QUALITY OF SERVICE AND CHANNEL-AWARE PACKET BUNDLING FOR CAPACITY IMPROVEMENT IN CELLULAR NETWORKS

A DISSERTATION IN
Telecommunications Networking
and
Computer Networking

Presented to the Faculty of the University
of Missouri-Kansas City in Partial Fulfillment of
the Requirements for the Degree

DOCTOR OF PHILOSOPHY

by
JUNG HWAN KIM

B.S., Korea Advanced Institute of Science and Technology, 1993
M.S., University of Missouri-Kansas City, 2002

Kansas City, Missouri
2011

QUALITY OF SERVICE AND CHANNEL-AWARE PACKET BUNDLING FOR CAPACITY IMPROVEMENT IN CELLULAR NETWORKS

Jung Hwan Kim, Candidate for the Doctor of Philosophy Degree

University of Missouri-Kansas City, 2011

ABSTRACT

We study the problem of multiple packet bundling to improve spectral efficiency in cellular networks. The packet size of real-time data, such as VoIP, is often very small. However, the common use of time division multiplexing limits the number of VoIP users supported, because a packet has to wait until it receives a time slot, and if only one small VoIP packet is placed in a time slot, capacity is wasted. Packet bundling can alleviate such a problem by sharing a time slot among multiple users.

A recent revision of cdma2000 1xEV-DO introduced the concept of the multi-user packet (MUP) in the downlink to overcome limitations on the number of time slots. However, the efficacy of packet bundling is not well understood, particularly in the presence of time varying channels.

We propose a novel QoS and channel-aware packet bundling algorithm that takes advantage of adaptive modulation and coding. We show that optimal algorithms are NP-complete and recommend heuristic approaches. We also show that channel utilization can be significantly increased by slightly delaying some real-time packets within their QoS requirements while bundling those packets with like channel conditions. We validate our study through extensive OPNET simulations with a complete EV-DO implementation.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Graduate Studies, have examined a dissertation titled "Quality of Service and Channel-aware Packet Bundling for Capacity Improvement in Cellular Networks," presented by Jung Hwan Kim, candidate for the Doctor of Philosophy degree, and certify that in their opinion it is worthy of acceptance.

Supervisory Committee

Baek-Young Choi, Ph.D., Committee Chair
Department of Computer Science Electrical Engineering

Cory Beard, Ph.D.
Department of Computer Science Electrical Engineering

Ken Mitchell, Ph.D.
Department of Computer Science Electrical Engineering

Xiaojun Shen, Ph.D.
Department of Computer Science Electrical Engineering

Chuanjun Zhang, Ph.D.
Department of Computer Science Electrical Engineering

CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ALGORITHMS

ACKNOWLEDGEMENTS

I am heartily thankful to my principal advisor, Dr. Baek-Young Choi, whose encouragement, guidance and support from the initial to the final level enabled me to develop my doctoral research. I would like to show my gratitude to Dr. Cory Beard during the completion of our Sprint project. It is an honor for me to acknowledge the advice and support given to me by the members of my advisory committee that includes Dr. Ken Mitchell, Dr. Xiaojun Shen, and Dr. Chuanjun Zhang. I would also like to thank Dr. Khosrow Sohraby for his guidance and constant support as well as John Kim and Shiva Narayanabhatla for their practical insights and information for the implementation of the EV-DO simulator.

On a personal note, I would like to thank my wife, Heejin, and my son, Justin, for their wonderful support during the course of my doctoral study. I would also like to offer my deepest gratitude to my mother for her encouragement throughout my whole life.

CHAPTER 1

INTRODUCTION

A growing demand for bandwidth intensive applications such as Web browsing and file transfers over wireless networks, urges the need to use the wireless channel efficiently. Moreover, an emerging strong demand for delay-sensitive data applications such as VoIP, wireless gaming, and push-to-talk (PTT) over cellular networks, poses challenges on a network system to support a large number of simultaneous users while meeting their desired delay requirements.

Since the capacity of wireless systems is particularly constrained by the nature of location dependent and time varying channel conditions, careful attention needs to be paid to algorithms over wireless links in order to use the channel as efficiently as possible. In this work, we study the problem of multiple packet bundling to improve spectral efficiency in cellular networks. The packet size of real-time data, such as VoIP, is often very small. However, the use of time division multiplexing (TDM) on the forward link limits the number of VoIP users supported, because a packet has to wait until it receives its own dedicated time slot. The time slot should not be made too small, however, due to the relative MAC layer overhead for each time slot. Packet bundling can alleviate such a problem by sharing a time slot among multiple users.

Most wireless standards define the QoS framework and various types of service

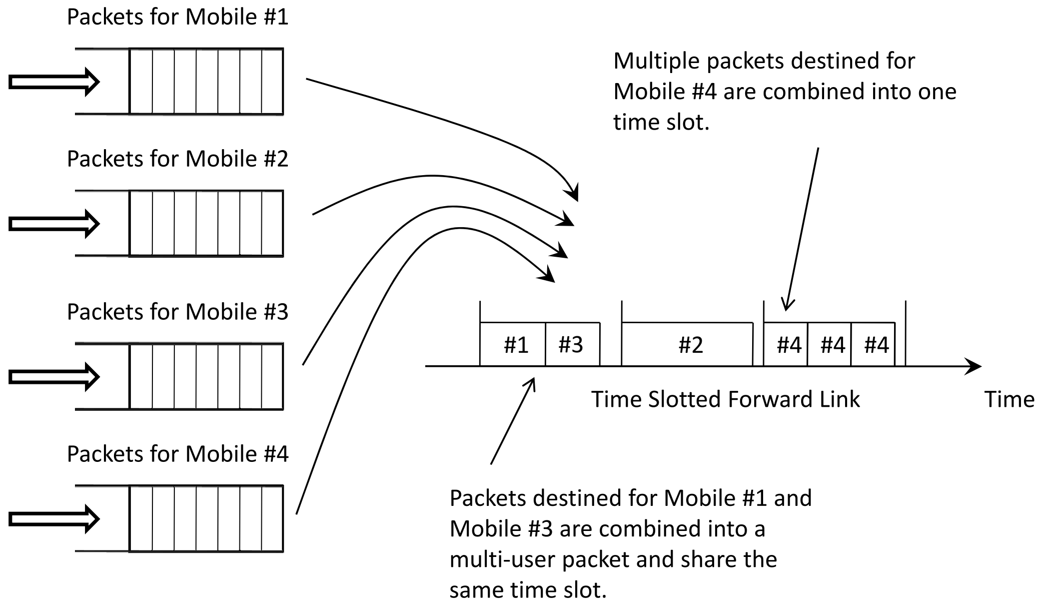

Figure 1: The concept of packet bundling

flows, but leave the QoS-based packet scheduling and radio resource assignment undefined. For example, a ’multi-user packet’ is among the improvements and expansions of EV-DO Rev. A and it permits the base station to serve multiple users with the same physical and MAC layer packet. However, there is no guideline or recommended strategy in multiple packet bundling, and the efficacy of multi-user packets is not well understood, especially in the presence of location dependent and time varying channels. The concept of packet bundling is illustrated in Figure 1. Packets from multiple users or multiple packets from a single user may be combined together in a single time slot. Intuitively, the bundling will increase channel utilization. Furthermore, it will decrease the average queueing delay of the VoIP packets, since later arriving VoIP packets do not have to wait for their own time slot.

An important aspect to consider, however, is bundling packets from mobile stations with different channel conditions. Advanced adaptive wireless systems employ channel measurement and feedback-based rate control mechanisms such as the cdma2000 1xEV-DO system. In EV-DO, in order for the bundled packet to be received reliably, the adaptive coding rate for the entire packet should correspond to the worst channel condition among the bundled users. However this may cause the channel utilization gain from packet bundling to deteriorate due to the lowest coding rate. One way to tackle the issue is to combine packets with the same or similar channel condition. A problem we observe from this approach, however, is that at the time of bundling if there are not enough packets with the same or similar channel condition, the gain in the channel utilization may be marginal.

Another disadvantage of packet bundling is the number of retransmissions. Hybrid ARQ [37] is one of the important elements in advanced adaptive wireless systems. The packet will be retransmitted when negative acknowledgement is received or no acknowledgement is received when hybrid ARQ is used. The uncorrected error probability of a multi-user packet is always higher than the uncorrected error probability of a single user packet whose channel condition is the worst among the bundled packets in the multi-user packet. Retransmissions in hybrid ARQ perform better than retransmissions in higher layer (e.g. transmission layer) but retransmissions in hybrid ARQ still reduce channel utilization. If a multi-user packet has a significantly higher retransmission probability than retransmission probability of a single user packet, the gain in the channel utilization may be marginal.

Our contributions are as follows. We first show that the optimal packet bundling algorithm that either maximizes channel utilization or minimizes queueing delay is an NP-complete problem. Secondly, we propose a novel QoS and Channel aware packet Bundling (QCB) algorithm to jointly optimize QoS requirements and channel utilization with a simple approximation. QCB may defer the bundling decision a little within the QoS requirement. In the meantime, the time slot can be used for best effort traffic, significantly increasing channel utilization.

We compare the QCB scheme with bundling algorithms for two individual objectives, namely QoS Aware packet Bundling (QAB) and Channel Aware packet Bundling (CAB) schemes. We show that QCB enables high throughput as well as low delay, achieving an optimal trade-off of the two extremes.

Finally, we validate our proposed algorithms through OPNET simulation of a complete EV-DO implementation. The complete EV-DO implementation simulates both downlink and uplink at the same time. This implementation includes a 19 cell wraparound model, slow and fast fading, and single and dual antenna modes. Five different channel models are implemented based on 3GPP2 recommendation [54]. On-off traffic sources are generally used for VoIP traffic but sixteen-state traffic sources are used for modeling VoIP traffic in our implementation. Uplink and downlink hybrid ARQ are also implemented in our simulator.

The remainder of this dissertation is organized as follows. Chapter 2 discusses related work on scheduling algorithms for general wireless networks. Chapter 3 presents the background on the physical and MAC layer of the cdma2000 1x EV-DO Rev. A system, the cdma2000 EV-DO Rev. B system, and the Worldwide Interoperability for Microwave Access (WiMAX) system. Chapter 4 discusses the hardness of a packet bundling problem and proposes approximation algorithms such as QCB, QAB, and CAB. Then an EV-DO OPNET simulation setup, explanation of EV-DO simulator components, and evaluation results are described in Chapter 5. Chapter 6 concludes the dissertation and explains future work.

CHAPTER 2

RELATED WORK

A number of scheduling algorithms are available for wired networks including fair queueing [40], virtual clock [53], and earliest deadline first [20]. However, these are not readily applicable to the wireless environment that has location dependent and time varying channel characteristics. Although there have been attempts to incorporate channel dependent features into schedulers from wired networks [9], they cannot effectively exploit the time-varying multiuser diversity gain. Therefore, several new algorithms for wireless systems have been developed to exploit multiuser diversity [38], but the corresponding issue of bundling that we consider here has not been fully addressed.

A great deal of research has been done for physical and link layer issues of wireless networks. For example, [14] proposes link layer retransmission schemes for the Code Division Multiple Access (CDMA) channel. The work in [32] addresses the issue of bandwidth allocation with guaranteed QoS for wireless networks using a fluid version of the Gilbert-Elliott channel model. The problem of multimedia data transmission in Multi-Code CDMA wireless systems is discussed in [10, 26, 34], where the authors proposed efficient error recovery schemes for physical or link layers.

