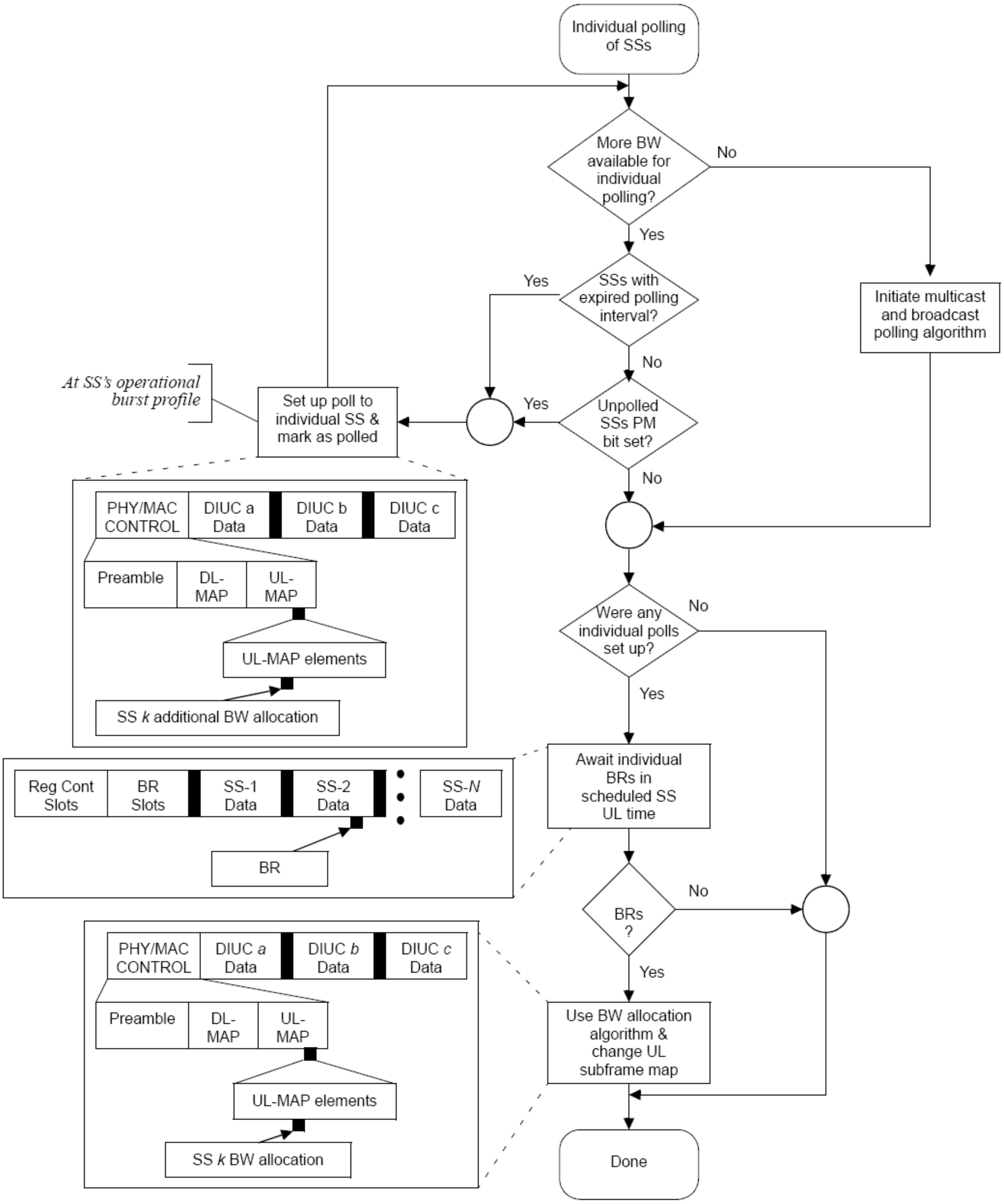

Figure 3-2 Unicast Polling

Multicast and Broadcast Polling

If insufficient bandwidth is available to individually poll many inactive MSs, some MSs may be polled in multicast groups or a broadcast poll may be issued. Certain CIDs are reserved for multicast groups and for broadcast messages. As with individual polling, the poll is not an explicit message, but bandwidth allocated in the UL-MAP. The difference is that, rather than associating allocated bandwidth with an MS's Basic CID, the allocation is to a multicast or Broadcast CID.

When the poll is a multicast or Broadcast CID, an MS belonging to the polled group may request bandwidth during any request interval allocated to that CID in the UL-MAP by a Request IE. They shall apply the contention resolution algorithm to select the slot in which to transmit the initial BR. The message flow of multicast and broadcast polling is shown in Figure 3-3.

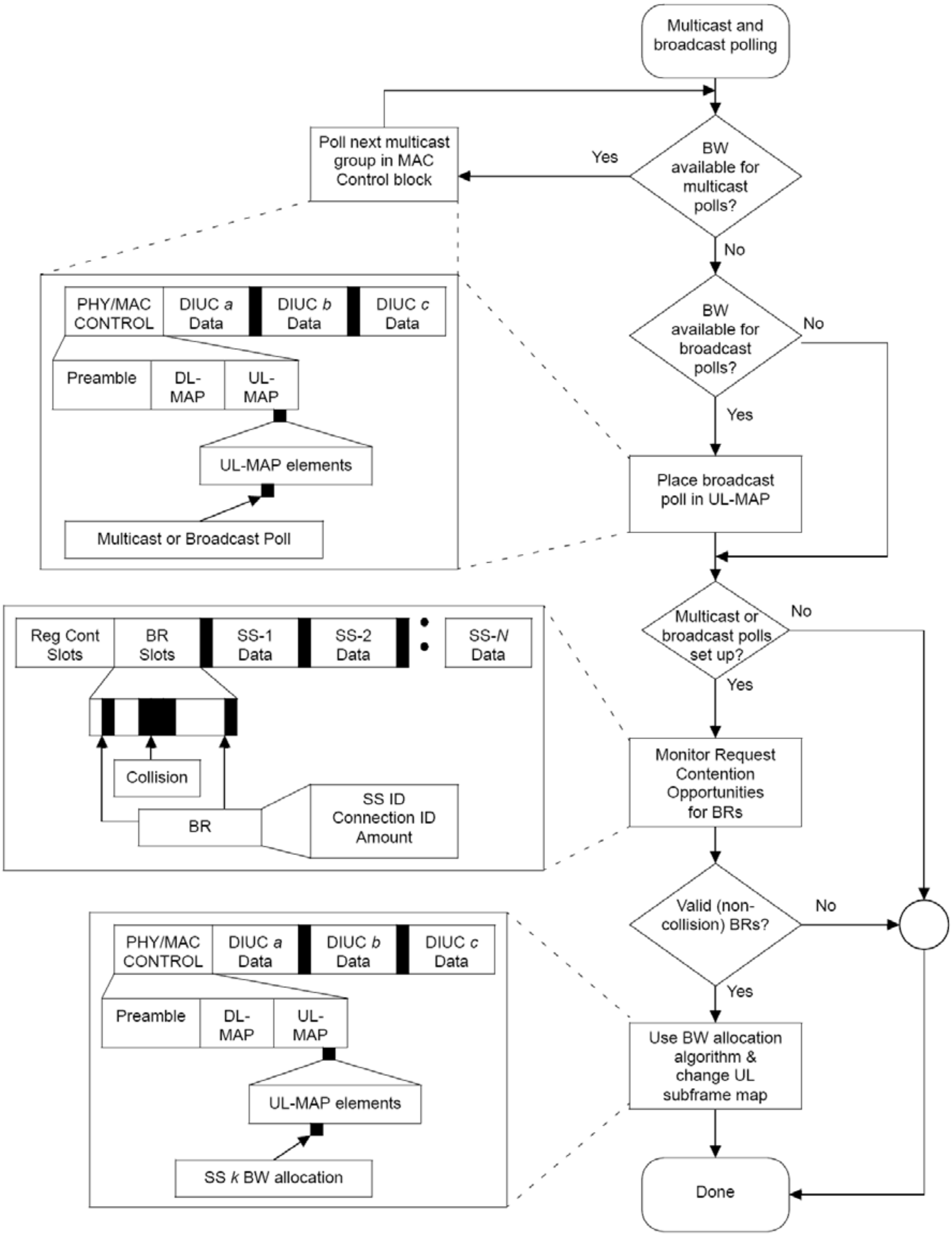

Figure 3-3 Multicast and broadcast polling

PM bit

MSs with currently active UGS connections may set the PM bit in a MAC packet of the UGS connection to indicate to the BS that they need to be polled to request bandwidth for non-UGS connections. To reduce the bandwidth requirements of individual polling, MSs with active UGS connections need be individually polled only if the PM bit is set. Once the BS detects this request for polling, the process for individual polling is triggered. The procedure is shown in Figure 3-4.

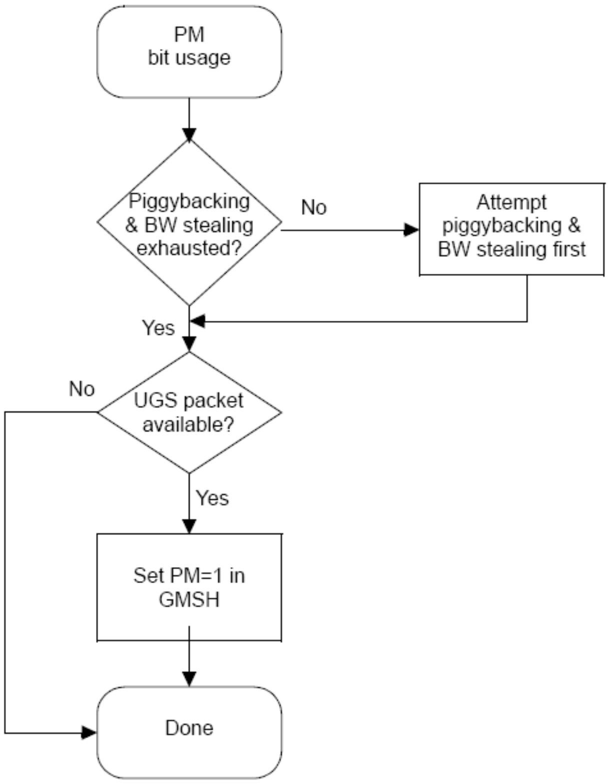

Figure 3-4 PM bit usage

Contention-based focused BRs for WirelessMAN-OFDMA

The OFDMA-based PHY specifies a ranging sub-channel and a subset of ranging codes that shall be used for contention-based BRs. The MS needing to request bandwidth shall select, with equal probability, a ranging code from the code subset allocated to BRs. This ranging code shall be modulated onto the ranging sub-channel and transmitted during a Ranging Slot randomly selected from the appropriate ranging region in a single frame. Upon detection, the BS shall provide UL allocation for the SS, and the Broadcast CID shall be sent in combination with a CDMA Allocation IE, which specifies the UL region and ranging code that were used by that specific MS. Thus, the selected CDMA code and the broadcast CID allow a MS to determine whether it has been given an allocation by matching these parameters with itself. The MS use the allocation to transmit a BR PDU. If the BS does not issue the UL allocation described above, or the BR MPDU does not result in a subsequent allocation of any bandwidth, the SS shall assume that the ranging code transmission resulted in a collision and follow the contention resolution

Contention Resolution Algorithm

The mandatory method of contention resolution that shall be supported is based on a truncated binary exponential backoff, with the initial backoff window and the maximum backoff window controlled by the BS. The values are specified as part of the UCD message and represent a power-of-two value. For example, a value of 4 indicates a window between 0 and 15; a value of 10 indicates a window between 0 and 1023.

When a MS has information to send and wants to enter the contention resolution process, it sets its internal backoff window equal to the request backoff start defined in the UCD message.

The MS shall randomly select a number within its backoff window. This random value indicates the number of contention transmission opportunities that the MS shall defer before transmitting. A MS shall consider only contention transmission opportunities for which this transmission would have been eligible. These are defined by Request IEs in the UL-MAP messages. Note that each IE may consist of multiple contention transmission opportunities.

Using BRs as an example, consider that an SS whose initial backoff window is 0 to

15 and assume it randomly selects the number 11. The SS must defer a total of 11 contention transmission opportunities. If the first available Request IE is for 6 requests, the MS does not use this and has 5 more opportunities to defer. If the next Request IE is for 2 requests, the MS has 3 more to defer. If the third Request IE is for 8 requests, the MS transmits on the fourth opportunity, after deferring for 3 more opportunities.

After a contention transmission, the SS waits for a Data Grant Burst Type IE in a subsequent UL-MAP. Once received, the contention resolution is complete. The MS shall consider the contention transmission lost if no data grant has been received in the number of subsequent UL-MAP messages specified by the Contention-Based Reservation Timeout parameter. The MS shall now increase its backoff window by a factor of two, as long as it is less than the maximum backoff window. The SS shall randomly select a number within its new backoff window and repeat the deferring process described above.

For BRs, if the MS receives a unicast Request IE or Data Grant Burst Type IE at any time while deferring for this CID, it shall stop the contention resolution process and use the explicit transmission opportunity.

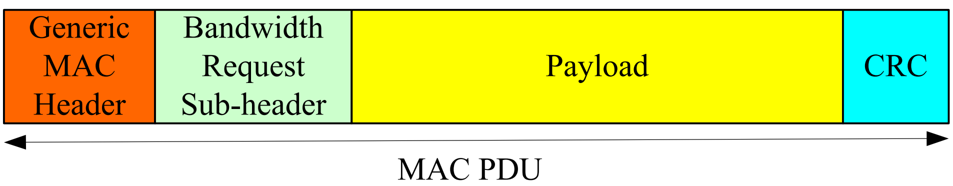

Figure 3-5 MAC PDU with BR Sub-header

For uplink transmission, the transmission opportunity of a mobile subscriber (MS) is assigned by Serving BS based on the bandwidth request and grant mechanism defined in the IEEE 802.16e-2005 standard. For some certain service flows, such as UGS, rtPS, nrtPS, and ertPS, Bandwidth Request Message is usually done by the insertion of Bandwidth Request Header (BRH) following Generic MAC Header (GMH), as shown in Figure 3-5. In order to assure the correct reception of Bandwidth Request Message in BS site, the most robust modulation and coding scheme, BPSK+CC 1/2, is applied for BR Messages no matter how good the channel condition is in the present simulation drop. In the simulation platform, VoIP and FTP traffic models are built to represent UGS and nrtPS respectively. For Unsolicited Grant Service (UGS), transmission opportunity is periodically granted in an unsolicited manner to accommodate the QoS quality, while nrtPS only offers unicast polls on a regular basis, which assures that the UL service flow in BS site receives request opportunities even during network congestion. Figure 3-6 and Table 3-1 shows the detail description of Bandwidth Request MAC Header.