There are several studies on CDMA downlink scheduling. [29] modifies various wireline packet scheduling algorithms for CDMA networks and discusses the performance characteristics of them. They found that algorithms that exploit request size outperform those that do not, and discrete bandwidth allocation and management can degrade users' performance. The problem of CDMA downlink scheduling with a probabilistic delay requirement is considered in [2]. They showed that the Largest Weighted Delay First (LWDF) scheduling scheme provides good QoS for the settings of discrete rate and discrete scheduling intervals. In [42], the authors study a scheduling rule called the exponential rule where scheduling selects a packet based on the current state of the channel and the queues, and proves that it is throughput-optimal. A scheduling algorithm that combines channel-based and round-robin schedulings is proposed in [17] for CDMA systems.

In recent years, research and development efforts have increased on adaptive wireless systems where higher rate and power levels are allocated as the channel quality increases. This enables physical layer Adaptive Modulation and Coding (AMC) (see [1] for example). Relying on AMC, opportunistic schedulers select the user with the best channel quality to maximize the channel utilization. However, QoS may be violated for some users in such schemes. A scheduling algorithm that takes adaptive rate control based on the reverse link feedback is proposed in [24]. The work in [25] shows that Delay-Margin-based Scheduling nested with User-Channel-based Scheduling performs well both in delay and utilization metrics.

Simulation studies on EV-DO VoIP capacity are presented in [7, 52]. [44] shows the trade-off between system throughput and delay with opportunistic scheduling with analysis and simulation of the EV-DO system. The authors in [13] developed a soft algorithm that has an additional step for VoIP packets in order to check the channel condition, that is, whether the current data rate is larger than or equal to the average data rate. They demonstrated that Proportional Fair (PF) scheduling combined with the soft PF algorithm (PFsoft) shows the best performance over MAX rate algorithms.

A forward link scheduling algorithm that supports a MUP scheme is proposed in [51]. This algorithm first selects the user's packet whose priority is the highest according to the PF algorithm. Then only packets with the same channel quality become the candidates for bundling with a higher priority given to VoIP packets. Otherwise, a single user packet (SUP) will be sent. We name this algorithm PF-MUP and compare the performance of our proposed scheme with it in Chapter 5. The bundling ratio is limited by the available packets with the same channel condition. Our work differs from the above in that our scheduling algorithm jointly considers QoS and channel quality for packet bundling, and packets to mobile stations of different channel conditions may be bundled.

There are several studies on performance of an EV-DO reverse link [11, 19, 39]. [35] shows the efficient reverse link scheduling algorithm cdma2000 1xEV-DV system. [50] proposes a PF algorithm for the EV-DO reverse link.

The scheduling algorithm for next generation wireless technology like WiMAX (Worldwide Interoperability for Microwave Access) is presented in many papers. Weighted round robin (WRR) has been applied for WiMAX scheduling in [15] because round robin (RR) cannot guarantee QoS for different service classes. Worst-case fair weighted fair queuing was introduced in [27] to keep the delay bound. Delay threshold priority queuing (DTPQ) was proposed in [31] to give priority to real-time traffic when the head-of-line (HOL) packet delay exceeds a given delay threshold. In [8], lower priority class packets move to a higher priority class queue when the queueing time of the packet is close to deadline. In [48], real-time data classes share the common queue to reduce complexity. The authors in [30] developed the scheduler that allocates the bandwidth for minimum QoS requirements first and optimizes slot allocations for each connection. A few studies focus on maximizing the total system throughput [41, 45]. Linear programming is used for finding optimal subcarrier allocations in [36].

CHAPTER 3

BACKGROUND ON WIRELESS SYSTEMS

In this chapter, we give an overview of the physical and MAC layers of the cdma2000 1xEV-DO Rev. A system. We also give an overview of the multicarrier EV-DO system (EV-DO Rev. B) and the WiMAX system.

3.1 EV-DO Revision A

In a wireless system, signal strength is location dependent and time varying. It is subject to slow fading, fast fading, and interference from other signals, resulting in degradation of the Signal to Interference-plus-Noise Ratio (SINR) [46]. A high SINR yields a high data rate and low errors. A good SINR in cellular systems is achieved by using the optimum rate and power control mechanisms.

In EV-DO networks, both direct sequence spread spectrum and Time Division Multiple Access (TDMA) are used in the downlink and CDMA is used in the uplink [22, 23]. The downlink channel is a single broadband link shared by all users in a cell. One user is allowed to receive data in a single time slot. The base station estimates each user's channel condition based on the feedback from individual mobile station's measurements. The channel quality indication (CQI) feedback from the mobile station and the corresponding Adaptive Modulation and Coding (AMC) schemes are used as in many current and future wireless standards. In time slotted systems, the number of users supported

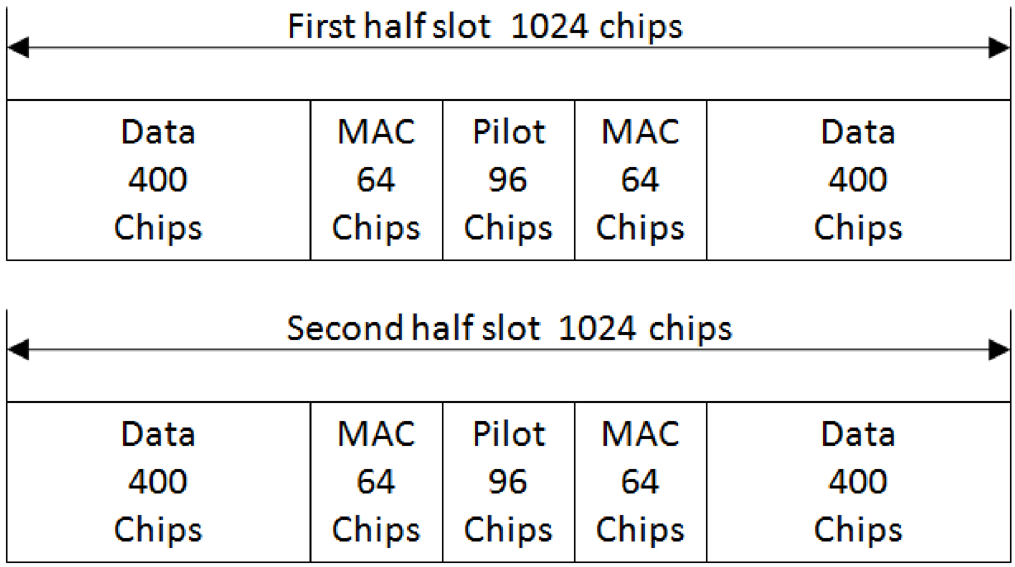

Figure 2: EV-DO downlink slot structure

is constrained theoretically. The maximum supported users or flows are limited by the number of time slots per second and packet arrival rates (Eq. (3.1)).

$$max\_supported\_users = \frac{no\_time\_slots/sec}{packet\_arrival\_rate/user} \tag{3.1}$$

For example, EV-DO revision A uses a 1.25 MHz bandwidth with direct sequence spread spectrum. The chip rate is 1.2288 Mchips/second, and the basic timing unit (slot) is 2048 chips. Figure 2 shows the EV-DO downlink slot structure. The downlink slot contains only three different channels. The pilot channel and the forward Medium Access Control (MAC) channel are always presented in a downlink slot. The forward traffic channel or the control channel can be included in the data part of the downlink slot.

The slot time is 1.667 ms, so there are 600 slots per second. Thus, it can serve a maximum of 600 packets per second (without bundling). Suppose a voice coder generates

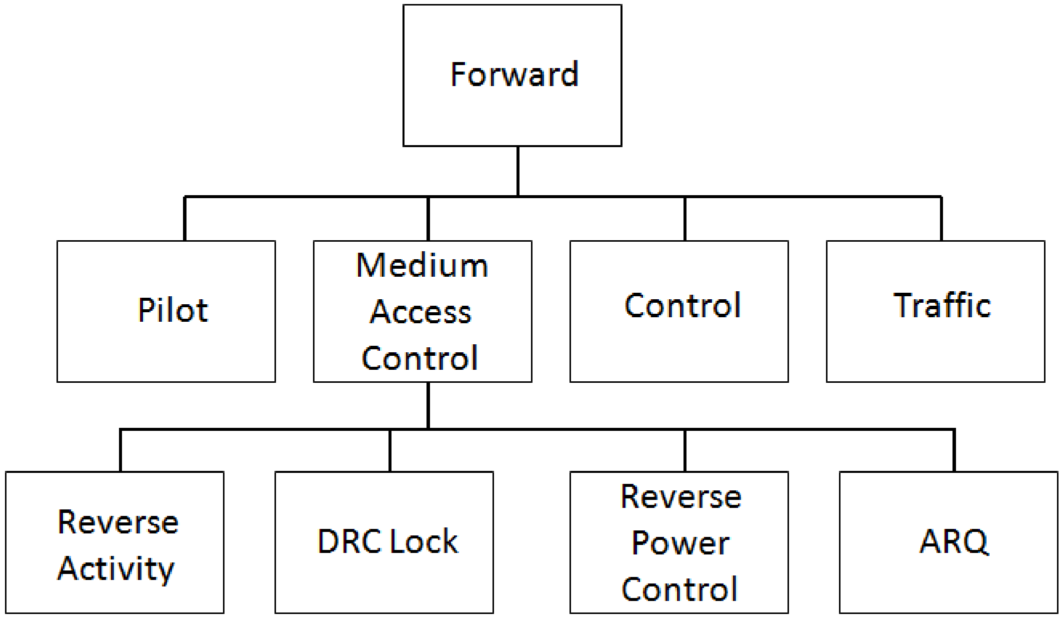

Figure 3: EV-DO downlink channel structure

a VoIP packet every 20 msec (i.e., a maximum of 50 packets/sec) and its average activity ratio is approximately 50%. Then, the maximum number of VoIP users supported in the EV-DO system is only 24 $(= 600/(50 \times 0.5))$. Meanwhile, the channel may go underutilized since the VoIP packet sizes are generally small (refer to Table 4), and not able to fill the entire time slot.

Figure 3 shows the EV-DO downlink (forward) channel structure. The EV-DO downlink channel consists of the pilot channel, the forward MAC channel, the forward traffic channel, and the control channel. The forward traffic channel is a packet-based, variable rate channel and carries user data. The control channel carries control messages and it may also carry user data. The pilot channel is used by the mobile station for initial acquisition, phase recovery, timing recovery, and maximal-ratio combining. An additional function of the pilot channel is to provide the mobile station with a means of predicting interference level for the purpose of forward data rate control of the forward data channel.

The forward MAC channel is composed of the reverse activity channel, the Data Rate Control Lock (DRC Lock) channel, the reverse power control channel, and the forward automatic repeat request (ARQ) channel. The reverse activity channel is used for indicating whether or not an uplink load exceeds a threshold. The mobile station adjusts the output power level of the pilot channel in response to each power control bit received on the reverse power control channel. The DRC Lock channel is used to indicate the channel state from the mobile station to the base station. The forward ARQ channel sends a positive acknowledgment or a negative acknowledgment in response to a physical layer packet.

Figure 4 shows the EV-DO uplink (reverse) channel structure. The EV-DO uplink channel structure is quite different when the mobile station is in a sleep state or an active state. In a sleep state, the mobile station uses the access channel. The access channel consists of the reverse pilot channel and the reverse data channel. In an active state, the mobile station uses the reverse traffic channel. The reverse traffic channel is composed of the primary pilot channel, the auxiliary pilot channel, the reverse MAC channel, the acknowledgement channel, and the reverse data channel. The reverse MAC channel consists of the Reverse Rate Indicator (RRI) channel, the DRC channel, and the Data Source Control (DSC) channel.