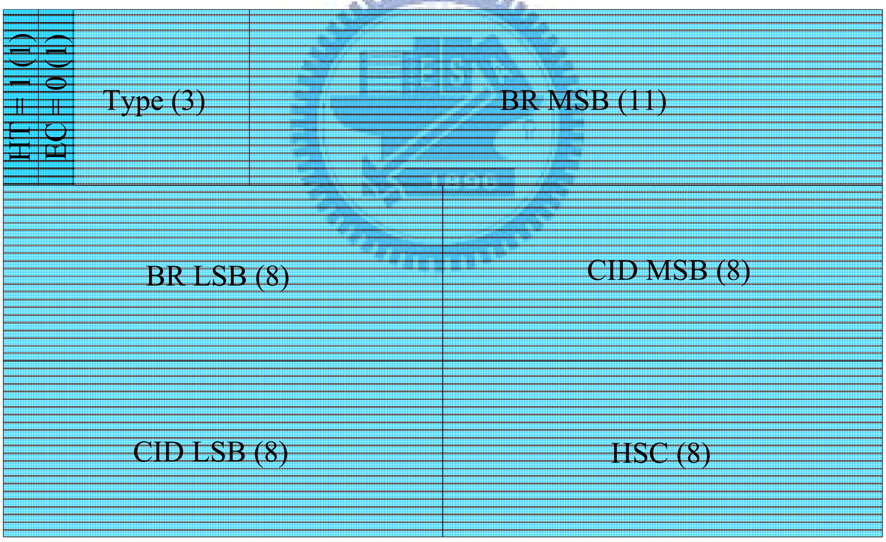

Figure 3-6 Bandwidth Request MAC Header

Table 3-1 Description of Bandwidth Request Header

Name	Size	Description
Type	3 bits	The type of BR and UL Tx Power Report header
BR	11 bits	Bandwidth Request: The number of bytes of uplink bandwidth requested by the MS. The bandwidth request is for the CID. The request shall not include any PHY overhead. It is incremental BW request. In case of the Extended rtPS, if the MSB is 1, the

		BS changes its polling size into the size specified in the LSBs of this field.
CID	8 bits	The CID shall indicate the connection for which uplink bandwidth is requested.
HCS	8 bits	Header Check Sequence

3.2 Uplink Scheduling Types and QoS services

Uplink QoS Classes

- **UGS (Unsolicited Grant Service)**

  The UGS is designed to support real-time UL service flows that transport fixed-size data packets on a periodic basis, such as T1/E1 and Voice over IP without silence suppression. The service offers fixed-size grants on a real-time periodic basis, which eliminate the overhead and latency of MS requests and assure that grants are available to meet the flow's real-time needs. The BS shall provide Data Grant Burst IEs to the MS at periodic intervals based upon the Maximum Sustained Traffic Rate of the service flow. The size of these grants shall be sufficient to hold the fixed-length data associated with the service flow (with associated generic MAC header and Grant management sub-header).

- **rtPS (real-time Polling Service)**

  The rtPS is designed to support real-time UL service flows that transport variable-size data packets on a periodic basis, such as moving pictures experts group (MPEG) video. The service offers real-time, periodic, unicast request opportunities, which meet the flow's real-time needs and allow the SS to specify the size of the desired grant. This service requires more request overhead than UGS, but supports variable grant sizes for optimum data transport efficiency.

- **ErtPS (extended real-time Polling Service)**

  Extended rtPS is a scheduling mechanism which builds on the efficiency of both UGS and rtPS. The BS shall provide unicast grants in an unsolicited manner like in UGS, thus saving the latency of a BR. However, whereas UGS allocations are fixed in size, ertPS allocations are dynamic.

The Extended rtPS is designed to support real-time service flows that generate variable-size data packets on a periodic basis, such as Voice over IP service with silence suppression.

- **nrtPS (non-real-time Polling Service)**

The nrtPS offers unicast polls on a regular basis, which assures that the UL service flow receives request opportunities even during network congestion. The BS typically polls nrtPS connections on an interval on the order of one second or less. The BS shall provide timely unicast request opportunities. In order for this service to work correctly, the Request/Transmission Policy shall be set properly. The mandatory QoS parameters for this scheduling service are Minimum Reserved Traffic Rate, Maximum Sustained Traffic Rate, Traffic Priority, Uplink Grant Scheduling Type, and Request/Transmission Policy.

- **BE (Best Effort Service)**

The intent of the BE grant scheduling type is to provide efficient service for BE traffic in the UL. In order for this service to work correctly, the Request/Transmission Policy setting shall be set so that the MS is allowed to use contention request opportunities. This results in the MS using contention request opportunities as well as unicast request opportunities and data transmission opportunities.

3.3 Implementation of Bandwidth Request and Grant Mechanism

Framework of Uplink Data Transmission

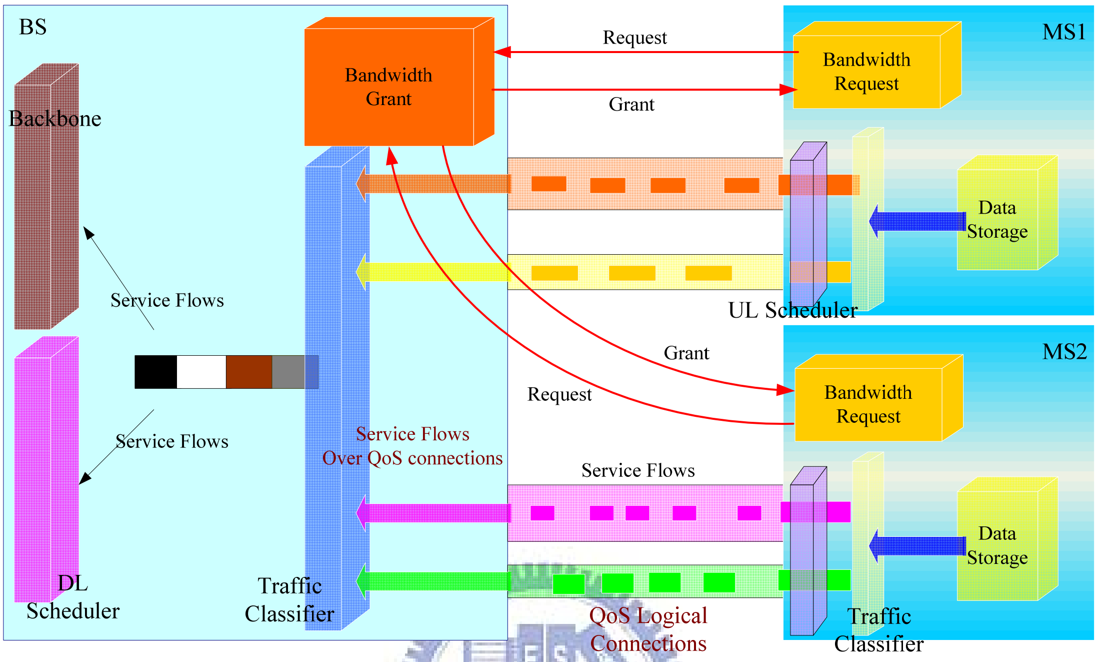

Figure 3-7 UL Service Flow Scheduler

For uplink service flow scheduler, it's similar to downlink scheduler in aspects of categories of service flows and QoS-based connections. However, while data traffic of mobile subscribers is transmitted by BS site via downlink resource allocation, MSs have to request Serving BS for transmission opportunity in uplink sub-frame via various bandwidth request methods as shown in Figure 3-7, which includes unicast polling service, contention based BR transmission opportunity, non-contention based BR transmission opportunity, unsolicited polling service, and unsolicited grant service. After negotiation with BS site, MS might either acquire desired bandwidth request transmission opportunity for requesting data transmission opportunity in UL sub-frame, or acquire data transmission opportunity in UL sub-frame directly. In this simulation platform, for the simplicity, only unicast polling service, unsolicited polling service, and unsolicited grant service are implemented to execute the attempted simulation of only VoIP and FTP traffic. QoS is provided via service flows as illustrated in Figure 3-7. An uplink service flow of packets is provided with a particular set of QoS parameters. Before identifying a certain type of data service, the base station and user-terminal first establish a logical link between the peer MACs called a connection, and a connection ID (CID) is assigned to each connection by MSs. Every MS might have more than one connection. The scheduler of MAC Layer then associates packets traversing the MAC interface into a service flow to be delivered over the connection. Based on scheduling algorithms, the QoS parameters associated with the service flow is the key for the determination of the transmission ordering and scheduling on the air interface. Therefore, the connection-oriented QoS can provide accurate control and maintain good QoS quality over the air interface based on their QoS parameters.

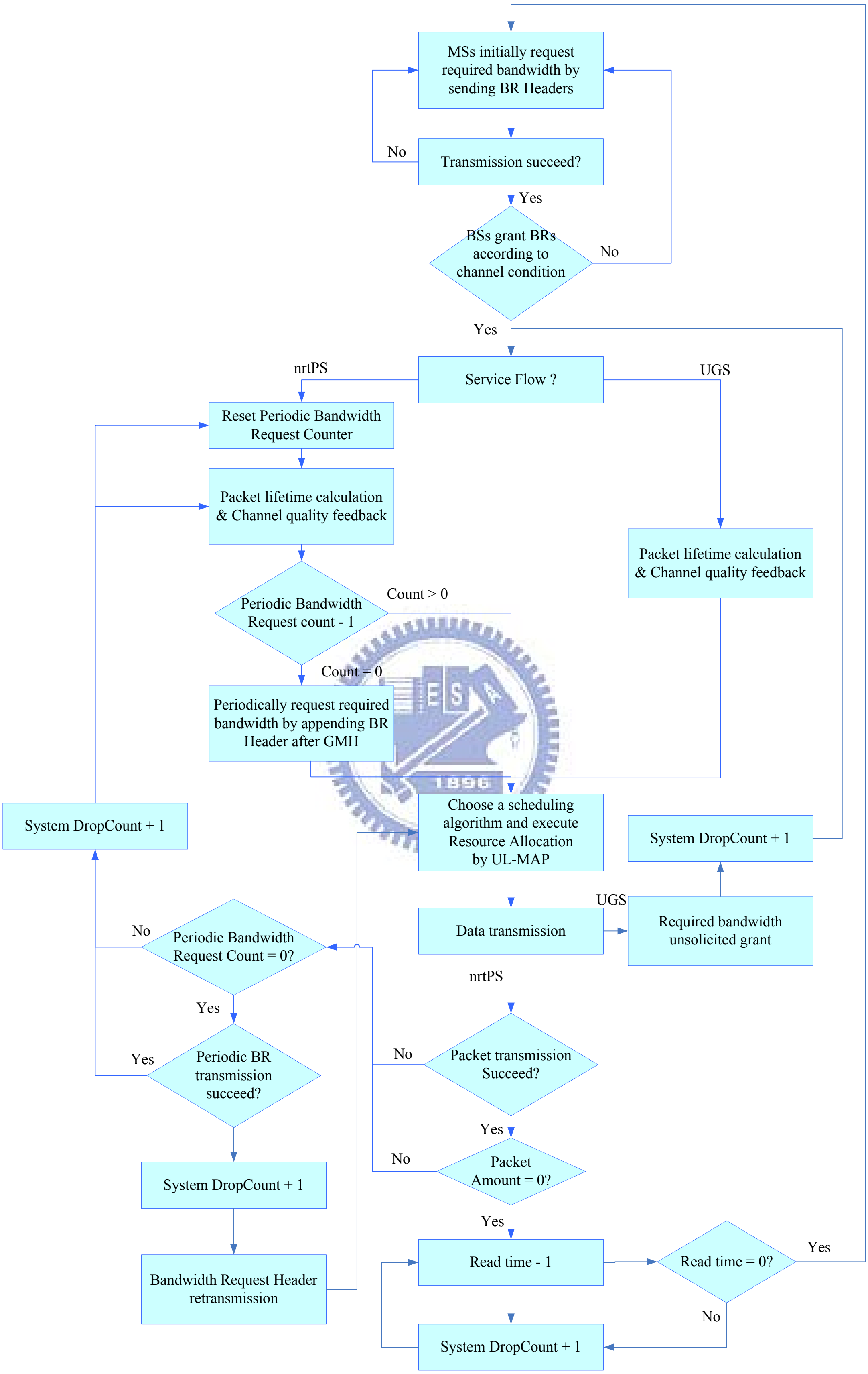

Figure 3-8 Flow Chart of bandwidth request and grant mechanism in VoIP and FTP traffic

Figure 3-8 shows the flow chart of bandwidth request and grant mechanism, which illustrates how the simulation of uplink scheduling was implemented step by step. Firstly, MSs which wants to initiate data traffic shall negotiate data transmission opportunity with Serving BS through sending initial BR Message. Until initial BR Message is successfully received by BS, data transmission opportunities will be assigned based upon resource allocation managing and ordering done by BS site according to the associated values of QoS control, such like signal strength or packet urgency.

In addition to initial BR Messages sent in the beginning of data service, FTP traffic, which belongs to nrtPS, shall send BR Message periodically by adding BR sub-header right after GMH to continuously request data transmission opportunity. MSs in FTP service might indicate the possible change of FTP packet size in the periodic BR Message to fulfill user's demand as well.

For MSs in VoIP traffic, without the transmission of BR Message like FTP service, data transmission opportunity is periodically given by Serving BS in an unsolicited manner. MSs shall report channel quality in a periodic basis. Also, the packet lifetime is calculated by BS site in every simulation drops. The received channel quality indicator and calculated packet lifetime shall be used for the purpose of managing the ordering of uplink transmission.