When the mobile station is transmitting a reverse traffic channel, it continuously transmits the primary pilot channel and the RRI channel. The reverse data channel is also

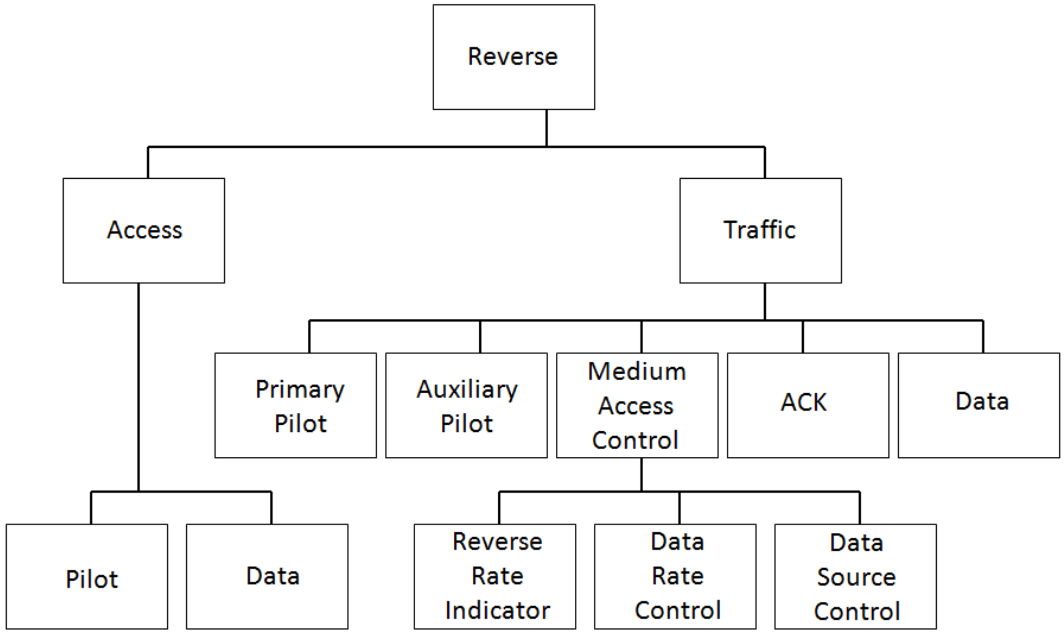

Figure 4: EV-DO uplink channel structure

a packet-based, variable rate channel and carries user data. The auxiliary pilot channel is used to provide uplink channel estimation for the large uplink physical layer packets. The RRI channel is used to indicate the data rate of the data channel being transmitted on the reverse traffic channel. The DRC channel is used by the mobile station to indicate to the base station the requested forward traffic channel data rate and the selected serving sector on the downlink channel. The ACK channel is used by the mobile station to inform the base station whether or not the physical layer packet transmitted on the forward traffic channel has been received successfully. The DSC channel indicates the sector of the base station from which the mobile station wishes to receive the forward traffic channel.

Figure 5 shows the EV-DO uplink subframe structure. One uplink subframe consists of four slots. Uplink channels are orthogonally spread by Walsh functions. Only the

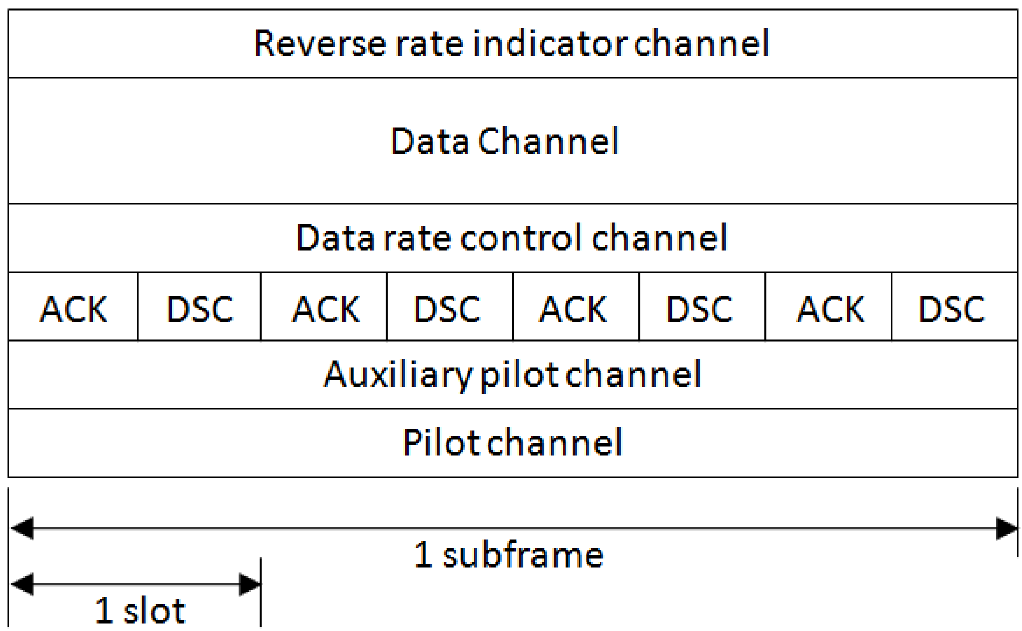

Figure 5: EV-DO uplink slot structure

ACK channel and DSC channel share slots using TDM.

The CQI from the mobile station is called the Data Rate Control (DRC) channel in the EV-DO system. The measured DRC value is fed back to the base station once every 1.667 msec using the reverse control channel. This slot size is short enough so that each user's channel quality stays approximately constant within one time slot, as it can be shown by computing the Doppler frequency of a mobile user at 2 GHz. In each time slot, one user is scheduled for transmission. Each user constantly reports to the base station its instantaneous channel capacity, i.e., the rate at which data can be transmitted if this user is scheduled for transmission.

Depending on the DRC feedback value, AMC schemes are adopted to support variable data rates for more reliable transmission for different mobile stations' channel

Table 1: Adaptive modulation and coding schemes in EV-DO Rev. A downlink

DRC	Data rate (kbps)	Bits per slot	Code Rate	Modulation
1	38.4	64	1/4	QPSK
2	76.8	128	1/4	QPSK
3	153.6	256	1/4	QPSK
4	307.2	512	1/4	QPSK
5	307.2	512	1/4	QPSK
6	614.4	1024	1/4	QPSK
7	614.4	1024	1/4	QPSK
8	921.6	1536	3/8	8-PSK
9	1228.8	2048	1/2	QPSK
10	1228.8	2048	1/2	16-QAM
11	1843.2	3072	1/2	8-PSK
12	2457.6	4096	1/2	16-QAM
13	1586.0	2560	1/2	16-QAM
14	3072.0	5120	1/2	16-QAM

environments. Modulation schemes are closely related to physical packet size. That is, if the physical packet size is less than or equal to 2048, QPSK is used; if the physical packet size is 3072, 8PSK is used; and if the physical packet size is 4096 or 5120, 16QAM is used. Table 1 shows modulation and coding options in the EV-DO Rev. A downlink.

On the reverse link where multiple mobile stations send transmissions concurrently, the EV-DO system capacity is limited by the interference level measured by RoT (Rise over Thermal). The RoT value is the total received power divided by the thermal noise value. The sector RoT value should be less than a threshold (99% of the time less than 7 dB is recommended) to stabilize the system. The base station measures the sector

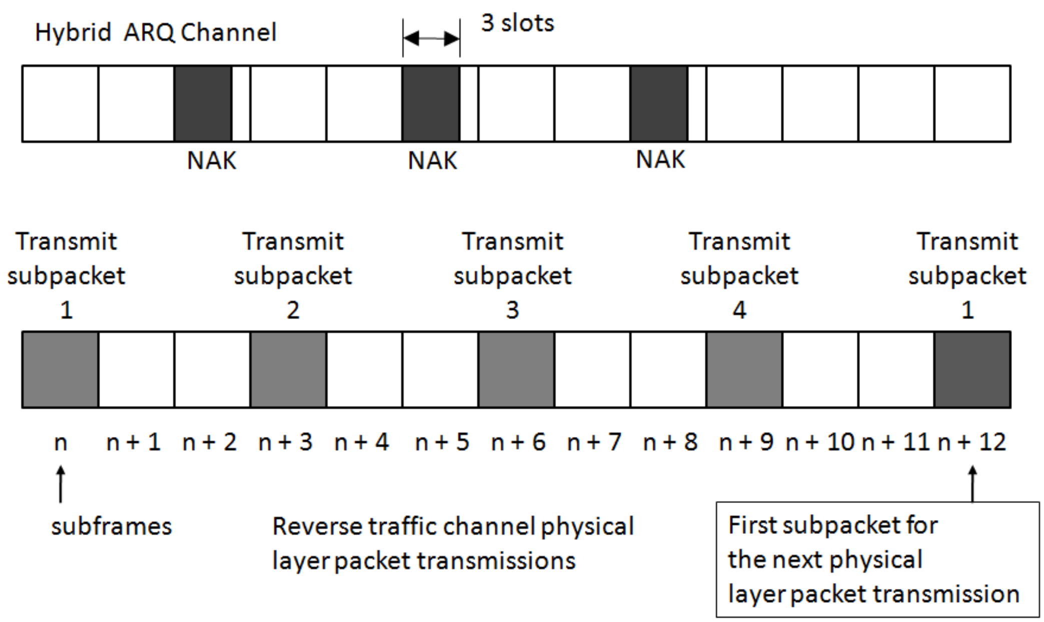

Figure 6: Reverse link Hybrid ARQ

RoT value and informs mobile stations with the RAB (Reverse Activity Bit) whether RoT is high or not, so that the uplink rate can be controlled.

In the EV-DO uplink link, four slots make one subframe and one frame consists of four subframes. One subframe sends one subpacket which is a basic sending unit in the EV-DO reverse link. In EV-DO rev. 0, Hybrid ARQ is not used because one reverse link frame uses 16 consecutive slots. Hybrid ARQ is used in EV-DO rev. A. Figure 6 illustrates usage of Hybrid ARQ in a reverse link. In EV-DO rev. A, four subpackets are not sent consecutively to use hybrid ARQ.

Speech is encoded using a variable rate vocoder via the Enhanced Variable Rate Codec (EVRC) that generates VoIP traffic depending on speech activity. Since a frame duration is fixed at 20 ms, the number of bits per frame varies according to the traffic rate.

171 bits, 80 bits, 40 bits, and 16 bits are generated for full, half, 1/4, and 1/8 rate coding, respectively [58, 60]. For silence periods, a 1/8 rate packet is sent in one out of every 12 slots (i.e., every 240 ms) for background comfort noise. A more detailed description can be found in cdma2000 specification [55].

The multi-user packet is a new feature of EV-DO Rev. A and it is designed to support more users per given time period. It is very important to support more users per given time period in real-time applications like VoIP because their delay deadlines can be better met with multi-user packets. The VoIP application is the best fit for the multi-user packet, since the VoIP packets are generated frequently (every 20 ms) and their sizes are small.

A bundled packet can be recognized by the preamble of the physical layer packet and the MAC header. Figure 7 shows single as well as bundled packet formats. A single user packet bundling of n packets (SUP-multiplex, (b)) has n bytes of header for the packet lengths of individual packets. With multi-user packet bundling (MUP-multiplex, (c)), 2 bytes are necessary for each packet to identify the mobile station within a sector and the packet length. All mobile stations decapsulate a frame as if it were a multicast packet, to extract the packet portion destined to itself.

To be specific, it works as follows. In a single user packet (SUP-simplex or SUP-multiplex), a preamble of the physical layer packet is used to hold the MAC Index that represents the destination of the packet. In a multi-user packet (MUP), a preamble of the physical layer packet is used to hold the modulation scheme and the MAC header is used to hold the MAC Index and packet size of individual packets(see Figure 7 (d)).