For uplink scheduling, QoS control parameters are calculated in every simulation drop, and used as the indicator for BS site to managing the order of data transmission opportunity in uplink sub-frame according to different scheduling algorithms, such as Round Robin, Early Deadline First, and etc. Serving BS indicates the resource allocation of uplink sub-frame in UL-MAP located in downlink sub-frame in the next OFDMA frame, and MSs shall get the information of uplink resource allocation by decoding UL-MAP in downlink sub-frame.

In the simulation platform, ARQ mechanism for BR Message is implemented to simulate its possible error caused by bad channel condition or traffic congestion. Users are not allowed to transmit any data if the transmission error of BR Message occurs, and need to retransmit BR Message until transmission succeeds.

3.4 QoS Parameters Calculation

As to real-time service, the QoS definition is the delay concerns. For each packet, it has its own delay bound and must be transmitted before the bound or the packet will be dropped. As to non-real-time service, the minimum reserved rate is set instead of packet delay bound.

In the simulation platform, due to the QoS control of VoIP traffic is packet delay bound, for the simplicity of uplink scheduler design, a so-called "soft delay bound' of non-real-time service (FTP) is being used. Serving BS can calculate the soft delay bound of each FTP packet based upon the minimum reserved rate. In both VoIP and FTP traffic, it is simpler to implement the scheduler design on the same time domain basis.

$$soft\ delay\ bound = \frac{packet\ size}{minimum\ reserved\ traffic\ rate} \quad (9)$$

$$\begin{aligned} soft\ delay\ bound &\le \frac{(P_1 + P_2 + \cdots + P_n)}{minimum\ reserved\ rate} \\ &= \frac{P_1}{minimum\ reserved\ rate} + \frac{P_2}{minimum\ reserved\ rate} + \cdots + \frac{P_n}{minimum\ reserved\ rate} \end{aligned} \quad (10)$$

The soft delay bound (9) indicates the minimum performance non-real-time service shall achieve. FTP packets accommodate the minimum reserved rate, while they're received by BS within soft delay bound.

Besides the drop and transmit issue, the soft delay bound of non-real-time service will be accumulated across the packets belonged to the same user. It can be expressed in the following formula, P denotes the packet size:

$$\frac{(P_1+P_2+\cdots+P_n)}{t} \geq minimum\ reserved\ rate \qquad (11)$$

The soft delay bound will accumulate across packets. If the first packet is transmitted earlier before its soft delay bound, the second packet of the user will have longer soft delay bound due to first packet’s fast transmission. However, while the packet doesn’t meet the soft delay bound constraint, the next packet will have shorter soft delay bound than normal due to the delay of the last packet.

Packet Urgency Calculation

$$num\_frame = \left\lceil \frac{life\_time}{frame\_length} \right\rceil \qquad (12)$$

$$transmit\_frame = num\_frame - \max(ARQretransmission) \qquad (13)$$

$$t_{availble} = UL\_subframe \times transmit\_frame \qquad (14)$$

$$\begin{aligned}
& if\left(\left\lceil \frac{packet\ size}{\min(bits\ per\ resource\ unit)} \right\rceil \leq num_of_ULtimeslots\right) \\
& \left\{ t_d = \left\lceil \frac{packet\ size}{\min(bits\ per\ resource\ unit) \times num_of_ULtimeslots} \right\rceil \times UL_subframe \right\} \qquad (15)
\end{aligned}$$

$$else$$

$$\left\{ \begin{array}{l}
if\left(\left\lceil \dfrac{packet\ size}{\min(bits\ per\ resource\ unit)} \right\rceil > num_of_ULtimeslots \times num_of_ULsubchannels\right) \\
\left\{ \begin{array}{l}
num_of_UL_subframe = \dfrac{\left\lceil \dfrac{packet\ size}{\min(bits\ per\ resource\ unit)} \right\rceil}{\left(num_of_ULtimeslots \times num_of_ULsubchannels\right)} \\
remainder = \left\lceil \dfrac{packet\ size}{\min(bits\ per\ resource\ unit)} \right\rceil \bmod \left(num_of_ULtimeslots \times num_of_ULsubchannels\right) \\
if\left(remainder \geq 6\right) \\
\left\{num_of_timeslot = 6\right\} \\
t_d = num_of_UL_subframe \times UL_subframe + \dfrac{num_of_timeslot}{num_of_ULtimeslots} \times UL_subframe
\end{array} \right\} \\
else \\
\left\{t_d = UL_subframe\right\}
\end{array} \right\} \qquad (15)$$

$$t_{edf} = t_{available} - t_d \qquad (16)$$

$$t = t_{edf} - UL\_subframe \qquad (17)$$

$$if\ t \leq 0,\ this\ service\ flow\ must\ be\ transmitted\ in\ this\ frame$$

Firstly, the number of frames the packet can wait within its delay bound is calculated in formula (12). Life time represents the rest time duration of the delay bound. Then, certain amount of frames is preserved for ARQ retransmission, described in (13). In (14), the available time for transmission is calculated. The time duration required to transmit the packet is calculated based upon required number of resource units in (15). According to the uplink resource allocation method, the associated number of uplink sub-frames needed to transmit the packet is calculated. In (16), using the value got in (15), t_{edf}, to get how many time does the system can wait for transmission without violating QoS requirements. Finally, the indicator, t, which represents the urgency of that packet, is calculated by t_{edf} minus the time duration of uplink sub-frame. If the value t is larger the zero, then it can wait for transmission in the next frame without violating QoS requirements. On the contrary, if the value t is less than zero, the packet is urgent, and supposed to be transmitted at this frame to accommodate QoS requirements.For earliest deadline first (EDF) scheduling algorithm, packets with t less than zero are transmitted first due to the urgency.

However, since packets of non-real-time service will not be dropped even if it exceeds the delay bound, real-time services, such as VoIP, video streaming are scheduled and transmitted first to maintain its strict QoS constraints.

3.5 Scheduling Algorithms

The following introduces some common used scheduling algorithms for uplink transmission. A good scheduling algorithm shall utilize the system resource well and accommodate different users' demand. In the earlier system, there is only voice or data service. It is much easier to design scheduling algorithm. However, nowadays, mixed service types are usually applied in the system evaluation.

In this section, descriptions of some scheduling algorithms proposed by others [8] [9] [10] [11] [12] [13] are given. Different scheduling algorithms are designed for different application scenario. Some of them are designed to achieve fairness, and some of them is capable of maximizing throughput or maintaining QoS quality.

- **Round-Robin (RR)**

  In round-robin scheduling algorithm, packets are scheduled user by user in a fixed ordering irrespective of the channel condition and services requirement. This algorithm provides fairness but ignores the channel condition and QoS requirement. It is hard to be used in mix-traffic system due to different QoS requirement.

- **MaxCINR**

  In MaxCINR scheduling, packets are scheduled based upon how good the channel condition (CINR) is. This kind of scheduling algorithm takes channel condition into consideration and may provide good multi-user diversity to enhance system throughput and shorter transmission time due to efficient transmission. However, it is less fairness. Without a good power control mechanism, some users might never get opportunity to transmit. Besides, it can't provide QoS guarantee, especially for real-time service, which QoS definition is usually delay bound consideration. System selects user m* that fulfill:

$$m_t^* = \arg\max_m CINR_m(t) \qquad (1)$$

- **Proportional Fair (PF)**

Channel condition and fairness are taken into consideration across users. It provides a tradeoff between system throughput and user fairness. The scheduling decision will follow a ratio which is instantaneous data-rate over average data-rate and pick larger user m*:

$$m^* = \arg\max_m \frac{R_m(t)}{S_m(t)} \qquad (2)$$

$$S_m(t) = (1-\frac{1}{W})S_m(t-1) + \frac{1}{W}R_m(t)\delta(t-m) \qquad (3)$$

, where $R_m(t)$ denotes the achievable instantaneous data-rate for user m, $S_m(t)$ denotes the moving average of data-rate at user m, and W donates the length of moving average. This algorithm provides both fairness and channel condition concerns, but it lacks QoS guarantee, especially for real-time service.

- **Modified Largest Weighted Delay First (m-LWDF)**

This algorithm, similar the Proportional Fair, considers fairness and system throughput. More than that, it adds another factor, delay, into consideration. System will pick user m* if

$$m_t^* = \arg\max_m a_m SVP_m(t) R_m(t) / S_m(t) \qquad (4)$$

, while a_m indicates the QoS level for user m, SVP_m indicates the starvation period, meaning the delay for user m. In this way, system can avoid users suffering from long delay.

- **Optimum Channel-Aware Scheduling with Service Differentiation (OCASD)**

This algorithm provides throughput gains while offering fairness and traffic delay constraints. However, this algorithm doesn't provide strict delay constraint which is important for delay-sensitive service. System picks connection x* while

$$x_t^* = \arg\max_x \{ [\left| w_i - \frac{B_i(t)}{B(t)} \right| \cdot B(t)]^{\operatorname{sgn}(w_i - \frac{B_i(t)}{B(t)})} R_m(t) / d_i \} \qquad (5)$$

$$B_i(t+1) = (1 - \frac{1}{W}) B_i(t) + \frac{1}{W} \times L_i(t) \delta(t-i) \qquad (6)$$

, where w_i is the weight depending on the QoS requirements of traffic class i. d_i is the delay bound of the HOL packet which size if $L_i(t)$ in queue i. B(t) is the aggregate of backlogged queues at time t. $R_m(t)$ denotes the achievable instantaneous data-rate for user m.

- **Traffic-Aided Opportunistic Scheduling (TAOS)**

This algorithm utilizes the file size information and user-diversity concept to reduce file receiving time, which is the sum of waiting time and transmission time. System will pick user m* if

$$m_t^* = \arg\min_m S_m(t) F_m / R_m(t) \qquad (7)$$

, where F_m donates the size of the HOL packet in user m's queue. This algorithm doesn't consider delay constraint and fail to ensure the QoS guarantee.

- **Early Deadline First (EDF)**

This algorithm considers merely the delay constraint without taking channel quality into consideration. Therefore, the bandwidth efficiency might not be good. However, strict QoS guarantee are provided for delay-sensitive service by serving urgent packets first. System transmits packet of user m if

$$m^* = \arg\min_m \{ DB - Age - T_t \} \qquad (8)$$

, where DB means delay bound, Age is the time that the packet stayed in MAC, and T_t is the transmission time for this packet in transmitted in the current frame.

Chapter 4 Simulation Setup

4.1 19-Cells Wrap-Around Implementation

Multi-Cell Layout

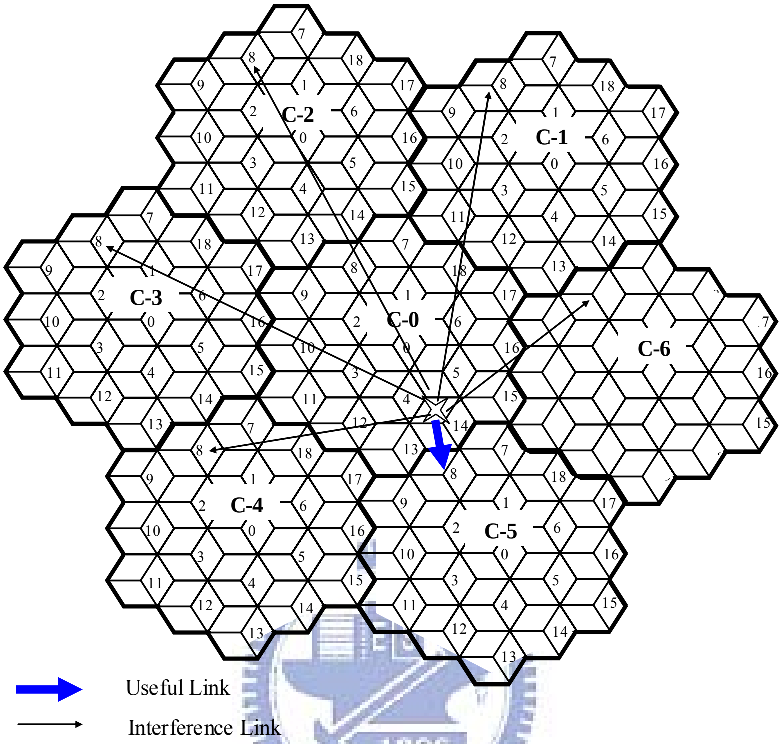

Figure 4-1 Multi-cell Layout and Wrap Around Example

In Figure 4-1, a network of cells is formed with 7 clusters and each cluster consists of 19 cells. For the cases where modeling outer-cells are necessary for accuracy of the results, the wrap around structure with the 7 cluster network can be used. In the wrap around implementation, the network is extended to a cluster of networks consisting of 7 copies of the original hexagonal network, with the original hexagonal network in the middle while the other 6 copies are around it symmetrically on 6 sides, as shown in Figure 4-1. There is a one-to-one mapping between cells/sectors of the center hexagon and cells/sectors of each copy, so that every cell in the extended network is identified with one of the cells in the central hexagonal network. Those corresponding cells have thus the same antenna configuration, traffic, fading etc. except the location as shown in Figure 4-2. The arrows in the Figure 4-2 show the directions that the antennas are pointing.

An example of the antenna orientations in case of a sectorized system is defined in Figure 4-2. The distance from any MS to any base station can be obtained from the following algorithm: Define a coordinate system such that the center of cell 1 is at (0, 0). The path distance and angle used to compute the path loss and antenna gain of a MS at (x, y) to a BS at (a, b) is the minimum of the following:

- Distance between (x,y) and (a,b);
- Distance between (x,y) and $(a+3R, b+8\sqrt{3}R/2)$;
- Distance between (x,y) and $(a-3R, b-8\sqrt{3}R/2)$;
- Distance between (x,y) and $(a+4.5R, b-7\sqrt{3}R/2)$;
- Distance between (x,y) and $(a-4.5R, b+7\sqrt{3}R/2)$;
- Distance between (x,y) and $(a+7.5R, b+\sqrt{3}R/2)$;
- Distance between (x,y) and $(a-7.5R, b-\sqrt{3}R/2)$,

The determination of serving cell for each MS is carried out by two steps due to the wrap-around cell layout. The first step is to determine the 19 shortest distance cells for each MS from all seven logical cells clusters, and the second step is to determine the serving cell/sector among the nearest 19 cells for each MS based on the strongest link according to the path-loss and shadowing. It should be noted that the shadowing experienced on the link between MS and cells located in different clusters is the same.

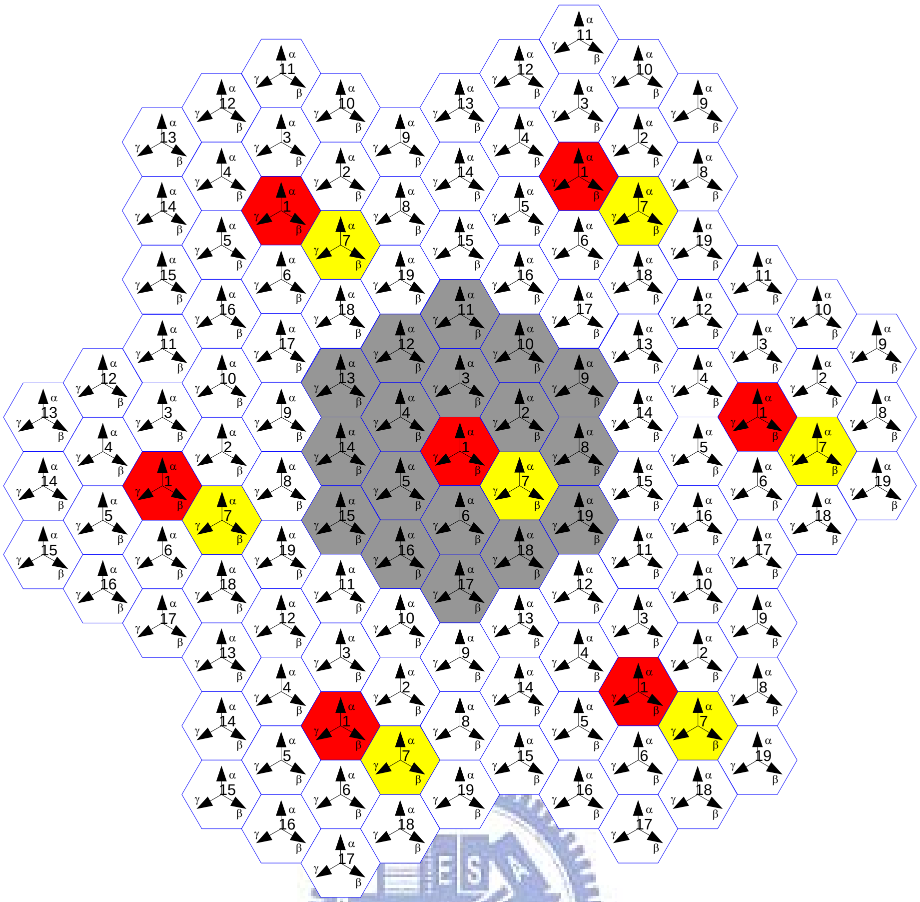

Figure 4-2 the antenna orientations for the system to be used in the wrap-around simulation.

4.2 OFDMA Sub-Carriers Parameters Setting

OFDMA uplink transmission with TDD mode is used in this thesis. Mobile WiMAX applying the IEEE 802.16e-2005 standard supports an asymmetric downlink and uplink transmission of TDD mode for dynamically adjust DL/UL time ratio according to the traffic loading of downlink and uplink transmission. In the platform, 48 OFDMA symbols in total with 1 guard symbol, 28 downlink symbols, and 19 uplink symbols is implemented according to Mobile WiMAX – Part I: A Technical Overview and Performance Evaluation [1]. DL/UL time ratio is 28:19 here. A total bandwidth 30 MHz with frequency reuse (3, 3, 3) is used here. The number of sub-carriers applied is 1024-FFT OFDMA sub-carriers allocations–PUSC mode defined in the IEEE 802.16e-2005 standard. Table 4-1 shows DL/UL ODFMA sub-carriers parameters setting.

Table 4-1 DL/UL OFDMA Sub-carriers Setting

Parameter	Downlink	Uplink
System Bandwidth	10 MHz	
FFT Size	1024	
Null Sub-Carriers	184	184
Pilot Sub-Carriers	120	280
Data Sub-Carriers	720	560
Sub-Channels	35	35
Symbol Period, T_s	102.9 microseconds	
Frame Duration	5 milliseconds	
OFDM Symbols/Frame	48	
Data OFDM Symbols	44	
DL/UL ratio	28:19	

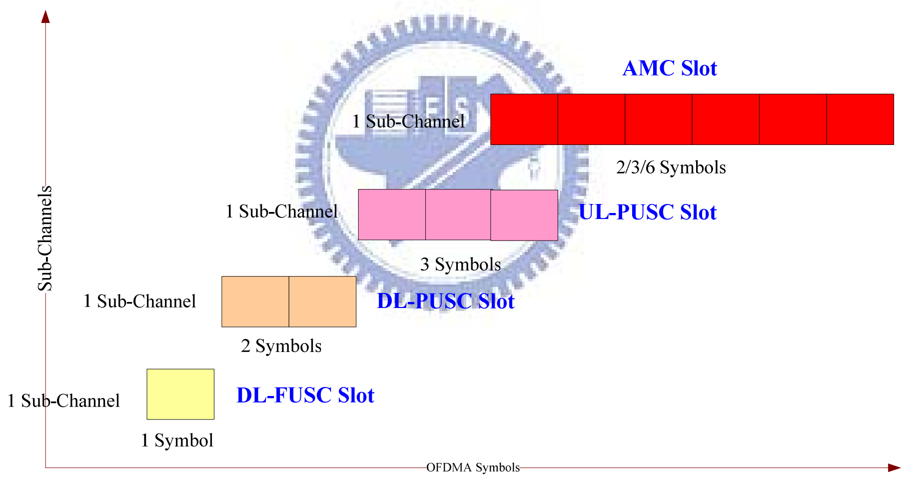

Figure 4-3 Slot Structure in different OFDMA Zones

Uplink PUSC mode, which is three symbols and 1 sub-channel per slot, is used here for the uplink slot structure as shown in Figure 4-3. One OFDMA symbol and one sub-channel are reserved for the usage of ACK channel, CQI channel, and Ranging Channel. According to the parameter setting, there are six time slots and 35 sub-channels in uplink sub-frame. 12 sub-channels per sector and 5 ms frame length are implemented in the simulation platform.

4.3 Link Budget

The link budget settings of uplink transmission in the simulation platform are as far as possible to accommodate the IEEE 802.16e-2005 standard. In IEEE S802.16e-03/23 document [16], it makes deployment scenario assumptions for 802.16e, like Table 4-2. In the simulation platform, the outdoor vehicular scenario is adopted, which the MS transmitted power is 27 dBm and the MS antenna gain is 3 dBi. The BS back off for the avoidance of the RF circuit working in the non-linear region due to the peak-to-average power ratio (PAPR) of OFDM system is 5 dB. The common usage value of thermal noise density is -173.93 dB / Hz. The receiver noise figure of MSs is 9dB [19].

Table 4-2 Link Budget parameters

Scenario / Parameter	Indoor	Outdoor to indoor	Outdoor vehicular
BS Tx power	27 dBm (0.5 W)	36 dBm (4 W)	46 dBm (40 W)
MS Tx power	17 dBm	17 dBm	27 dBm
BS ant gain	6 dBi	17 dBi	17 dBi
MS ant gain	0 dBi	0 dBi	3 dBi
BS ant height		15 m	30 m

In wireless channel, the transmitted signals will suffer the fading effect, which might significantly affect the quality of received signal power. There are three categories of fading effects which are pathloss, shadow fading, and fast fading (Multi-path effect and Doppler effect). Pathloss and shadow fading are considered in the simulation platform. The pathloss model is used to simulate the degradation of signal strength with increasing distance between transmitter and receiver. In Winner D5.4 document [20], it provides several pathloss models, such as Table 4-3.

According to the system parameters setting, C2 scenario is more suitable in the platform.

Table 4-3 Path-loss Model Scenarios

Scenario		Path-loss [dB]	Shadow fading standard dev.	Applicability range
A1	LOS	$18.7 \log_{10}(d[\mathrm{m}])+46.8$	σ = 3.1 dB	$3\mathrm{m} < d < 100\mathrm{m}$
A1	NLOS	$PL(d) = 36.8 \log_{10}(d[\mathrm{m}])+38.8$	σ = 3.5 dB	$3\mathrm{m} < d < 100\mathrm{m}$
B1	LOS	$22.7 \log_{10}(d[\mathrm{m}])+41.0$	σ = 2.3 dB	$10\mathrm{m} < d < 650\mathrm{m}$
B1	NLOS	$0.096\ d_1[\mathrm{m}]+65+$ $(2.8-0.024\ d_1[\mathrm{m}]) \log_{10}(d_2[\mathrm{m}])$	σ = 3.1 dB	$10\mathrm{m} < d_1 < 550\mathrm{m}$ $w/2 < d_2 < 450\mathrm{m}$
C2	NLOS	$35.0 \log_{10}(d[\mathrm{m}])+31.5$	σ = 8 dB	$50\mathrm{m} < d < 5\mathrm{km}$
D1	LOS	$PL(d) = 21.5 \log_{10}(d[\mathrm{m}])+44.6$ $= 40.0 \log_{10}(d/d_{BP})+44.6+$ $21.5 \log_{10}(d_{BP})$	σ = 3.5 dB σ = 6.0 dB	$30\mathrm{m} < d < d_{BP}$ $d_{BP} < d < 10\mathrm{km}$
D1	NLOS	$PL(d) = 25.1 \log_{10}(d[\mathrm{m}])+55.8$	σ = 8.0 dB	$30\mathrm{m} < d < 10\mathrm{km}$

Shadow Fading

Shadow fading effect is the increase or decrease of signal strength due to the shelters, like buildings or mountains, on the signal transmitted path. According to the test result of the real wireless environment, it's known that the variant of shadow fading is a log-normal distribution statistically. Hence, log-normal distribution is applied to simulate shadow fading effect. The standard deviation of this distribution is based on different simulation environments. In the simulation platform, the standard deviation is set to 8 dB [20]. When the user is stationary, shadow fading effect will not alter. Shadow fading effect changes at different locations. However, due to the time duration between two simulation drops is tiny, the distance a MS moves is very short, so that the location of the MS does not change much even if it's at high mobile speed. Therefore, the variance of shadow fading has a correlated relationship associated with the distance the MS moves during the time duration between two neighbor simulation drops. Hence, a correlation model, called Gudmundson's correlation model [21] is used in the simulation platform, in (22).

$$\rho(\Delta x)=e^{-\frac{|\Delta x|}{d_{cor}}\ln 2} \qquad (22)$$

,where ρ is the auto-correlation constant between two simulation drops, Δx is the distance of two simulation time drops and is a function of sampling times, drop duration, and user speed. The d_{cor} is de-correlation distance and the values in the suburban macro, urban macro, and urban micro environments are 200m, 50m, and 5m, respectively.

4.4 Uplink SINR Computation

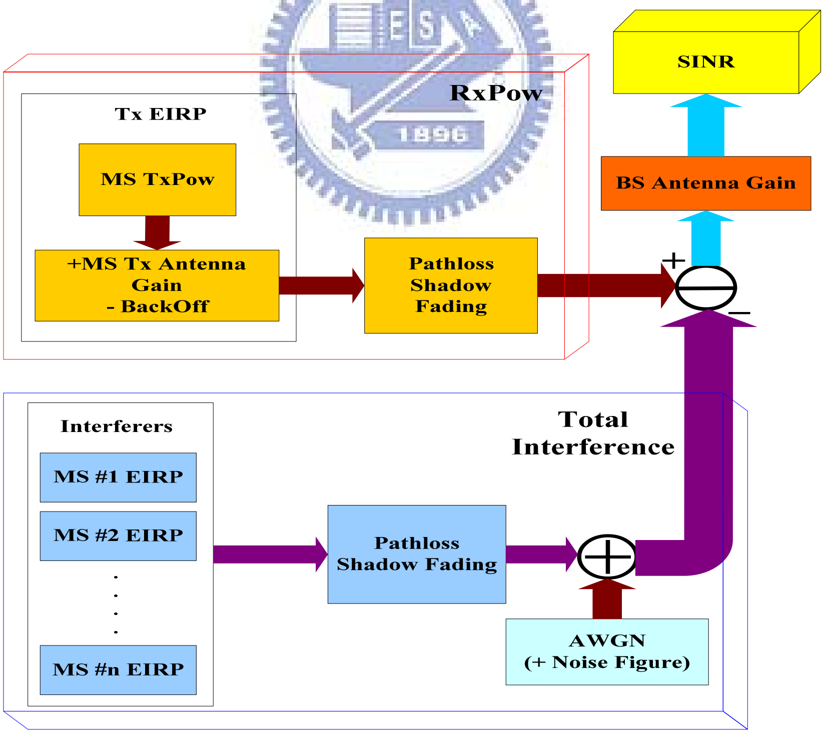

Figure 4-4 Uplink SINR computation

As shown in Figure 4-4, uplink SINR is determined by the received power in BS site and the total interference power caused by the MS interferers in other cells or sectors operating on the same RF channel as the mobile victim with the consideration of path-loss and shadow fading.

The received power is equal to MS transmitted power plus MS antenna gain minus back-off gain through fading effects. For the calculation of the total interference, with the consideration of fading effects, the power of different MS interferers operating on same RF channel in other cells or sectors are accumulated. Additional White Gausian Noise (AWGN) channel with BS noise figure is considered in the platform.

4.5 Basic Functions of Radio Resource Management

In this simulation platform, the basic functions of radio resource managements are implemented as follow.