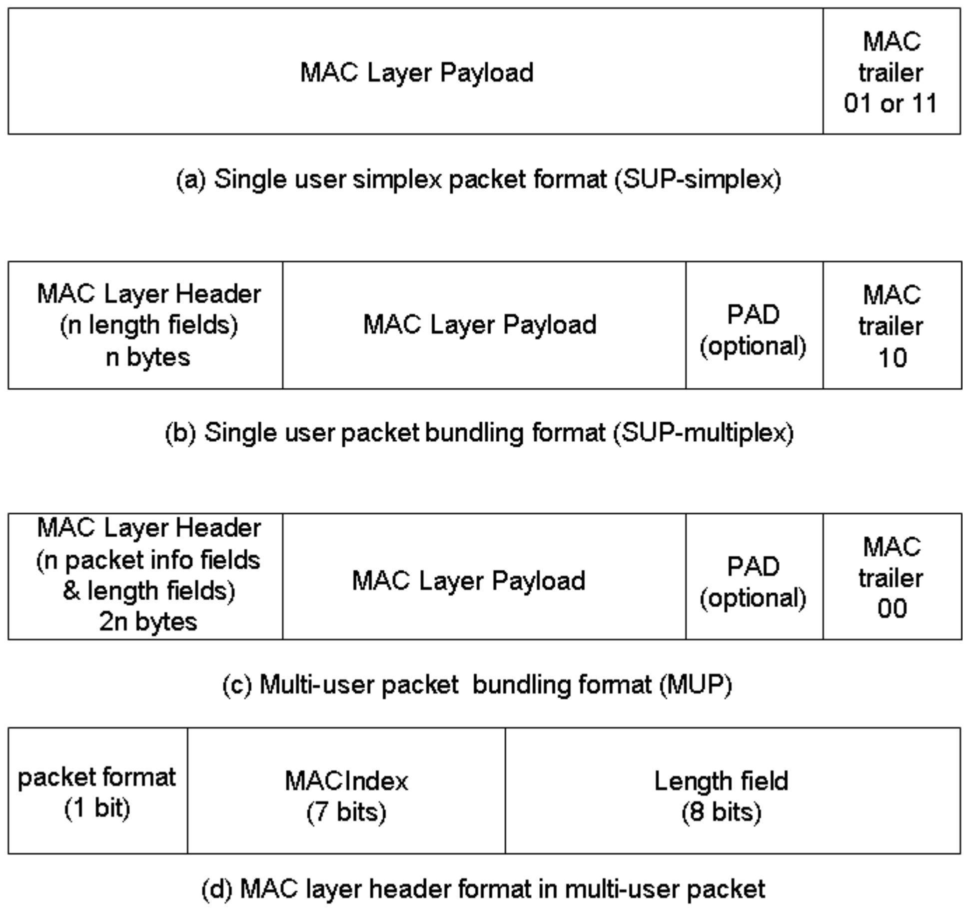

Figure 7: Packet formats.

Table 2 shows compatible single user packet formats and multi-user packet formats on DRC values. More compatible formats are supported by the multi-user packet because a smaller packet can reduce packet error rate. An ARQ mechanism is also used in the multi-user packet. The base station retransmits the multi-user packet until all mobile stations send positive acknowledgements to the base station or duration of the multi-user packet ends. When we compare uncorrected error probability of the multi-user packet to uncorrected error probability of the single user packet for the same DRC value, uncorrected error probability of the multi-user packet is always bigger than uncorrected error probability of the single user packet because the multi-user packet needs more acknowledgements than the single user packet. If packet size is small enough to use a lower rate packet format, the lower rate packet format will reduce an error rate and a lower error rate will reduce the number of retransmissions.

Table 2: Transmission formats in EV-DO Rev. A downlink

DRC index	Data rate (kbps)	Single user transmission format (packet size, duration, preamble)	Multi-user transmission format (packet size, duration, preamble)
1	38.4	(128, 16, 1024), (256, 16, 1024), (512, 16, 1024), (1024, 16, 1024)	None
2	76.8	(128, 8, 1024), (256, 8, 1024), (512, 8, 1024), (1024, 8, 1024)	None
3	153.6	(128, 4, 256), (256, 4, 256), (512, 4, 256), (1024, 4, 256)	(128, 4, 256), (256, 4, 256), (512, 4, 256), (1024, 4, 256)
4	307.2	(128, 2, 128), (256, 2, 128), (512, 2, 128), (1024, 2, 128)	(128, 4, 256), (256, 4, 256), (512, 4, 256), (1024, 4, 256)
5	307.2	(512, 4, 128), (1024, 4, 128), (2048, 4, 128)	(128, 4, 256), (256, 4, 256), (512, 4, 256), (1024, 4, 256), (2048, 4, 128)
6	614.4	(128, 1, 64), (256, 1, 64), (512, 1, 64), (1024, 1, 64)	(128, 4, 256), (256, 4, 256), (512, 4, 256), (1024, 4, 256)
7	614.4	(512, 2, 64), (1024, 2, 64), (2048, 2, 64)	(128, 4, 256), (256, 4, 256), (512, 4, 256), (1024, 4, 256), (2048, 4, 128)

Table 2: – Continued.

DRC index	Data rate (kbps)	Single user transmission format (packet size, duration, preamble)	Multi-user transmission format (packet size, duration, preamble)
8	921.6	(1024, 2, 64), (3072, 2, 64)	(128, 4, 256), (256, 4, 256), (512, 4, 256), (1024, 4, 256), (2048, 4, 128), (3072, 2, 64)
9	1228.8	(512, 1, 64), (1024, 1, 64), (2048, 1, 64)	(128, 4, 256), (256, 4, 256), (512, 4, 256), (1024, 4, 256), (2048, 4, 128)
10	1228.8	(4096, 2, 64)	(128, 4, 256), (256, 4, 256), (512, 4, 256), (1024, 4, 256), (2048, 4, 128), (3072, 2, 64), (4096, 2, 64)
11	1843.2	(1024, 1, 64), (3072, 1, 64)	(128, 4, 256), (256, 4, 256), (512, 4, 256), (1024, 4, 256), (2048, 4, 128), (3072, 2, 64)
12	2457.6	(4096, 1, 64)	(128, 4, 256), (256, 4, 256), (512, 4, 256), (1024, 4, 256), (2048, 4, 128), (3072, 2, 64), (4096, 2, 64)
13	1586.0	(5120, 2, 64)	(128, 4, 256), (256, 4, 256), (512, 4, 256), (1024, 4, 256), (2048, 4, 128), (3072, 2, 64), (4096, 2, 64), (5120, 2, 64)

Table 2: – Continued.

DRC index	Data rate (kbps)	Single user transmission format (packet size, duration, preamble)	Multi-user transmission format (packet size, duration, preamble)
14	3072.0	(5120, 1, 64)	(128, 4, 256), (256, 4, 256), (512, 4, 256), (1024, 4, 256), (2048, 4, 128), (3072, 2, 64), (4096, 2, 64), (5120, 2, 64)

3.2 Multicarrier EV-DO

Multicarrier EV-DO is backward compatible with 1xEV-DO Rev. A systems and specifies up to a 20 MHz wide system with each carrier 1.25 MHz wide and terminals supporting one or more carriers. A multicarrier operation achieves higher efficiencies relative to a single-carrier because a multicarrier operation can exploit channel frequency selectivity, improve transmit efficiencies on the reverse link, and use adaptive load balancing across carriers.

New concepts introduced in a multicarrier EV-DO are as follows [4]:

- multilink radio link protocol (ML-RLP) for channel aggregation;
- symmetric multicarrier mode and asymmetric multicarrier modes of operation;
- multicarrier reverse traffic channel MAC;
- adaptive load balancing;
- flexible duplex carrier assignment.

If EV-DO Rev. A channel cards are used in a multicarrier EV-DO network, ML-RLP is required when a terminal is assigned carriers on channel cards that do not communicate with each other and operate an independent scheduler. ML-RLP is only necessary on the forward link and is not required if a single scheduler is responsible for scheduling packets across multiple carriers. The base station controller splits packets into segments and sends the segments to the base stations. Each channel card independently schedules the segments and sends them to mobile stations and ML-RLP aggregates and reassembles them to restore the original packets.

Multicarrier EV-DO supports a symmetric multicarrier mode, a basic asymmetric multicarrier mode, and an enhanced asymmetric multicarrier mode. The number of forward channels is equal to the number of reverse channels in the symmetric multicarrier mode. The symmetric mode of operation may be used for applications with symmetric data rate requirements on the forward and reverse links. The symmetric multicarrier mode enables a multicarrier operation using an aggregation of EV-DO Rev. A channel cards. If application uses more forward links than reverse links, the asymmetric multicarrier mode can be used. The asymmetric mode of operation results in reduced reverse link overhead as the pilot channels for the additional reverse link carriers are not transmitted. In the basic asymmetric multicarrier mode, a single reverse channel may carry feedback for more than one forward channel using unique long codes for each feedback channel. In an enhanced asymmetric multicarrier mode, a single reverse channel may carry feedback for up to four forward channels using the same long code.

The main features of the multicarrier reverse traffic channel MAC are flow to carrier mapping, reverse link load balancing and efficient reverse link transmission. The reverse link MAC specifies fixed allocation flows and elastic flows. Fixed allocation flows have high priority and can use network resources up to the limit. Elastic allocation flows use excess resources when the demands of all fixed allocation flows have been met. The mobile station attempts to achieve efficient transmission while achieving load balancing by favoring the reverse link carrier with the least interference.

The base station can achieve packet-based load balancing on the forward link. If the near uniform load is maintained across carriers on the forward link, the base station can assign carriers to mobile stations to maximize capacity utilization and spectral efficiency gains. The mobile station can also achieve carrier-based load balancing on the reverse link. If nearly equal interference is maintained across carriers on the reverse link, the mobile station can pick the instantaneously best carrier for transmitting the packet.

Any reverse channel from a band class can be coupled with any forward channel from the same band class or with a forward channel from another band class based on the capabilities of the mobile station with flexible duplex spacing. Flexible duplex spacing also allows using a reverse channel from a paired spectrum with forward channels from both the paired spectrum as well as the unpaired spectrum providing operators further flexibility in spectrum allocation.

3.3 Worldwide Interoperability for Microwave Access (WiMAX)

The WiMAX system and Long Term Evolution (LTE) system are currently deploying in the United States [16]. The current WiMAX system is based on IEEE 802.16e (802.16-2005) specification. There are five different physical layers, WirelessMAN-SC, WirelessMAN-SCa, WirelessMAN-OFDM, WirelessMAX-OFDMA, and WirelessHUMAN. They are defined in IEEE 802.16e specification [47, 56, 57]. The WirelessMAN-SC physical layer uses a single carrier in line of sight (LOS) environments. It is targeted for operation in the 10-66 GHz frequency band. All other physical layers support non-line of sight (NLOS) environments in frequency bands between 2-11 GHz. The WirelessMAN-OFDM physical layer uses the 256 point fast Fourier transform (FFT) orthogonal frequency-division multiplexing (OFDM). The WirelessMAN-OFDMA physical layer uses 128, 256, 512, 1024, or 2048 point FFT orthogonal frequency division multiple access (OFDMA) based on system bandwidth. Most of the current WiMAX products implement the OFDMA physical layer [43]. Multiple subscribers use a TDMA method to share the wireless channel in this physical layer.

The key features of the WiMAX system are as follows:

- time division duplex (TDD), frequency division duplex (FDD), and half-duplex FDD (H-FDD) are supported;
- several frame sizes (2 ms - 20 ms) are supported;
- supported bandwidth ranges are from 1.25 MHz to 28 MHz;
- multiple modulation and coding schemes are supported; QPSK, 16QAM, and 64QAM are supported modulation schemes; convolutional codes, convolutional turbo codes, block turbo codes, and low density parity check codes are supported coding schemes;
- downlink and uplink hybrid ARQ are supported;
- multi-antenna operations are supported including the advanced antenna subsystem (AAS) mode; the open-loop space time coding (STC) modes; the closed-loop

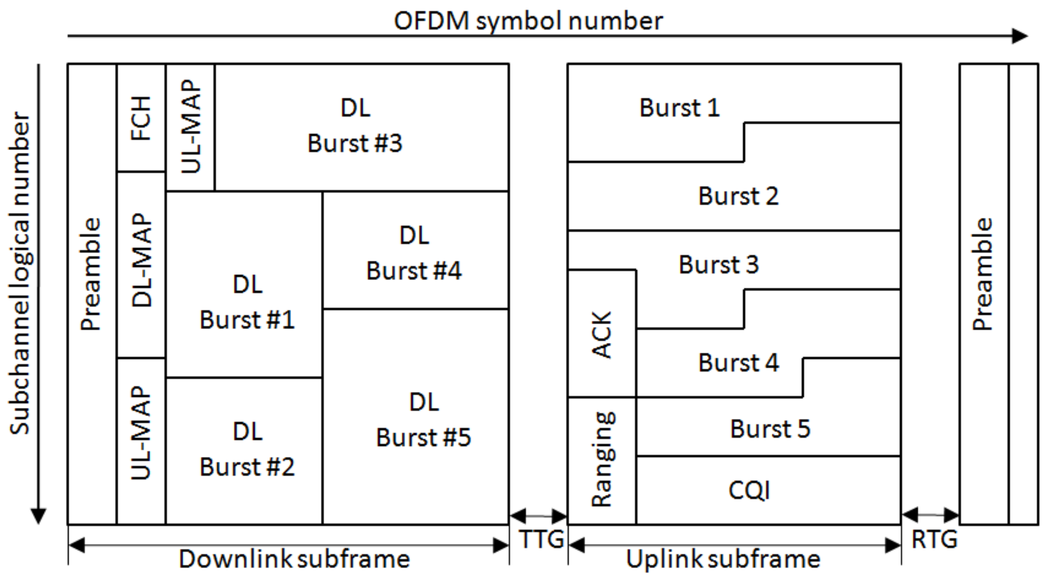

Figure 8: WiMAX frame structure

multiple-input multiple-output (MIMO) modes; and the uplink space division multiple access (SDMA);

- multiple QoS classes can be defined; those multiple QoS classes are suitable for combination of data, voice, and video services;
- efficient multicast-broadcast transmission schemes using single frequency network (SFN) concepts are supported;
- fast scheduling is possible based on flexible CQI.

Figure 8 shows the WiMAX frame structure. There are several options for frame size. Generally, each frame is configured to be 5 ms long and is time division duplexed into downlink and uplink subframes. There are two gaps between a downlink subframe and an uplink subframe. A transmit-receive transition gap (TTG) occurs when the base station switches from transmit to receive mode and a receive-transmit transition gap (RTG) occurs when the base station switches from receive to transmit mode. The mobile stations use these gaps in the opposite way.

At the beginning of each frame, downlink control information is transmitted consisting of the preamble, the frame control header (FCH), the downlink map (DL-MAP), and the uplink map (UL-MAP). The preamble is used for time synchronization and downlink channel estimation. The FCH carries information of the subchannels being used by the sector in the current frame, coding, and size of the DL-MAP. The DL-MAP and the UL-MAP define the burst start time, burst end time, modulation types, and forward error control (FEC) information. The DL-MAP and UL-MAP consist of information elements. The size of the DL-MAP and the UL-MAP are proportional to the number of downlink and uplink users scheduled in that frame. The DL-MAP and UL-MAP are modulated with reliable modulation and coding such as BPSK or QPSK. The DL-MAP and UL-MAP usually require two or four repetitions depending on the channel condition.

Uplink control channels are composed of the ranging channel, the CQI channel, and the acknowledge (ACK) channel. The ranging channel provides the random access capability for initial entry, timing adjustment, periodic synchronization, bandwidth request, and hand-off entry. The uplink CQI or fast feedback channel is used by a mobile station to report the measured carrier to interference and noise ratio (CINR) back to the base station. This information is used for functions such as selecting the downlink modulation and coding rates. Another uplink control channel, the uplink ACK channel, transports ACK and negative acknowledge (NACK) feedback for the downlink hybrid ARQ data transmission and occupies half an uplink slot.

The WiMAX system supports full usage of subchannels (FUSC) and partial usage of subchannels (PUSC). In FUSC, all subcarriers have to be allocated in one cell or sector. PUSC is designed so that only a set of the subcarriers can be allocated to one cell or sector, depending on the traffic conditions, and to reduce interference. Other options include tile usage subchannelization (TUSC) and optional full usage subchannelization (O-FUSC) for downlink, and optional partial usage subchannelization (O-PUSC) for uplink.

The burst allocation is important in WiMAX scheduling. The shape of the downlink burst should be rectangular and slots will be wasted if the allocation algorithm is not efficient. In an uplink burst, the number of subchannels should be minimized because of power limitation. The shape of the uplink burst has no limitation.

CHAPTER 4

MULTIPLE PACKET BUNDLING

In this chapter, we first show the hardness of the bundling problem [33]. We then discuss channel error probability of multi-user packets. We finally discuss QAB, CAB, and QCB algorithms that approximately optimize QoS requirements, utilization, and both QoS and channel utilization, respectively. The analytic models of QAB and CAB are described in [12].

Efficient packet bundling may not always be beneficial due to diverse channel conditions observed at the mobile stations. For EV-DO, when multiple packets are bundled, the modulation is used for a destination with the worst channel condition. A modulation for a worse channel means a lower effective bit rate to compensate for a higher error rate. Thus, a multi-user packet may sacrifice data rates for all users based on the worst channel. Therefore, bundling is not a trivial task and careful decisions need to be made.

4.1 Hardness of the Problem

We show that given a set of packets, finding a packet bundling assignment with a minimal number is NP-complete.

<u>Packet bundling assignment problem:</u> Given a set of packets of varying size, time slot, and an integer b, is there a bundling assignment or partition of the packets into time slots, with a partition size less than b?

To prove that it is NP-complete, we prove that the following Bin Packing Problem that is known to be NP-complete [21, 28] can be reduced to our packet bundling problem in polynomial time.

Bin packing problem: Find a partition and assignment of a set of objects such that a constraint is satisfied or an objective function is minimized (or maximized). Specifically, determine how to put the most objects in the least number of fixed space bins. More formally, given a bin size V and a list $a_1, \ldots, a_n$ of sizes of the items to pack, find an integer B and a B-Partition of a set $S_1 \cup \cdots \cup S_B$ such that $\sum_{i \in S_k} a_i \leq V$ for all $k = 1, \ldots, B$.

The reduction is trivial in that the object and bin sizes correspond to the packet size and time slot interval, respectively, and the partition relates to the packet bundle assignment. Notice that this problem is easier than the problem of finding an optimal or minimal packet partition. If a minimal partition is known, simply computing its size and comparing it to B allows us to answer the question.

4.2 Uncorrected Error Probability of a Multi-user Packet

The hybrid ARQ is used in EV-DO rev. A downlink and uplink. In the EV-DO system, a downlink hybrid ARQ is introduced from EV-DO rev. 0 and it shows substantial capacity gains [49]. The channel error probability of the hybrid ARQ depends on channel conditions. Before we discuss comparisons between uncorrected error probability of the multi-user packet and that of the single user packet, we define basic probabilities.

- $p_{i,j}$ is channel error probability of j-th transmission using data rate control index i which is the best data rate control index in the current channel condition. Channel error probability $p_{i,j}$ is always greater than channel error probability $p_{i,k}$ if $k > j$ because of hybrid ARQ.

- $p^k_{i,j}$ is channel error probability of j-th transmission using data rate control index i which is not the best data rate control index in the current channel condition. The data rate control index k is the best data rate control index in the current channel condition. Channel error probability $p^k_{i,j}$ is always less than channel error probability $p_{i,j}$ because reliability of packet delivery increases if a lower data rate control index is used in the same channel condition.
- $d(n)$ is the best rate of n-th packet in the current channel condition.

If the multi-user packet is composed of two single user packets with data rate control index i, the data rate control index of multi-user packet is i. An uncorrected error probability of multi-user packet P^m_1 follows after the first transmission.

$$P^m_1 = 1 - (1 - p_{i,1})(1 - p_{i,1}) = 2p_{i,1} - {p_{i,1}}^2 > p_{i,1} \tag{4.1}$$

The uncorrected error probability of the multi-user packet is greater than the uncorrected error probability of the single user packet because the uncorrected error probability of the single user packet is $p_{i,1}$.

The following equation is an uncorrected error probability of multi-user packet P^m_2 after the second transmission. The uncorrected error probability of the single user packet is $p_{i,1}p_{i,2}$.

$$P^m_2 = 1 - (1 - p_{i,1}p_{i,2})(1 - p_{i,1}p_{i,2}) = 2p_{i,1}p_{i,2} - p_{i,1}{p_{i,2}}^2 > p_{i,1}p_{i,2} \tag{4.2}$$

The uncorrected error probability of the multi-user packet is also greater than the uncorrected error probability of the single user packet in the second transmission.

In the s-th transmission case, the uncorrected error probability of the single user packet is $\prod_{t=1}^{s} p_{i,t}$. The uncorrected error probability of multi-user packet P_s^m follows after the s-th transmission.

$$P_s^m = 1 - (1 - \prod_{t=1}^{s} p_{i,t})(1 - \prod_{t=1}^{s} p_{i,t}) > \prod_{t=1}^{s} p_{i,t} \tag{4.3}$$

This result shows that the multi-user packet has a higher uncorrected error probability compared to the single user packet.

If the multi-user packet is composed of two single user packets with a data rate control index i and k respectively ($i < k$), then the data rate control index of the multi-user packet is i. An uncorrected error probability of multi-user packet P_1^m follows after the first transmission.

$$P_1^m = 1 - (1 - p_{i,1})(1 - p_{i,1}^k) = p_{i,1} + p_{i,1}^k - p_{i,1}p_{i,1}^k > p_{i,1} \tag{4.4}$$

The uncorrected error probability of the multi-user packet is greater than the uncorrected error probability of the single user packet whose data rate control index is the worst among bundled packets because the uncorrected error probability of the single user packet is $p_{i,1}$.

The following equation is an uncorrected error probability of multi-user packet P_2^m after the second transmission. The uncorrected error probability of the single user packet is $p_{i,1}p_{i,2}$.

$$P_2^m = 1 - (1 - p_{i,1}p_{i,2})(1 - p_{i,1}^k p_{i,2}^k) = p_{i,1}p_{i,2} + p_{i,1}^k p_{i,2}^k - p_{i,1}p_{i,2}p_{i,1}^k p_{i,2}^k > p_{i,1}p_{i,2} \tag{4.5}$$

The uncorrected error probability of the multi-user packet is also greater than the uncorrected error probability of the single user packet in the second transmission.

In the s-th transmission case, the uncorrected error probability of the single user packet is $\prod_{t=1}^{s} p_{i,t}$. The uncorrected error probability of the multi-user packet P_s^m follows after the s-th transmission.

$$P_s^m = 1 - (1 - \prod_{t=1}^{s} p_{i,t})(1 - \prod_{t=1}^{s} p_{i,t}^k) > \prod_{t=1}^{s} p_{i,t} \tag{4.6}$$

This result shows that the multi-user packet has a higher uncorrected error probability compared to the single user packet.

If the multi-user packet is composed of q single user packets with a data rate control index $d(n)$ respectively, the data rate control index of the multi-user packet is $L = \min_{n=1}^{q} d(n)$. The uncorrected error probability of the multi-user packet P_1^m follows after the first transmission.

$$P_1^m = 1 - \prod_{n=1}^{q} (1 - p_{L,1}^{d(n)}) > p_{L,1} \tag{4.7}$$

The uncorrected error probability of the multi-user packet is greater than the uncorrected error probability of the single user packet whose data rate control index is the worst among bundled packets because the uncorrected error probability of single user packet is $p_{L,1}$.