- **Power Control:** MSs' transmission with the maximum and fixed power is assumed. And the power of per sub-carrier is the same.
- **Rate Control (AMC):** The adaptive modulation and coding scheme is a major method to keep the quality of wireless transmission. The IEEE 802.16e standard supports a variety of modulation and coding scheme. In the simulation platform platform, BPSK, QPSK, 16-QAM, and 64-QAM modulation schemes are used, and BPSK is only used on BR Header transmission. For channel coding, convolution code (CC) with 1/2 code rate is used. 16 sub-carriers per sub-channel and three OFDMA symbols represent an UL-PUSC symbol. Therefore, BPSK, QPSK, 16-QAM, and 64-QAM with CC 1/2 carry 24, 48, 96, and 144 bits respectively.
- **Channel assignment:** The OFDMA frame structure has two dimensions, the slot with two OFDMA symbols and one sub-channel, for channel assignment. In the simulation platform, we obey the definition of 802.16e standard that we introduce in 2.1.2. The basic principles are to segment the data after the

modulation block into blocks sized to fit into one OFDMA slot, and map the slots in the sub-channels with higher priority than that in the OFDMA symbols. In other words, the data mapping method is frequency first.

- **Sub-carrier permutation:** In the simulation platform, the distributed sub-carrier permutation, UL-PUSC, is implemented. For simplicity, instead of implementing the complex permutation formula, the statistic method is implemented to simulate the permutation effect. Interference from other MS interferers operating on the same RF channel will be dispersed equally to the uplink sub-frame.
- **Scheduling method:** Scheduling algorithms, such as robin (RR), proportional fair (PF), MaxCINR, early deadline first (EDF), are build for the baseline performance analysis in the simulation platform.
- **Handoff method:** Handoff is not a weight-bearing point here. Therefore, the simplest method, hard handoff, is implemented. This method is "Break-Before-Make".
- **PDU segmentation:** In the simulation platform, SDU is generated for different data traffic, and segmented as PDUs according to different requirements of target PDU Error Rate of different traffic models (23).

$$PER_{target} = 1 - (1 - BER)^{bits} \qquad (23)$$

BER is the bit error rate which is different for different channel condition feedbacks. Target PDU error rate is different from traffic types. Hence, the PDU segmentation size is calculated based on the PDU error rate. However, an upper bound of PDU size is set to 180 bytes to prevent high error rate caused by too big PDU size.

- **ARQ retransmission:** ARQ is implemented. If the system finds out that the retransmission must exceed the delay bound of real-time service packet before retransmission, the packet will be dropped to save RF resources for other users. For streaming service and non-real-time service, the retransmission timer is set to three. For voice service, the retransmission timer is set to one since smaller packet size.

4.6 Traffic Models

In IEEE 802.16e standard, the uplink data traffics have five different QoS categories which are UGS, rtPS, ertPS, nrtPS, and BE. The details are described in 2.2.4. In the simulation platform, FTP traffic is build as the nrtPS, and VoIP without silence compression is build as UGS. The FTP traffic model adopts 3GPP2 model [22] as shown in Table 4-4. The minimum reserved rate of the non-real-time service is set to 60kbps according to [23]. The VoIP traffic model corresponds to G729-1 codec [24] as shown in Table 4-5. FTP services use TCP/IP protocol to transmit, so the FTP packet needs to add 20 bytes TCP header and 20 bytes IP header. VoIP service uses RTP/UDP/IP protocol to transmit. So, 12 bytes RTP header, 8 bytes UDP header, and 20 bytes IP header are added in a VoIP packet.

Table 4-4 FTP Traffic Model

File size (S)	Truncated Lognormal	A=0.35, u =14.45, M=5m
Packet size	1400 bytes	
Reading time (Dpc)	Exponential	L=1/180

Table 4-5 VoIP Traffic Model

Codec	Framesize(byte)	Interval(ms)	Rate(bps)	Delay bound(ms)
G729-1	20.0	20.0	8k	20.0

Finally, Table 4-6 is the summary of the parameter setting in the simulation platform.

Table 4-6 System parameters setting

Parameters	Value/Comment
Cell layout	Hexagonal grid, 19 cells (wrap around)
Sectors per cell	3
Frequency reuse factor	(3,3,3)
Available Bandwidth	30 MHz with reuse factor (3,3,3)
Antenna Pattern	$70°$ with 20 dB front-to-back ratio, according to [16]
Cell Radius	1 km
Transmitter/Receiver	Uplink (from MSs to BSs)
Duplex	TDD mode
DL/UL Sub-Frame ratio	28:19
Frame Length	5ms, according to [1]
Frame Structure	1024-FFT OFDMA Uplink PUSC, according to [1]
OFDMA Symbol Length	102.9 μs, according to [17] [18]
OFDMA Symbols per slot	3 Symbols (UL-PUSC Slot)
BS Tx Power	46dBm (40 Watt), according to [15]
BS Antenna Gain	17 dBi, according to [15]
BS Back-Off Gain	5 dB, according to [20]
BS Noise Figure	5 dB, according to [20]
Thermal Noise Density	-173.93 dB/Hz, according to [20]
MS Tx Power	27dBm
MS Noise Figure	9dB, according to [18]
MS Antenna Gain	3 dBi, according to [15]

Pathloss Model	35.0log(d[m])+31.5, 50m<d<5km, according to [20]
Shadow Fading	Log-normal distribution with STD=8dB and Gudmundson's correlation model, according to [21]
Mobility Model	MS speed : 30 km/hr Probability to change direction : 0.2 Max. angle for direction update : 45^o
BS Power Control	Maximum and fixed power consumption
MS Power Control	Maximum and fixed power consumption
AMC	BPSK+CC 1/2, QPSK+CC 1/2, 16-QAM+CC 1/2, 64-QAM+CC 1/2, according to [2]
Channel Assignment	Uplink, time first then frequency
Scheduling Algorithms	Round Robin (RR), Proportional Fair (PF) Max CINR (MC), Early Deadline First (EDF)
Handoff	Hard handoff
Traffic Models	FTP [22][23] VoIP [24]

Chapter 5 System Level Simulation Result

In this chapter, we show the system level simulation results of different scheduling algorithms with bandwidth request mechanism in uplink transmission. The scheduling algorithms include Early Dead First, Proportional Fair, Maximum CINR, and Round Robin. We investigate the throughput performance and some QoS-associated factors such as minimum reserved rate, packet delay rate for non-real-time services, packet loss rate for real-time services and AMC Usage.

5.1 Throughput Performance

In this thesis, the system-level simulation is used to investigate the performance of Mobile WiMAX uplink transmission. In this section, the MAC throughput performance and AMC Usage percentage in uplink transmission are discussed in terms of different scheduling algorithms.

5.1.1 MAC Throughput & AMC Usage for full buffer FTP Users Only

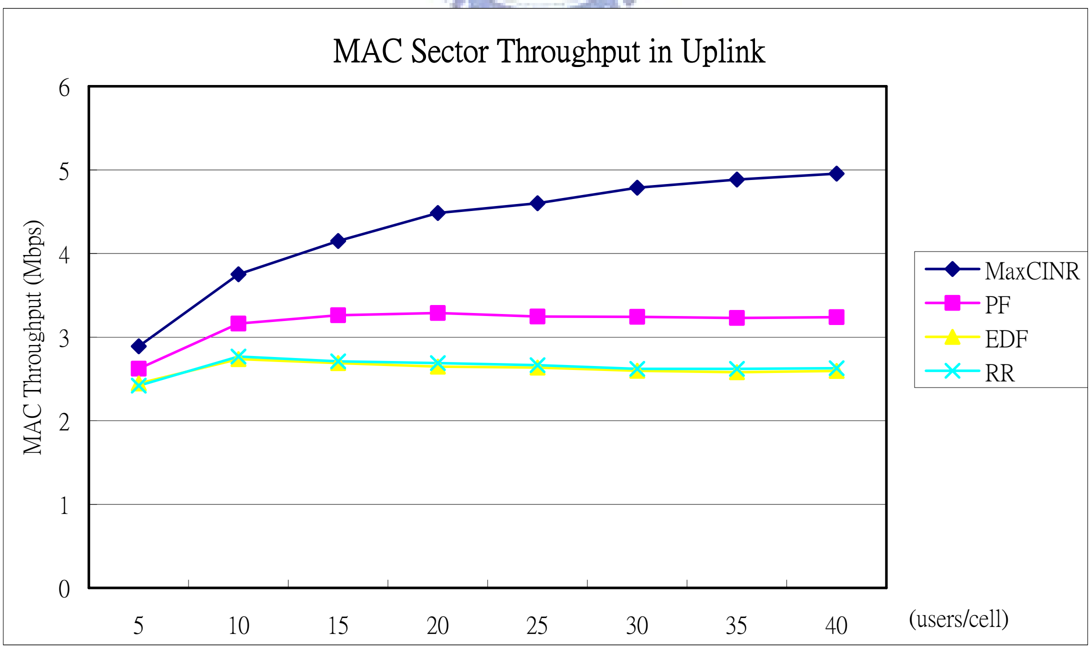

Figure 5-1 MAC Sector Throughput of uplink transmission with different scheduling algorithms

Figure 5-1 shows MAC Sector Throughput with different scheduling algorithm simulation in uplink transmission. MaxCINR has the highest MAC throughput since BS serves MSs having best carrier to noise ratio (CINR) first, thus users using highest order modulation and coding scheme are always served in the first priority. At the simulation point of five users, since there're not so many users having good CINR, the difference of throughput performance is not so different from other scheduling algorithms. While more users having higher order AMC are in the environment, the advantage of throughput performance in MaxCINR in comparison with others is more apparent as shown in blue line in Figure 5-1. In addition, Figure 5-5 shows the percentage of AMC usage in MaxCINR. The percentage of 64QAM users is increasing with number of users, since 64QAM users are always having first priority for uplink transmission in MaxCINR algorithm.

As referred to Proportional Fair (PF), it takes the instantaneous user data rate, which represents the channel condition, into consideration in order to enhance the system throughput performance while moving average phy throughput is considered meanwhile to provide the fairness of transmission opportunity. As shown in pink line in Figure 5-1, Proportional Fair algorithm performs the second highest MAC Sector throughput of all scheduling algorithms, since EDF and RR, without considering users' channel condition, are not designed for throughput enhancement. In Figure 5-4, the percentage of AMC Usage is shown in PF scheduling.

Throughput performance of Round-Robin algorithm is shown by green light in Figure x-x. It fully considers the fairness of transmission opportunity while MaxCINR and PF are focusing on enhancing throughput performance.

Throughput performance of EDF algorithm is shown by yellow line in Figure 5-1. While it performs bad throughput performance, it provides the strongest QoS Guarantee of all scheduling algorithms. For RR and EDF algorithm, the AMC percentage, shown in Figure 5-2 and 5-3, remains similar while the users increase from 5 to 60. So do the throughput performance. That is because they schedule packets regardless channel condition. Also, RR and EDF have more percentage of QPSK users and much less percentage of 64QAM users than PF and MaxCINR.

In conclusion, the MAC Sector Throughput reaches the highest 5Mbps in MaxCINR, 3.1 Mbps in PF, and 2.8 Mbps in EDF and RR while there're 40 users in each cell..

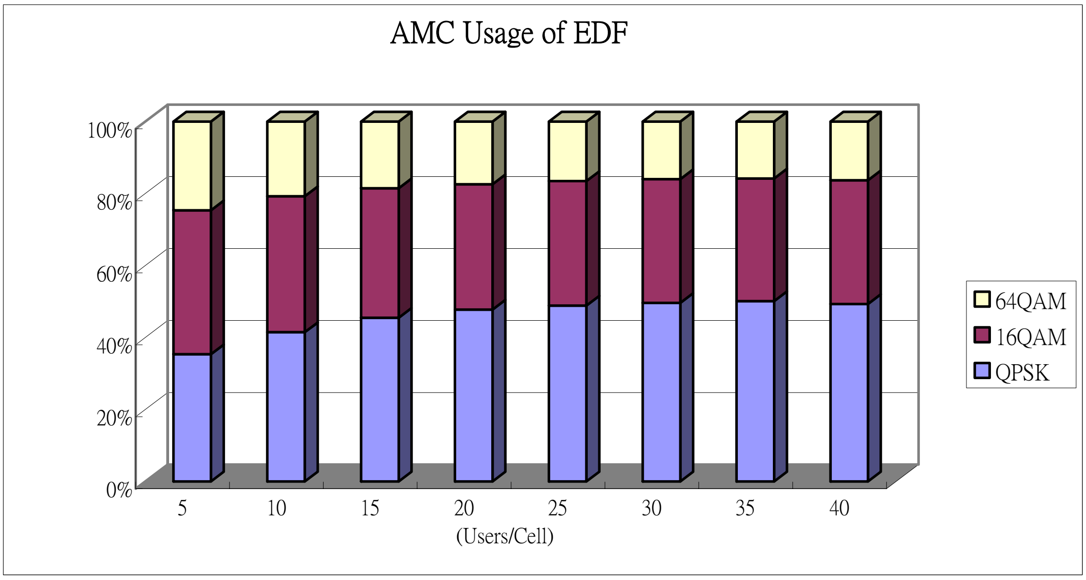

Figure 5-2 AMC Usage of EDF

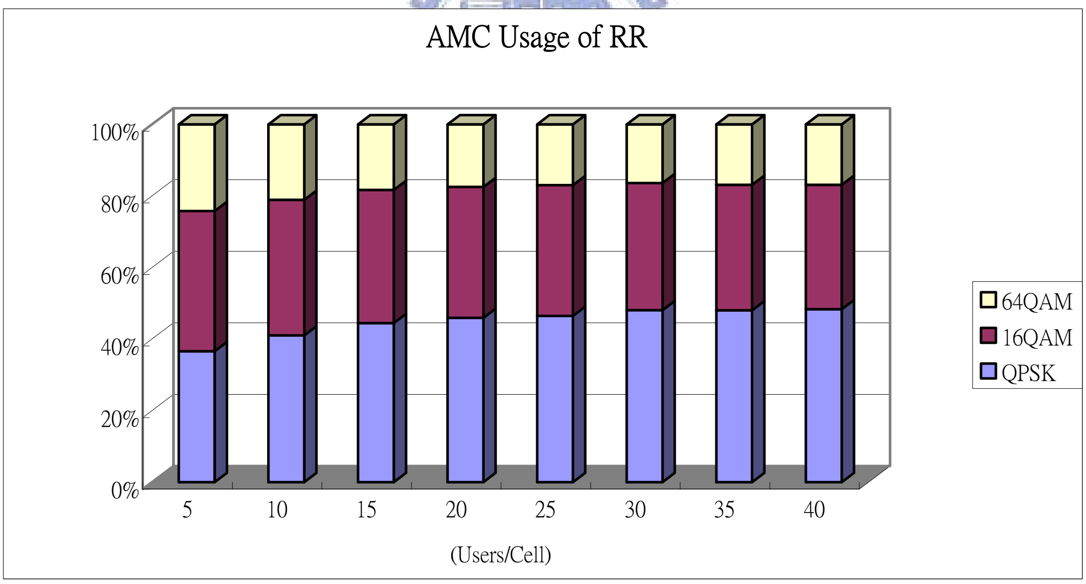

Figure 5-3 AMC Usage of RR

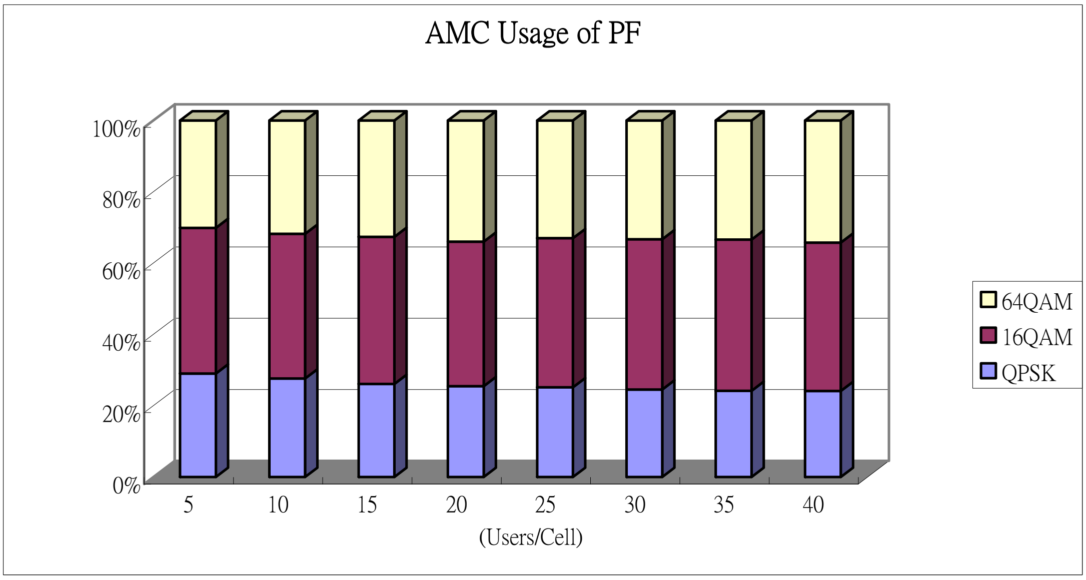

Figure 5-4 AMC Usage of EDF

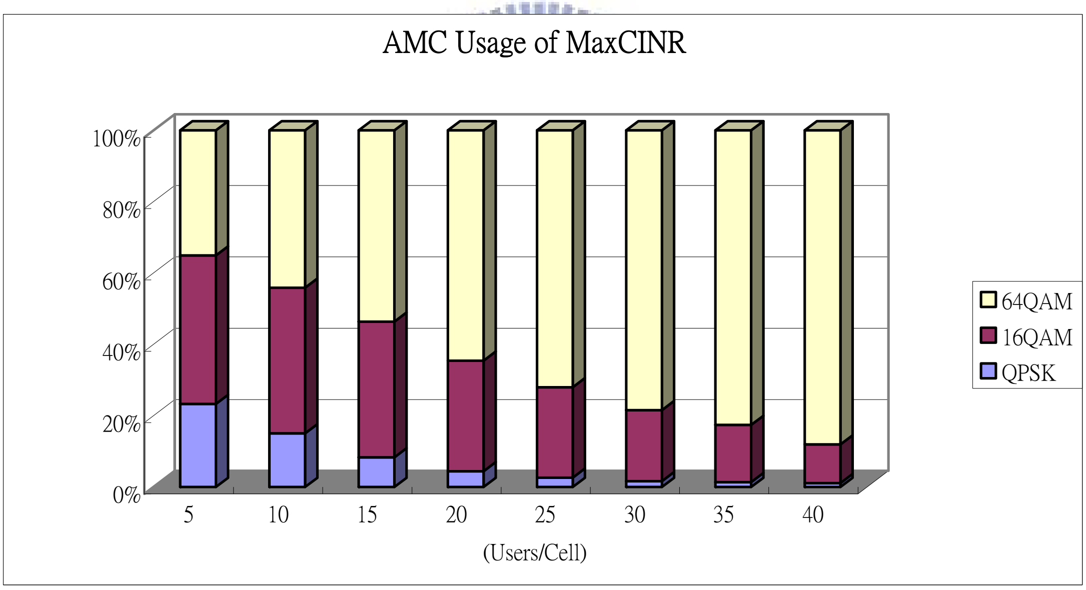

Figure 5-5 AMC Usage of EDF

5.1.2 MAC Throughput for FTP Users in mixed-traffic Environment

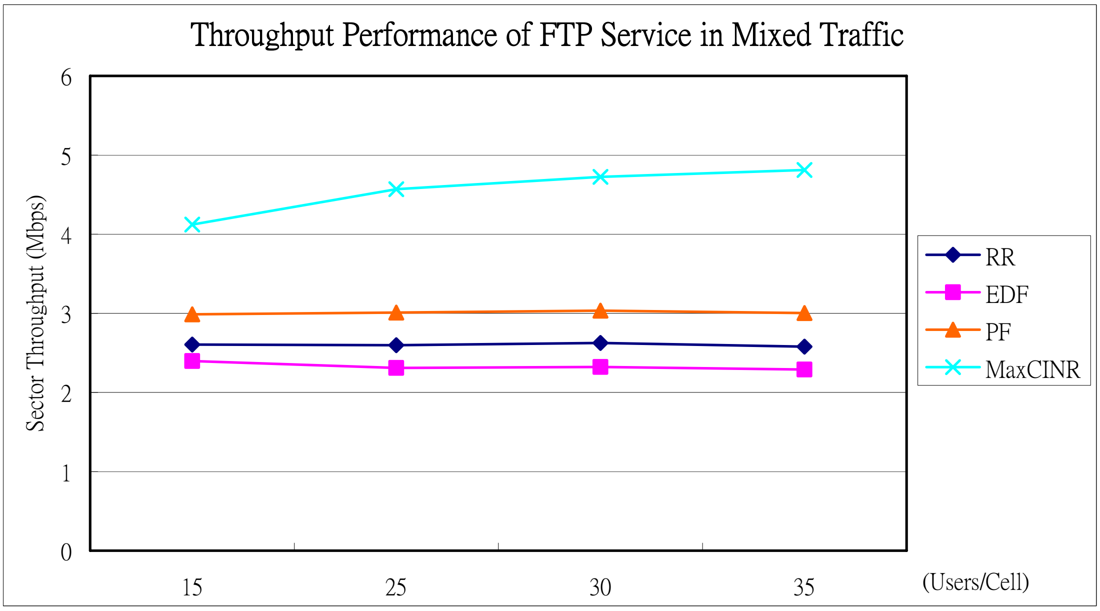

Figure 5-6 Throughput Performance of FTP Service in Mixed Traffic (10 VoIP Users)

Figure 5-6 shows the throughput performance of FTP Service in the mixed traffic environment (10 VoIP Users). Compared to the throughput performance of FTP service only, we observe that FTP throughput of each scheduling algorithm slightly degrades due to VoIP traffic's occupation of resource units. EDF algorithm has the lowest sector throughput since it always serves real-time service first, following with non-real-time service, in order to prevent packet loss.

5.2 Performance of Quality of Service

In this thesis, the uplink performance of Mobile WiMAX is investigated in terms of VoIP, FTP, and Mixed Traffic. The indication of QoS is defined as packet delay rate (PDR) and QoS Failure Rate (QFR) for non-real-time service (FTP) and packet loss rate (PLR) for real-time service (VoIP). PDR represents the percentage of packets exceeding its soft delay bound, and QoS Failure Rate is the percentage of users who don't meet the QoS requirements, which is minimum reserved rate in FTP Service. PLR indicates the percentage of packets exceeding its delay bound.

In the section, the uplink system level simulation results of PDR, QoS Failure Rate (QFR) for non-real-time traffic and mix-traffic are shown and discussed.

5.2.1 QoS for full buffer FTP Users Only

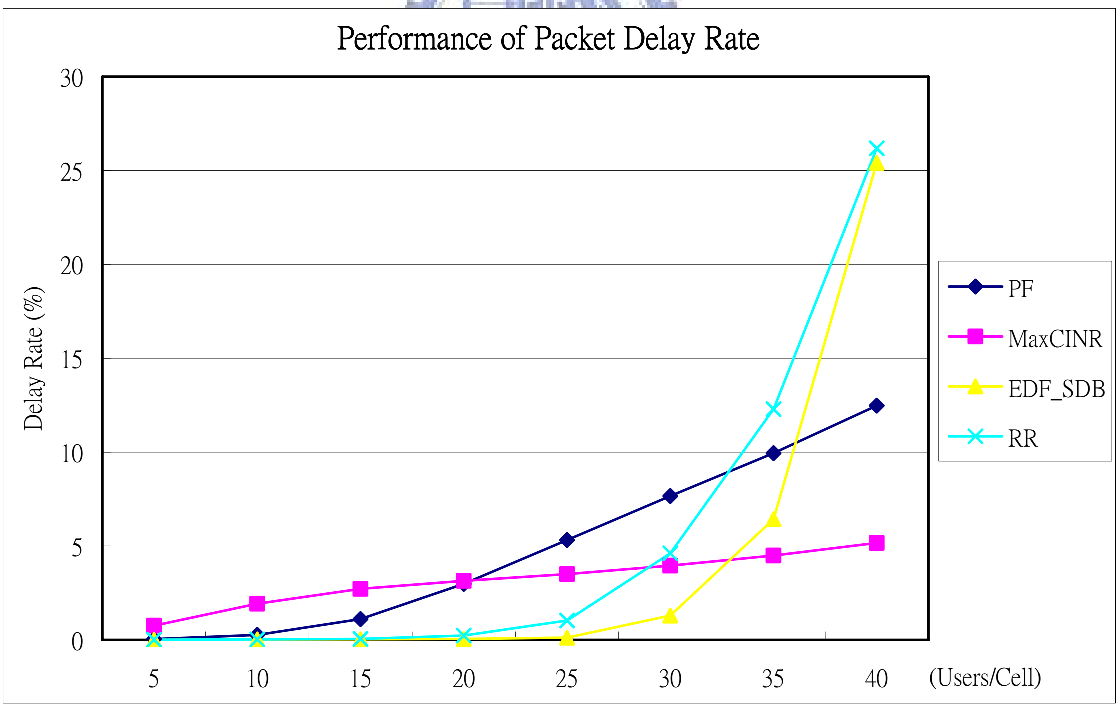

Figure 5-7 Packet Delay Rate in uplink transmission

Figure 5-7 shows the performance of Packet Delay Rate in different scheduling algorithms. EDF can perform the best PDR while number of users per cell is less than 35, and the PDR goes up rapidly to 25% with 40 users per cell due to the congestion of data traffic. The plot of RR is similar to the plot of EDF. For PF and MaxCINR, the PDR goes high in the beginning, but has a small slope compared to the other two algorithms. Therefore, the PDRs of MaxCINR and PF are lower then RR and EDF while number of users per cell exceeds 35. The reason MaxCINR begins with a high PDR and goes up with a small slope is that users having higher order AMC are served in the first priority without consideration of QoS requirement, so the PDR is quite bad as expected in the beginning of the plot. However, with number of users per cell increases, users with good channel condition increase, and almost same users who always have best CINR and best modulation scheme are served in every frame. Thus, other users don't even have a chance to transmit. Due to transmission of same users with best AMC, transmitted packets are hard to exceed its soft delay bound. This leads to a low PLR in the end of the plot. So does PF algorithm.

Although Packet Delay Rate shows the number of packets exceeding the soft delay bound, people don't know exactly how many users are not achieving their QoS requirement. In the simulation, FTP service is used as a non-real-time service with the minimum reserved rate as the QoS requirement. Hence, we use another indicator to show the percentage of users who don't meet the minimum reserved rate as shown in Figure 5-8.

Unlike the performance of PLR, EDF algorithm has the best QoS failure rate from the beginning to the end of the plot due to the strict consideration of user packet delay bound. MaxCINR performs worst QoS failure rate since the order of transmission in this algorithm is not associated with users' QoS requirement but system throughput performance. PF algorithm performs a not bad QoS failure rate compared to other three algorithms while it enhances the system throughput at the same time.