The following equation is an uncorrected error probability of the multi-user packet P_2^m after the second transmission. The uncorrected error probability of the single user packet is $p_{L,1}p_{L,2}$.

$$P_2^m = 1 - \prod_{n=1}^{q} (1 - p_{L,1}^{d(n)} p_{L,2}^{d(n)}) > p_{L,1}p_{L,2} \tag{4.8}$$

The uncorrected error probability of the multi-user packet is also greater than the uncorrected error probability of the single user packet in the second transmission.

In the s-th transmission case, the uncorrected error probability of the single user packet is $\prod_{t=1}^{s} p_{L,t}$. The uncorrected error probability of multi-user packet P_s^m follows after the s-th transmission.

$$P_s^m = 1 - \prod_{n=1}^{q}(1 - \prod_{t=1}^{s} p_{L,t}^{d(n)}) > \prod_{t=1}^{s} p_{L,t} \tag{4.9}$$

This result shows that the multi-user packet has a higher uncorrected error probability compared to the single user packet whose data rate control index is the worst among bundled packets.

We reduce the retransmission of a multi-user packet using the following method. If a member of a multi-user packet is sent using a single user packet format, the maximum retransmission count of this single user packet is C_s. The C_s is less than or equal to the maximum retransmission count of the multi-user packet. If the retransmission count of the multi-user packet is greater than or equal to C_s, the scheduler assumes that acknowledgement is received even if negative acknowledgement is received. For example, the multi-user packet is composed of two member packets and the maximum retransmission count of the first member is 2 and that of the second member is 4. The maximum retransmission count of the multi-user packet is 4. After two transmissions, negative acknowledgement is received for the first member and positive acknowledgement is received for the second member. In this case, the scheduler stops sending the multi-user packet because the scheduler assumes that all positive acknowledgements are received for both member packets.

4.3 QoS Aware Packet Bundling (QAB)

With QoS aware scheduling, a packet with the longest delay, p_u^{*d}, will be selected for service as below, when packet bundling is not used.[1] [2]

$$p_u^{*d} = \arg\max_u d(p_u) \tag{4.10}$$

where $d(p_u)$ is the delay of a packet of user u.

The above equation can be extended to a set of bundled packets B^{*d} as follows:

$$\begin{aligned} B^{*d} &= \arg\max_{B^d} |B^d| \sum_{u \in B^d} d(p_u) && (4.11) \\ & \text{such that} \sum_{u \in B^d} L(p_u)/AMC(u) \leq T && (4.12) \end{aligned}$$

where T is the size of time slot, $L(p_u)$ is the size of user u's packet, and $AMC(u^*)$ is the AMC rate of the user with the worst channel condition. The bundle set is composed of packets where the sum of their delays is the longest. The constraint is to ensure the set of packets are bundled within a time slot when adaptive coding and modulation is applied. As discussed earlier, finding such a set of packets for bundling is an NP-complete problem. Thus, we use an approximation algorithm called QAB as shown in Algorithm 1. The input is a queue of VoIP packets and the output is a packet bundling assignment. The QAB algorithm is similar to the Earliest Deadline First (EDF) algorithm. Both algorithms are designed to serve real-time applications like VoIP. When there is not a real-time packet to bundle, a BE packet will be sent along. The packet size of BE traffic is often big enough for an entire time slot. For handling BE traffic, we use the PF algorithm for fairness.

[1]A packet will be dropped if the delay is greater than the requirement.

[2]We discuss QoS mainly in the context of a delay parameter. However, it can be easily applied to other QoS parameters.

4.4 Channel Aware Packet Bundling (CAB)

As the channel condition varies depending on the time and the location of a user, the transmission data rate that a base station can send to a mobile station changes, depending on the channel condition. Opportunistic scheduling that maximizes the channel utilization is to choose a packet p_u^{*c} whose channel rate $CQI(u)$ is the maximum. That is

$$p_u^{*c} = \arg\max_u CQI(u) \tag{4.13}$$

A natural extension of the scheme to packet bundling is to choose the set of packets , B^c that gives the maximum sum of CQIs within the time slot.

$$B^{*c} = \arg\max_{B^c} \sum_{u \in B^c} CQI(u) \tag{4.14}$$

subject to $\sum_{u \in B^c} L(p_u)/AMC(u) \leq T$.

Since an algorithm that finds such a set of packets is NP-complete, a heuristic algorithm can be used to approximate the maximum rate bundling. A sketch of the CAB algorithm is shown in Algorithm 2. In order to better utilize the channel, packets from the same or similar channel conditions are bundled together. Since the worst AMC rate of the bundled packets will be the same or similar to the users' channel condition, the bundling ratio is high, resulting in efficient channel utilization. Also, for efficient handling of small size real-time packets, it defines a bundling threshold, B_{thresh}, which is the minimum real-time data size or the number of packets that should be bundled. By limiting B_{thresh} to a small number, packets can be scheduled without being deferred, particularly when there are few real-time packets to be bundled. A large B_{thresh} forces the real-time packets to be bundled with a high bundling ratio, in order to better share the channel with BE traffic.

Note that since the objective is only to maximize the utilization, it is impossible to provide any delay guarantees. Thus, a packet may wait for a long time for a chance of bundling. Real-time packets that exceed the maximum allowed delay, or packets arriving when the queue is full, will be dropped.

4.5 QoS and Channel Aware Packet Bundling (QCB)

The QCB scheme seeks to gain the benefits of both the QAB and CAB methods. The main objectives of the QCB scheme are first to satisfy delay requirements of real-time packets, and then to utilize the wireless channel efficiently. We first define a maximum allowed delay, D_{thresh} that scheduling of real-time packets can be deferred in the queue without sacrificing QoS. If there are packets whose delays are greater than or equal to D_{thresh}, those packets should be bundled first in order to meet the delay requirement. When the packets' delays are less than D_{thresh}, they attempt to utilize the channel efficiently by gathering packets of similar channel conditions that can be bundled together. The deferred scheduling of real-time packets makes room for opportunistic scheduling. For our experiments in Chapter 5, we set D_{thresh} to be 25 ms, and B_{thresh} to be 4. We have varied the parameters and found that those values provide a good tradeoff between QAB and CAB. When D_{thresh} is 0, the QCB algorithm is the same as QAB. When D_{thresh} is bigger than the delay threshold of dropping a packet, the QCB algorithm is reduced to the CAB algorithm. The pseudo-codes of the QCB algorithm are illustrated in Algorithm 3.

To fully understand the comparative benefits of QAB, CAB, and QCB, we now study them from simulation results from a full EV-DO implementation.

Algorithm 1 QoS Aware packet Bundling (QAB)

```
if no VoIP packet
    Add a BE packet to a SUP
else
    Remove the oldest VoIP packet from the queue and make a MUP
    while the queue is not empty and the MUP is not full
        if the coding of next oldest VoIP packet is not compatible with the MUP
            if the MUP is too small to add the VoIP packet
                Exit the while loop
            else
                Change the MUP with the new coding
                Remove the VoIP packet from the queue and add it to the MUP
            end if
        else
            Remove the next oldest VoIP packet and add it to the MUP
        end if
    end while
    if the MUP is not full
        Add a BE packet to the MUP
    end if
end if
```

Algorithm 2 Channel Aware packet Bundling (CAB)

```
if no VoIP packet
    Add a BE packet to a SUP
else
    while the queue is not empty and the MUP is not full
        Remove a VoIP packet from the queue and add it to the MUP
        with corresponding coding format
    end while
    foreach defined MUP format
        if the number of VoIP packets ≥ B_thresh
            Create a MUP using VoIP packets
            if the MUP is not full
                Add a BE packet to the MUP
            end if
            Exit foreach loop
        end if
    end foreach
    if no MUP created
        Add a BE packet to a SUP
    end if
end if
```

Algorithm 3 QoS and Channel aware packet Bundling (QCB)

```
if delay of a VoIP packet ≥ D_thresh
    Run QAB algorithm
else
    Run CAB algorithm
end if
```

CHAPTER 5

EVALUATION

We have implemented the complete cdma2000 1xEV-DO system recommended by the 3GPP2 evaluation methodology [54] using OPNET. We have built an EV-DO link layer because OPNET does not support an EV-DO link layer. To the best of our knowledge, it is the first simulation that includes both a downlink and an uplink of the EV-DO system.[1] Even though our work is focused on downlink resource allocation and scheduling, the performance of the downlink is tightly coupled with uplink feedback and control mechanisms. Therefore, our implementation provides practical insights from the interplay of both links. In this chapter, we describe simulation setup used in our study, our EV-DO simulator, and discuss several prominent results of our extensive simulations.

5.1 Simulation Setup

We discuss important aspects of a simulation environment in the following subsections.

5.1.1 Wraparound Model

As for cell interference, we implement a 19 cell wraparound model as depicted in Figure 9. It makes the interference environment more realistic than with a 7 cell model as it considers second level interferences, and is recommended in [54]. With the wraparound

[1] Previous EV-DO evaluation studies [6, 52] have been conducted on each link separately.

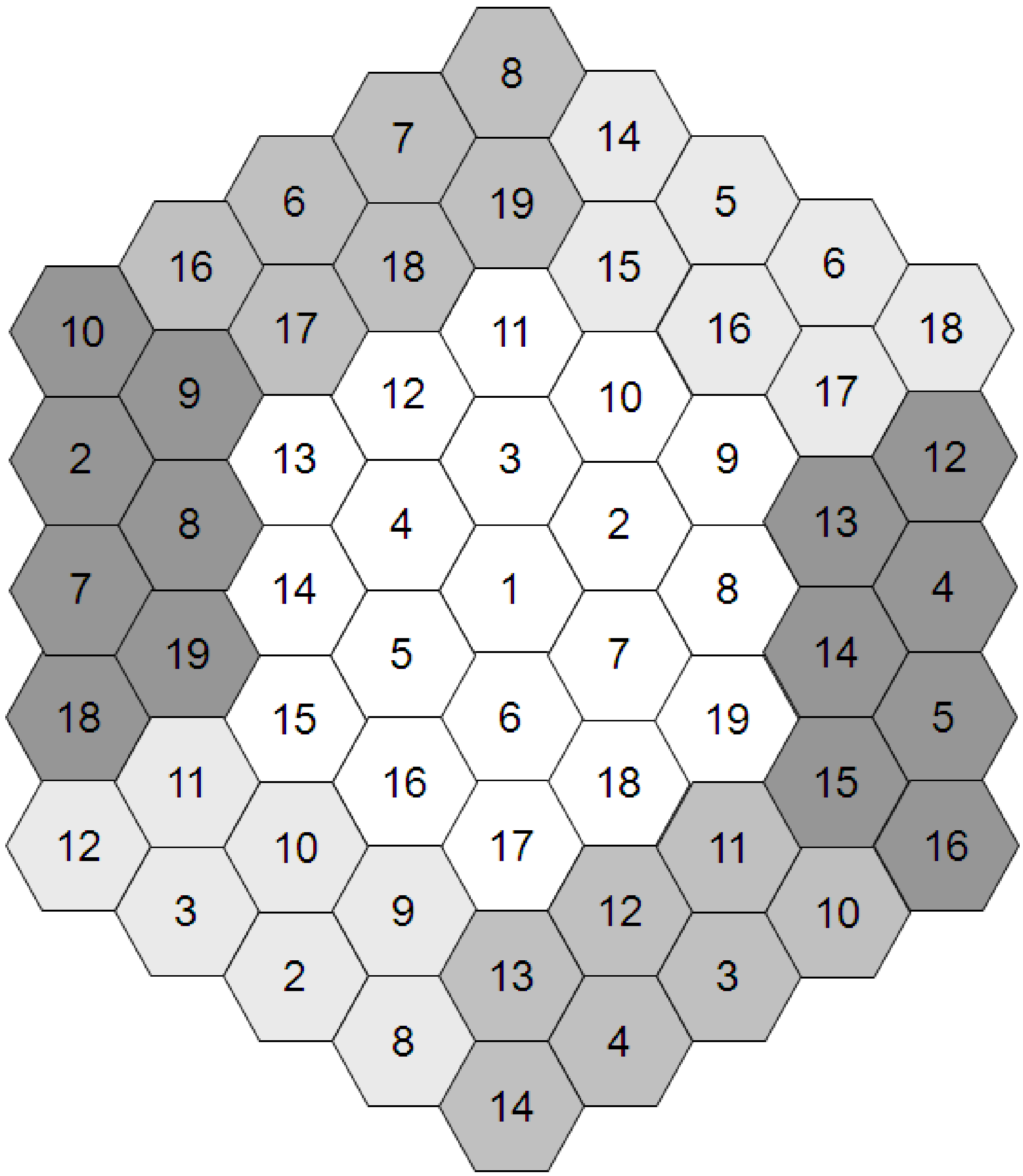

Figure 9: 19 cell wraparound model (The 19 white cells in the center are our modeled cells. The other gray cells are imaginary cells that give interferences)