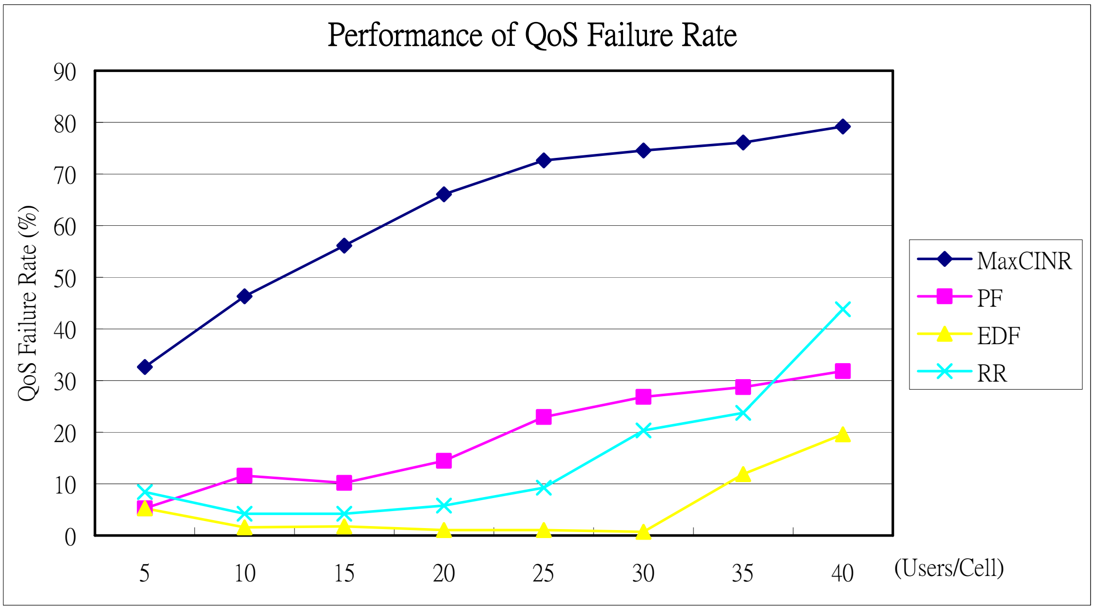

Figure 5-8 QoS Failure Rate of FTP users in uplink

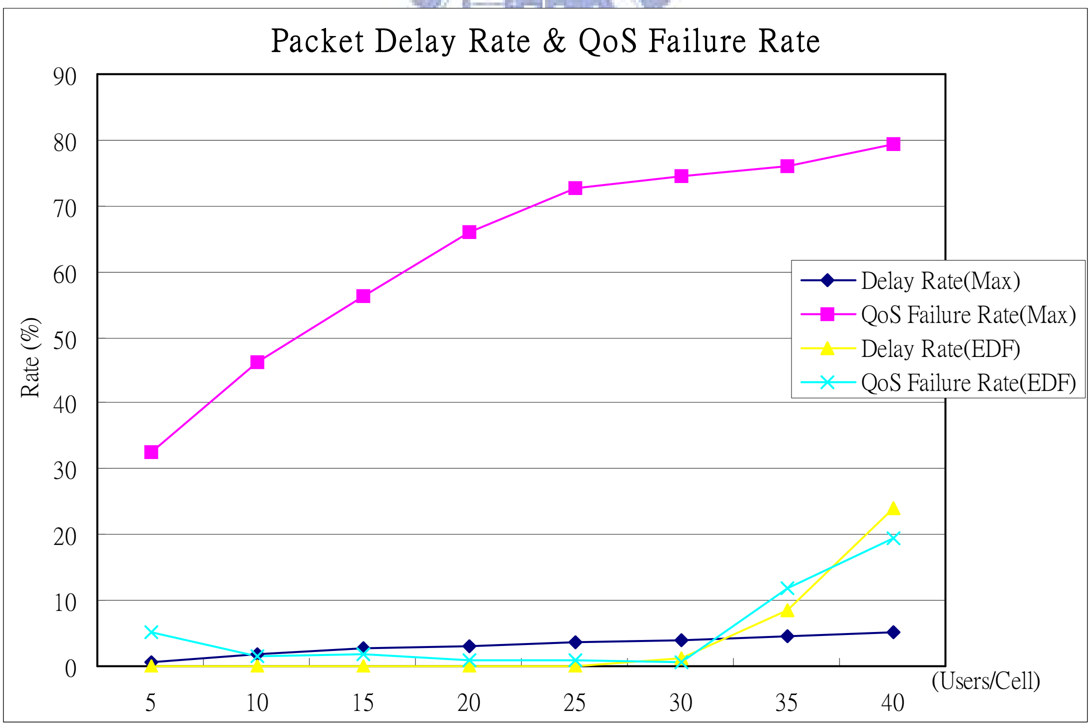

Figure 5-9 Comparison of Packet Delay Rate & QoS Failure Rate

Figure 5-9 compares Packet Delay Rate and QoS Failure Rate together to show the relationship between the two QoS indicators. We can clearly see that while MaxCINR sustain a low packet delay rate, it also performs a high QoS failure rate meanwhile. For the case of EDF algorithm, the packet delay rate and QoS failure rate are in a consistent trend, but the two curves are still quite different.

In conclusion, although PLR is consistent with QFR in some cases, it is still not capable of demonstrating the real QoS performance. However, the usage of soft delay bound concept in non-real-time service still has some benefits and advantages in scheduler design. It is discussed in the following.

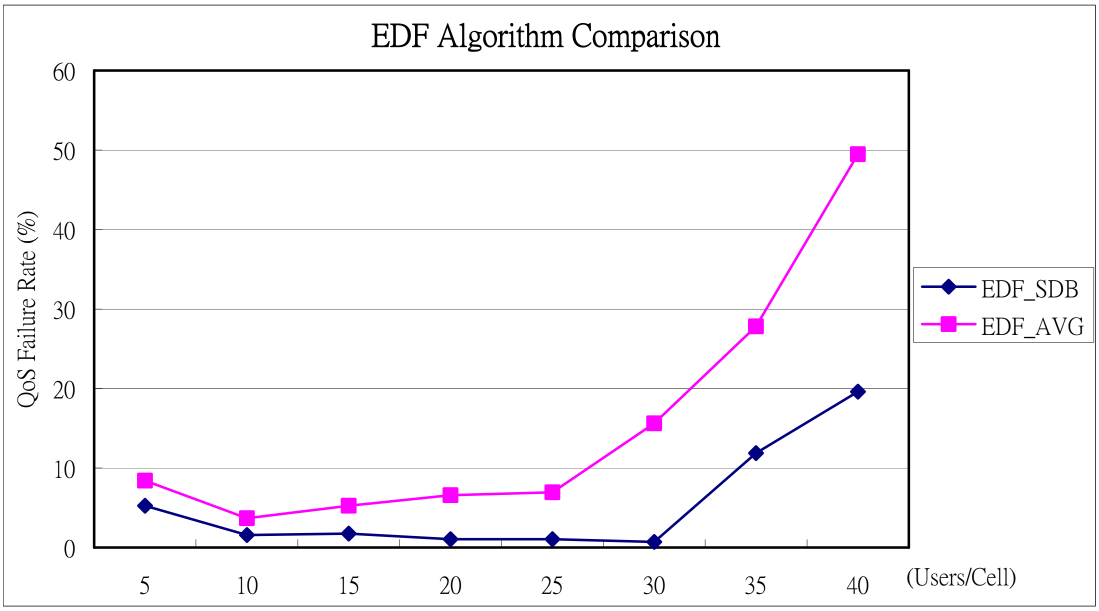

Figure 5-10 Performance Comparison of Soft Delay Bound and Average Data Rate

Figure 5-10 shows the QoS performance of Early Deadline First (EDF) algorithm using two different methods for the order of packet transmission. One is to arrange the order of packets based upon packets' soft delay bound (SDB), and packets having shorter delay bound are served with higher priority. The other method is to arrange the order based upon user average throughput (AVG), and connections transmitting with a lower throughput are served with higher priority. The SDB method is obviously better than the AVG method in terms of QoS failure rate. While apply EDF as the scheduling algorithm for non-real-time service like FTP, indicators for arrange order of packet transmission are significant. With AVG method, the scheduler might choose connections which transmit in a lower throughput, but are not urgent to transmit data to accommodate with the minimum reserved rate. In addition, AVG method might cause the scheduler to serve users which have high data rate, but are urgently necessary to transmit for achieving the minimum reserved rate. However, with SDB method, the scheduler can find packets which are urgent to meet the minimum reserved rate, and allocate transmission opportunity for them first.

5.2.2 Performance of QoS for full buffer VoIP Users Only

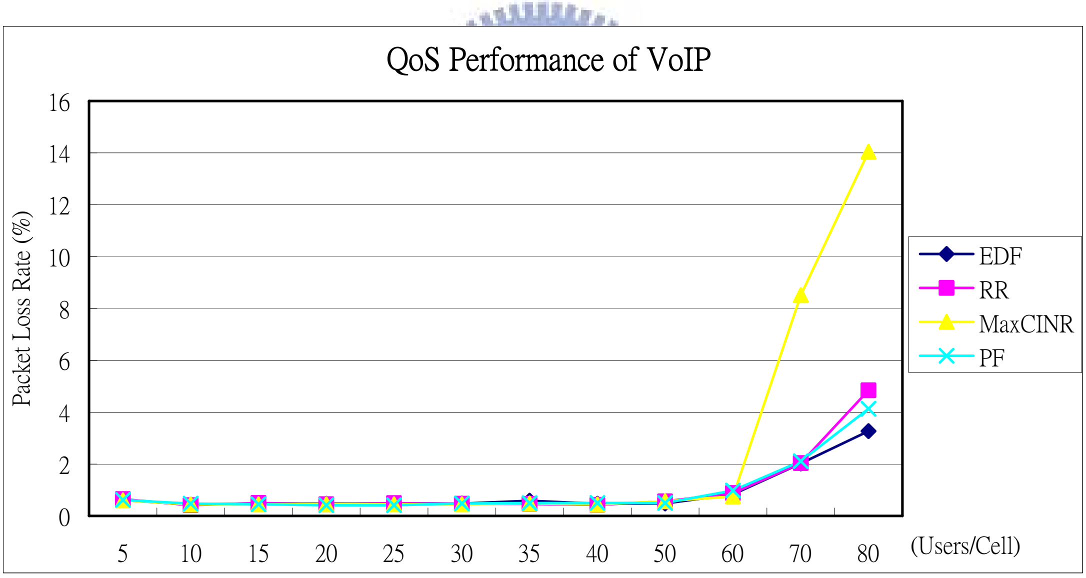

Figure 5-11 Packet Loss Rate of VoIP service only

For real-time-service, such as Voice Over IP (VoIP), data rate is no more an appropriate indicator to measure the QoS control. People concern the Quality of voice during communication. So, the QoS requirement of VoIP is packet loss rate. Figure 5-11 shows packet loss rate of VoIP users only. As expected, MaxCINR performs worst,, and EDF achieve the best QoS performance.

5.2.3 Performance of QoS in mixed-traffic Environment

In this section, we investigate and discuss the QoS performance of real-time service and non-real-time service in terms of different scheduling algorithm in a mixed traffic environment. And we also discuss and compare the result of single traffic QoS with that of mixed traffic QoS to understand the level of difference. For the simulation of non-real-time service, the mixed traffic is set to be 10 and 20 full buffer VoIP users with 15, 25, 30, 35 FTP users respectively. For simulating real-time service, the mixed traffic is set to be 10 FTP users co-existing with 10, 20, 25, 30 VoIP users per cell respectively.

5.2.3.1 Performance of QoS for FTP users in mixed-traffic Environment

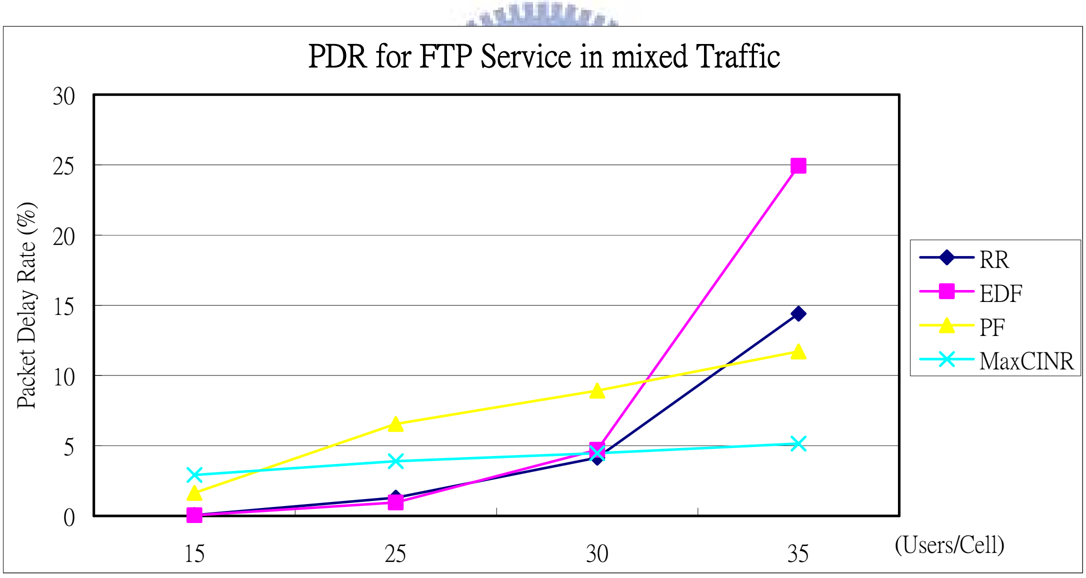

Figure 5-12 Packet Delay Rate of FTP users in mixed Traffic (10 VoIP Users/Cell)

In a mixed traffic environment, we consider FTP traffic and VoIP traffic both in full buffer for non-real-time and real-time service respectively. As shown in Figure 5-12, performance of FTP service in mixed traffic is shown. As aforementioned, MaxCINR in light-blue line keeps with a small slope since almost same users having highest order modulation scheme are served in every frame, so that transmitted packets are not easy to exceed its soft delay bound. Again, this reveals PDR is not capable of demonstrating the real performance of QoS satisfaction. EDF in pink line goes from the lowest PDR to the highest PDR of all scheduling algorithms due to traffic congestion and that every user has chance to transmit based upon its delay bound. Proportional Fair in yellow line goes in the middle because it considers both throughput performance and transmission fairness. We observed that EDF has the highest packet delay rate at the simulation point of 35 users per cell. That is because the EDF algorithm in this platform serves real-time service first then non-real-time service in order to meet the strict timing constraint of its QoS requirement, so users in non-real-time service are expected to have less radio resources to transmit.