model, the interference affects other cells nearby in the simulation. In Figure 9, the 19 white center cells are our modeled cells. The other gray cells are imaginary cells that show how wraparound models work. For example, when we calculate interference of white cell 11, cells 10, 3, 12, 18, 19, and 15 give first level interferences and cells 16, 9, 2, 1, 4, 13, 17, 6, 7, 8, 14, and 5 give second level interferences. Each cell has three sectors. From this model, we can collect the results from all cells rather than only from the center cell. We also evaluate the performance repeatedly and from all the 57 sectors to provide the statistical results.

5.1.2 Path Loss and Antenna Gain

The path distance and angle used to compute the path loss and antenna gain of a mobile station at (x, y) to a base station at (a, b) are computed with Equations (5.1) and (5.2), respectively from [54].

$$Path\_loss = 28.6 + 35log10(d)dB \tag{5.1}$$

where d is the distance between the base station and the mobile station in meters.

$$A(\theta) = -min(12 * (\theta/70.0)^2, 20)dB \tag{5.2}$$

where $-180 \leq \theta \leq 180$.

The distance used in the path loss between a mobile station at (x, y) to the nearest base station in a group of cells centered at (a, b) is the minimum of the following.

$$
\begin{aligned}
min \quad \{ \quad & Dist\{(x,y),(a,b)\} \\
& Dist\{(x,y),(a+3R, b+8\sqrt{3}R/2)\} \\
& Dist\{(x,y),(a-3R, b-8\sqrt{3}R/2)\} \\
& Dist\{(x,y),(a+4.5R, b-7\sqrt{3}R/2)\} \\
& Dist\{(x,y),(a-4.5R, b+7\sqrt{3}R/2)\} \\
& Dist\{(x,y),(a+7.5R, b+\sqrt{3}R/2)\} \\
& Dist\{(x,y),(a-7.5R, b-\sqrt{3}R/2)\}\}
\end{aligned}
$$

where R is the radius of a circle that connects the six vertices of the hexagon.

5.1.3 Channel Model

We used the five channel models as recommended in [54]. Channel models are randomly assigned to each mobile station. The probabilities that mobile stations take the channel models A, B, C, D, and E are 0.3, 0.3, 0.2, 0.1 and 0.1, respectively. Table 3 summarizes the channel models that were used.

5.1.4 Traffic Generation

As the effectiveness of the scheduling algorithms would depend on the traffic mix, we evaluate the algorithms under various scenarios. We vary the number of VoIP sessions from 5 to 45 users. Additionally, 10 Best Effort (BE) sessions are added to observe the interplay of VoIP and BE traffic. For VoIP traffic, EVRC is used as mentioned in Chapter 3.

Table 3: Channel models used

Channel model	Multi-path model	No. of fingers (paths)	Speed (km/h)	Fading	Model assignment probability
Model A	Pedestrian A	1	3	Jakes	0.30
Model B	Pedestrian B	3	10	Jakes	0.30
Model C	Vehicular A	2	30	Jakes	0.20
Model D	Pedestrian A	1	120	Jakes	0.10
Model E (Stationary)	Single path	1	0, f_D=1.5 Hz	Rician Factor K = 10 dB	0.10

We also use silence suppression for VoIP packets, where a 1/8 rate packet is generated every 240 ms in a silence mode. Robust Header Compression (RoHC) [18] is used as recommended in [55]. RoHC reduces IP, UDP, and RTP headers from 40 bytes to just 3 bytes, which leads to significant bandwidth savings. For BE traffic, FTP file downloads are performed for large files, so that the channels do not go idle for the duration of the simulation.

5.1.5 Reverse Link Implementation

The uplink activity includes reverse activities of applications such as reverse direction VoIP (two way conversation) and TCP acknowledgements. Our implementation supports reverse link Hybrid ARQ. Reverse link Hybrid ARQ is explained in Figure 6. Our implementation also supports two reverse link transmission modes, LoLat and HiCap modes. LoLat mode represents a low latency mode and HiCap mode represents a high capacity mode. LoLat mode generally uses less than 4 subpackets. LoLat mode is more

Table 4: Summary of parameters used for simulation

Parameter	Value
# of VoIP users/sector	5, 10, 15, 20, 25, 30, 35, 40, 45
# of BE users/sector	10
Bandwidth	1.25 MHz
Cell radius	1 Km
Maximum BS transmission power	20W (43 dBm)
Slot length	1.667 ms
VoIP packet length	5B $\sim$ 23 B after RoHC
Interval of VoIP packet generation	20 ms
Path loss exponent	3.5

susceptible to channel variation because it uses less time diversity and fewer ARQ rounds compared to the HiCap mode. LoLat mode also requires a larger traffic-to-pilot power ratio. HiCap mode generally uses 4 subpackets.

5.1.6 Other Simulation Parameters

Table 4 summarizes other simulation parameters.

5.2 EV-DO simulator

The layout of the EV-DO simulator is shown in Figure 10. The simulator is composed of the application configuration, the application profile, the simulation configuration, the radio channel configuration, the base station controller, base stations, and mobile stations. The relationship of simulator components is shown in Figure 11.

Traffics are defined in the application configuration. In our simulation, Voice over

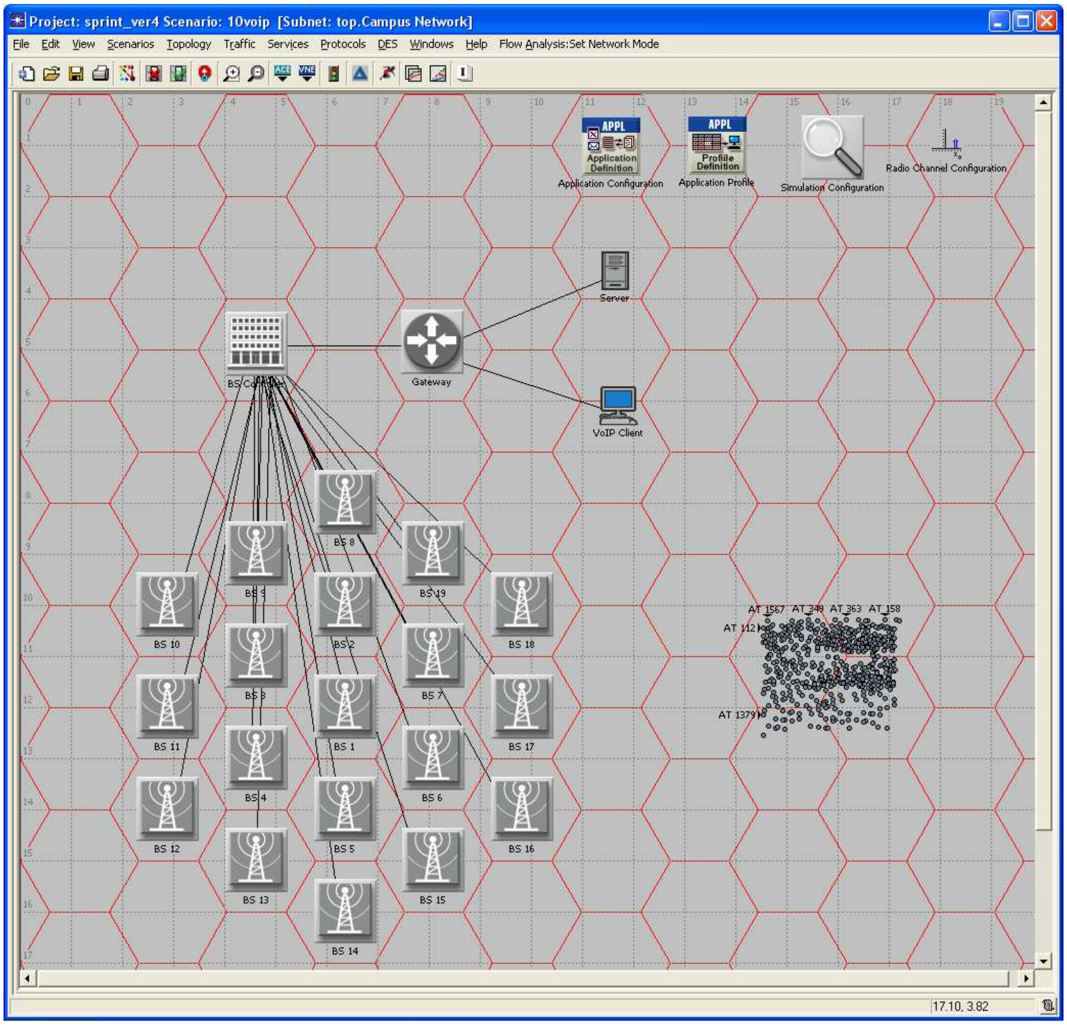

Figure 10: EV-DO simulator layout

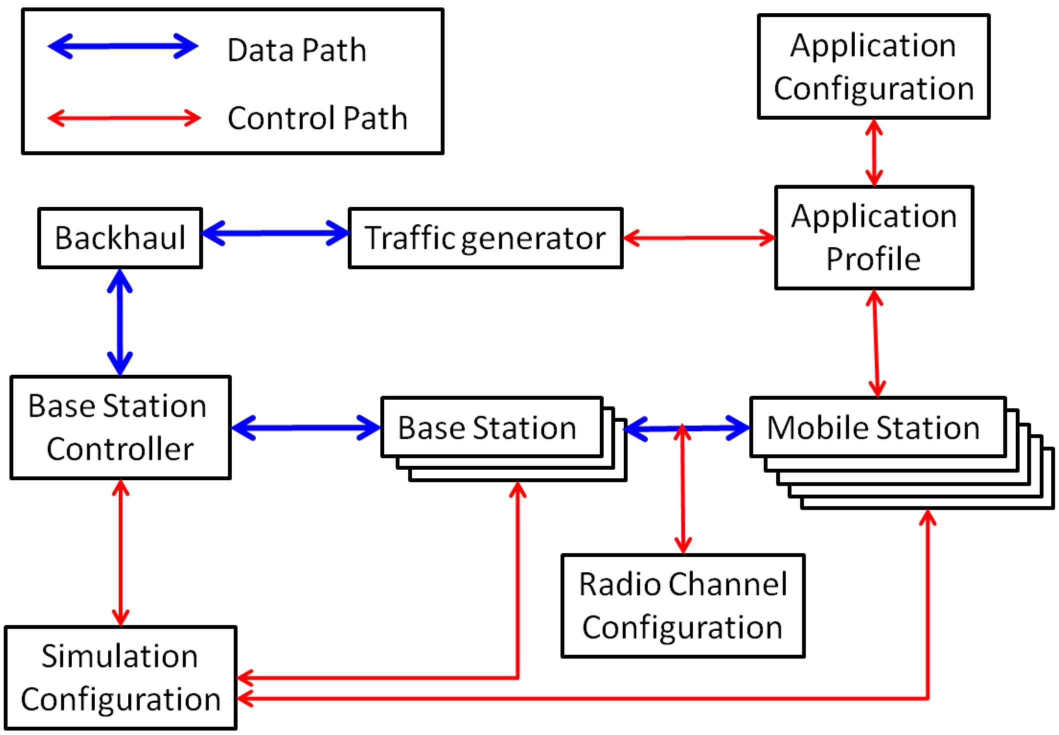

Figure 11: Relationship of simulator components

Internet Protocol (VoIP) and Best Effort (BE) traffics are used. Codec type, source rate, silence suppression, voice activity length, silence activity length, and voice activity factor are defined for VoIP traffic. Hypertext Transfer Protocol (HTTP), File Transfer Protocol (FTP), Database queries, E-mail, and Remote Login are supported as types of BE traffics. Based on the type of BE traffic, the supported parameters are different. Generally request size and inter-request time are defined for BE traffic.