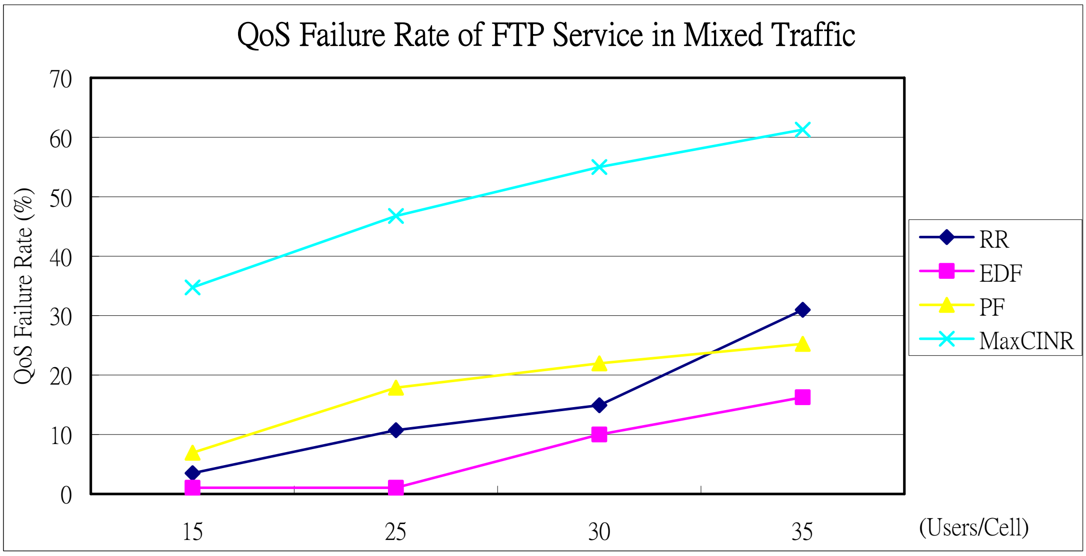

Figure 5-13 QoS Failure Rate for FTP users in Mixed Traffic (10 VoIP Users/Cell)

QoS Failure Rate of FTP Service in the mixed traffic environment is shown in Figure 5-13. As expected, while MaxCINR shown by light-blue line has a low packet delay rate and a flat plot, it performs the highest QoS failure rate due to the ignorance of packet delay bound. EDF shown by pink line still has the lowest QoS failure rate , the same as the simulation result of single-traffic environment.

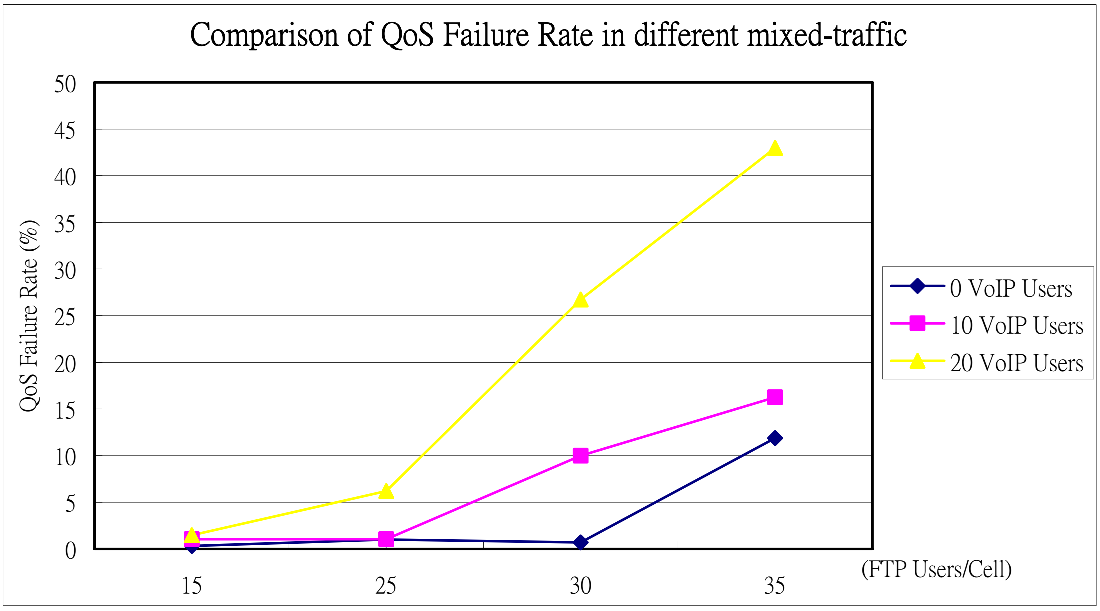

Figure 5-14 Comparison of FTP QoS Failure Rate in different mixed-traffic

Figure 5-14 shows the QoS Failure Rate of QoS Failure Rate of FTP users in different mixed-traffic environment. The blue line represents zero VoIP users co-existing with 15, 25, 30, and 35 FTP users in each cell respectively. The pink line represents 10 VoIP users co-existing with 15, 25, 30, and 35 FTP users in each cell respectively, and the yellow line represents 20 VoIP Users co-existing with 15, 25, 30, and 35 FTP users in each cell respectively. As expected, the QoS Failure Rate increases significantly while number of co-existing VoIP users increase in a cell. Since EDF algorithm serves real-time packets first in order to ensure a qualified packet loss rate, with the increase of co-existing VoIP users, the amount of resource units for FTP users decreases.

5.2.3.2 Packet Loss Rate for VoIP users in mixed-traffic Environment

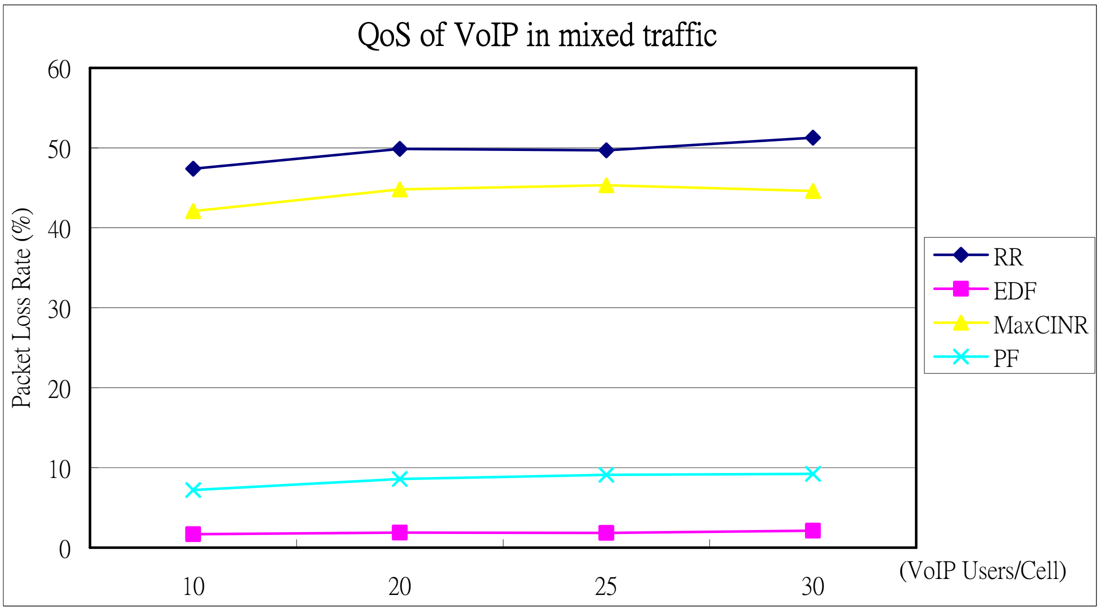

Figure 5-15 Packet Loss Rate of VoIP service in mixed traffic (10 FTP Users/Cell)

Figure 5-15 investigates the overall packet loss rate of VoIP users in the mixed traffic environment, which is set to be 10 FTP users per cell and respectively 10, 20, 25, 30 VoIP users per cell. Due to the large packet size of FTP service, one uplink sub-frame might run out of resource units with one or two FTP users. Since Round-Robin and MaxCINR algorithm don't tempt to serve VoIP users first, it's observed that both of them have high packet loss rate even with only 10 FTP users in one cell. Due to the same reason, Proportional Fair algorithm also fails to meet the QoS requirement of VoIP service, which is 2% packet loss rate. However, Early Deadline First algorithm performs the best QoS for VoIP users in mixed traffic environment since VoIP and FTP packets are served in a separate order, and VoIP packets are always allocated resources first. Due to the higher priority of VoIP service, the number of VoIP users per cell doesn't affect the performance much. The PLR slightly goes up because number of FTP packets in ARQ increases, to which resource units are allocated even before VoIP packets.

Chapter 6 Conclusion and Future Work

Fist of all, we introduce the Physical and MAC layer of Mobile WiMAX, which applies the IEEE 802.16d-2004 and IEEE 802.16e-2005 standard as reference. Then, we build up the system-level simulator for Mobile WiMAX and the signaling control plane for the uplink bandwidth request Mechanism. By exploiting the simulator, we focus on studying the overall and complete uplink performance and investigating the advantages and disadvantages of different scheduling algorithms, including Early Deadline First (EDF), Proportional Fair (PF), Maximum CINR (MaxCINR), and Round-Robin (RR). In addition, we further discuss the capacity and QoS issue of different scheduling algorithms in terms of different traffic type in both single traffic and mixed traffic environment. By the way, the critical positions making Mobile WiMAX one of the 3G communication systems in IMT-2000 are its coverage and high data throughput. Most of researches and papers for WiMAX are focusing on downlink performance study or metrics of enhancement. However, the uplink performance also plays an important role nowadays while bandwidth-thirsty applications, such as video streaming, and web3.0-based Mobile Web-Clubs become more and more prevalent. Mobile subscribers indeed have the intense desire of sharing their things with public or friends. This makes uplink scheduling and transmission mechanism much more important than past. Through our system-level simulation, the uplink MAC throughput in Mobile WiMAX is clearly revealed, and meanwhile several common scheduling algorithms are implemented to get a complete performance analysis of uplink. In addition, a soft delay bound (SDB) method is proposed to properly schedule non-real-time service, such as FTP. It transforms the minimum reserved rate to time domain delay bound. This method facilitates the scheduler design and performs better QoS failure rate than just using data rate as the

QoS indicator.

For the next generation of IEEE 802.16 standard, the system throughput and capacity are at least double than those in the IEEE 802.16e-2005 standard. Therefore, the uplink scheduler shall be more sophisticated in order to satisfy the need of high data throughput and accommodate their QoS meanwhile. In order to leverage our research in this thesis, we will focus our future study on downlink and uplink scheduler design for a mixed traffic user environment.

Reference

[1] IEEE, "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), 2004 Page(s):0_1 - 857

[2] IEEE, "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1",IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005 (Amendment and Corrigendum to IEEE Std 802.16-2004), 2006 Page(s):0_1 - 822

[3] David Paranchych, "WiMAX / 802.16 Overview", NORTEL, Wireless Technology & Standards, Feb. 10, 2006.

[4] Vijay Dube, "CAN OFDM ENHANCEMENT DRIVE WIMAX MOBILITY FORWARD?", Wavesat, Oct 1, 2005.

[5] GuoSong Chu; Deng Wang; Shunliang Mei, David, "A QoS Architecture for the MAC Protocol of IEEE 802.16 BWA System", IEEE International Conference on Communications Circuits & System and West Sino Expositions, vol.1, Page(s): 435--439,

[6] Dong-Hoon Cho, Jung-Hoon Song, Min-Su Kim, Ki-Jun Han "Performance Analysis of the IEEE 802.16Wireless Metropolitan Area Network", Proc. First Int'l Conf. Distributed Frameworks for Multimedia Applications (DFMA '05), Page(s): 130-137, Feb. 2005.

[7] 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 on Communications (ICC2005)

[8] Wengerter, C.; Ohlhorst, J.; von Elbwart, A.G.E., "Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA", Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st Volume 3, 30 May-1 June 2005 Page(s):1903 - 1907 Vol. 3

[9] Seokhyun Yoon; Youngkwon Cho; Chan-Byoung Chae; Hyunwoo Lee, "System level performance of OFDMA forward link with proportional fair scheduling", Personal, Indoor and Mobile Radio Communications, 2004. PIMRC 2004. 15th IEEE International Symposium on Volume 2, 5-8 Sept. 2004 Page(s):1384 - 1388 Vol.2

[10] Seungwan Ryu; Byunghan Ryu; Hyunhwa Seo; Mooyong Shin, "Urgency and Efficiency based Packet Scheduling Algorithm for OFDMA wireless system", Communications, 2005. ICC 2005. 2005 IEEE International Conference on Volume 4, 16-20 May 2005 Page(s):2779 - 2785 Vol. 4

[11] Chiussi, F.M.; Sivaraman, V., "Achieving high utilization in guaranteed services networks using early-deadline-first scheduling", Quality of Service, 1998. (IWQoS 98) 1998 Sixth International Workshop on 18-20 May 1998 Page(s):209 - 217

[12] Meng-Shiang Lin, “Radio Resource Management in Mobile WiMAX System”

[13] Tony Klinkert, "802.16e System Requirements Feedback from an operator perspective", IEEE 802.16 Presentation SubmissionTemplate (Rev. 8.3), IEEE S802.16e-03/23, 2003-03-10.

[14] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Spatial channel model for Multiple Input Multiple Output (MIMO) simulations (Release 6)", TR 25.996 V6.1.0, 2003-09.

[15] “Mobile WiMAX – Part I: A Technical Overview and Performance Evaluation”, June 2006

[16] “Mobile WiMAX – Part II: A Comparative Analysis”, May 2006

[17] Harri Holma; Antti Toskala,"WCDMA for UMTS – Radio Access For Third Generation Mobile Communication", ISBN 0-471-72051-8 , John Wiley & Sons , 2000, pp.157~159.

[18] Winner, "Final Report on Link Level and System Level Channel Models", IST-2003-507581 WINNER D5.4 v. 1.4, Nov. 18th, 2005.

[19] Gudmundson, M., "Correlation model for shadow fading in mobile radio systems", Electronic Letters, pp.2145-2146, Vol. 27, Issue 23, November 1991.

[20] “cdma2000 Evaluation Methodology Revision 0”, 3GPP2 C.R1002-0 Version 1.0, December 10, 2004

[21] ”3G & The Future of Wireless”, PDAs & Cell Phones, Vol.8 Issue 5,Page(s) 48-51, May 2002

[22] Winner, “Intermediate requirements per scenario”, IST-2003-507581 WINNER D1.2, 02/03/2005.

[23] “IEEE 802.16m System Requirements Document,” *IEEE Standard Draft Document*.

[24] “IEEE 802.16m Evaluation Methodology,” *IEEE Standard Draft Document*.