Traffic patterns are defined in the application profile. Applications that are defined in the application configuration can be used to define a profile of application. After an application is selected, operation mode, start time, duration, and repeatability are defined for a profile of application.

All simulation parameters, except radio channel parameters, are defined in the simulation configuration. Scheduler type, location of mobile stations, cell radius, error rate of data rate control packets, use of fast fading, use of shadowing, use of wraparound model, and wireless channel assignment are notable parameters in the simulation configuration.

Radio channel parameters are defined in the radio channel configuration. The radio channel configuration also determines packet errors during simulation. In order to determine packet errors, the radio channel configuration loads modulation tables and fast fading tables. In each slot time, the radio channel configuration calculates path losses, antenna gains, interferences, and signal to noise ratios (SNRs). Packet errors are simulated based on SNRs and random number generations.

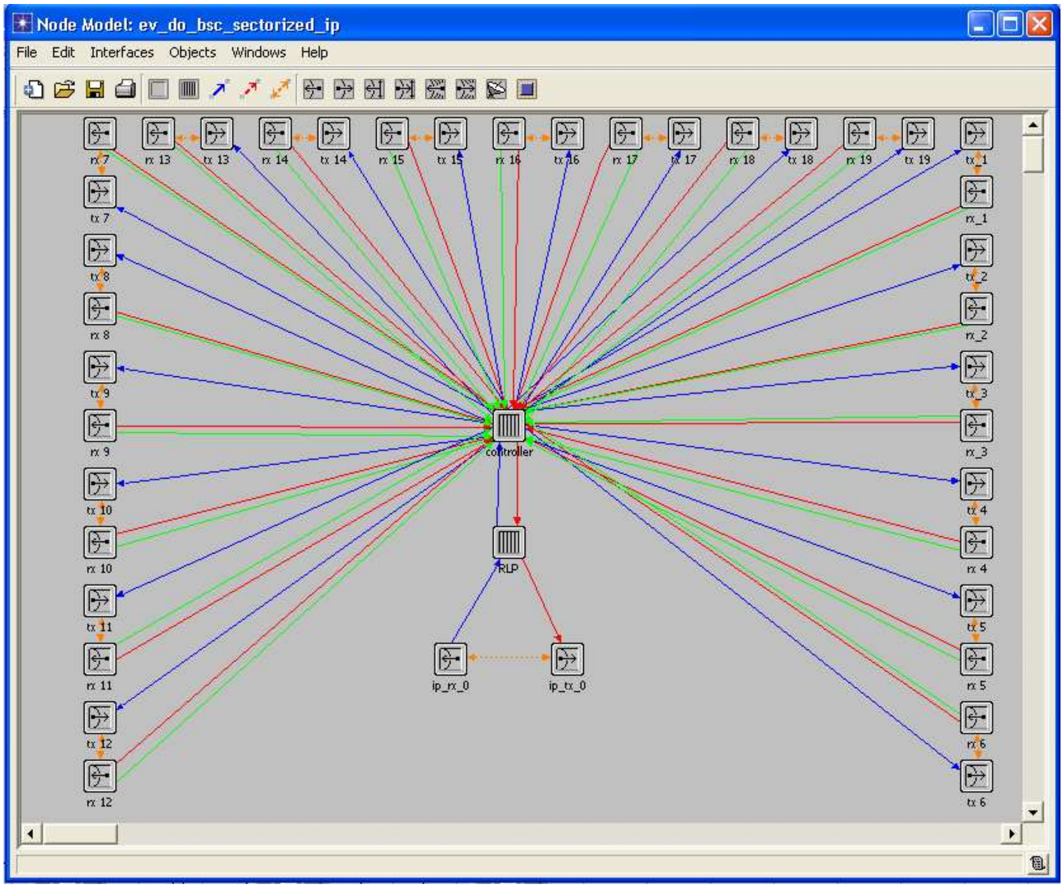

Figure 12: Base station controller components

5.2.1 The Base Station Controller

The base station controller receives packets from the backhaul and routes packets to the base stations. The base station controller also sends packets to the backhaul when the base station controller receives packets from the base stations. The base station controller needs to handle inter-cell hand-offs. When inter-cell hand-offs happen, the base station controller holds new packets from the backhaul during inter-cell hand-offs, transfers packets between the base stations, and updates the routing information.

The internal components of the base station controller are shown in Figure 12. The Radio Link Protocol (RLP) module, the BSC controller module, transmitters, and receivers are internal components of the base station controller. Transmitters and receivers are standard OPNET modules and we reuse those modules.

The RLP module inspects the IP packets and decides the traffic class of packets. The RLP module also determines destination, which is one of the mobile stations, of the IP packets based on an IP address. Then the RLP module reduces the IP header using Robust Header Compression. The RLP module holds all uplink packets and sends the IP packets to backhaul if all segmented IP packets are found in its buffers.

The BSC controller module routes packets to base stations. If a mobile station is not in active mode, the BSC controller module will page the mobile station. If a mobile station is currently in an inter-cell hand-off mode, the BSC controller will store packets and deliver them to another base station after the inter-cell hand-off is done. If a mobile station is in an inter-cell hand-off mode and an old base station forwards packets, the BSC controller module forwards packets to a new base station. The BSC controller also handles all uplink control packets.

5.2.2 Base Stations

A base station schedules downlink packets and sends downlink packets to mobile stations. A base station receives uplink packets from mobile stations and forwards the uplink packets to the base station controller. A base station keeps maintaining channel quality information of all mobile stations and updates channel quality information when

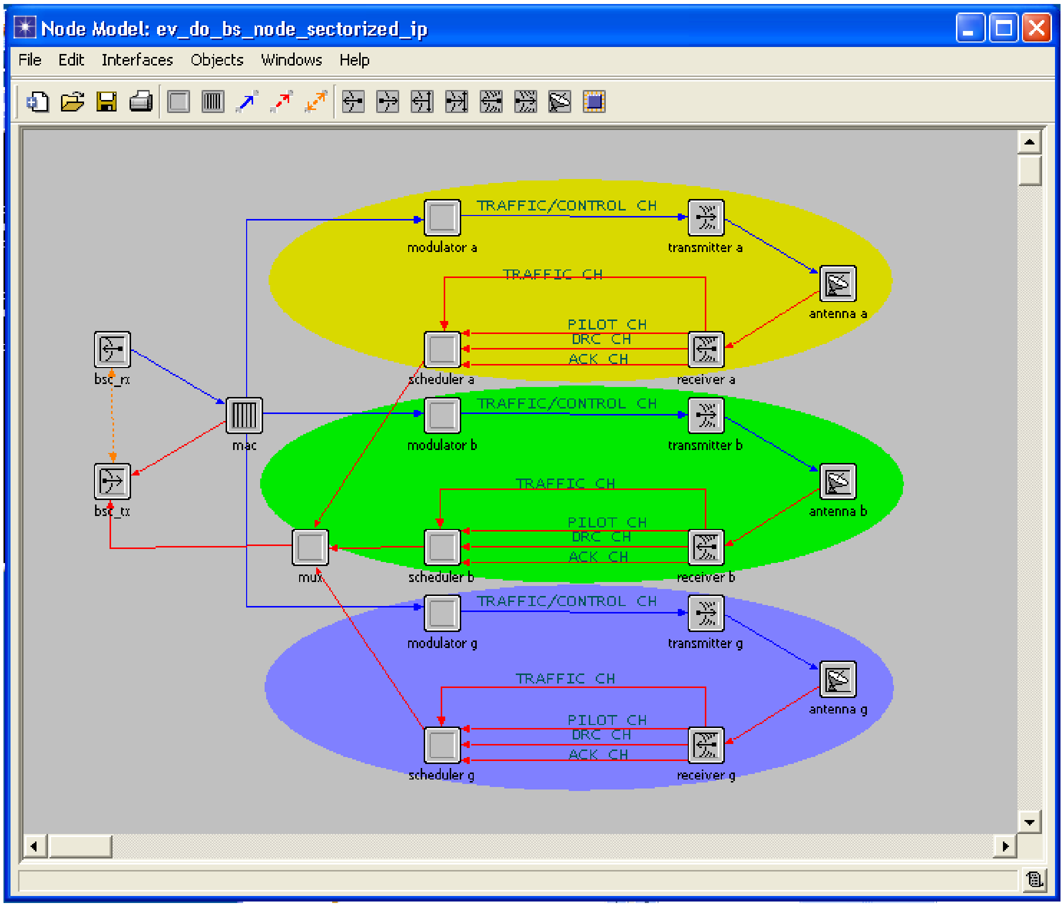

Figure 13: Base station components

the base station receives data rate control packets from the mobile stations.

The internal components of the base station are shown in Figure 13. Transmitters, receivers, antennas, Media Access Control (MAC) module, MUX module, scheduler modules, and modulator modules are internal components of the base station. The base station has three scheduler modules and three modulator modules because the base station divides the cell area into three sectors (alpha, beta, and gamma). Transmitters, receivers, and antennas are standard OPNET modules and we reuse those modules.

The MAC module receives downlink packets and holds the packets to different queues based on sectors and traffic classes. If a class of a packet is real-time class and a packet is old enough to drop, the MAC module will drop the packet. When intra-cell hand-off happens, the MAC module moves packets that are affected by intra-cell hand-off between queues. When the inter-cell hand-off happens, the MAC module sends all packets that are affected by inter-cell hand-off to the base station controller. The MAC module also gives the state information of packets and queues to the scheduler modules. The MAC module creates single user packets or multi-user packets based on the scheduler module's decision and sends the packet to the modulator module.

If a current slot is not scheduled yet, the scheduler module requests the state information of queues and packets to the MAC module. Then the scheduler module decides to send a single user packet or a multi-user packet and which VoIP or BE packets are included in a single user packet or a multi-user packet. After the decision is done, the scheduler module publishes the decision to the MAC module. In each slot boundary, the scheduler module processes the Data Rate Control packets and updates the channel conditions. When the scheduler module receives an acknowledgement from the mobile station, the scheduler module cancels the retransmissions if they are already scheduled for the packet that the acknowledgement is related to. When uplink packets are received in the scheduler module, those packets are sent to the MUX module.

The modulator module holds packets from the MAC modules and sends the packets to the mobile stations. The modulator module also schedules retransmissions up to four times based on the data rates of a packet.

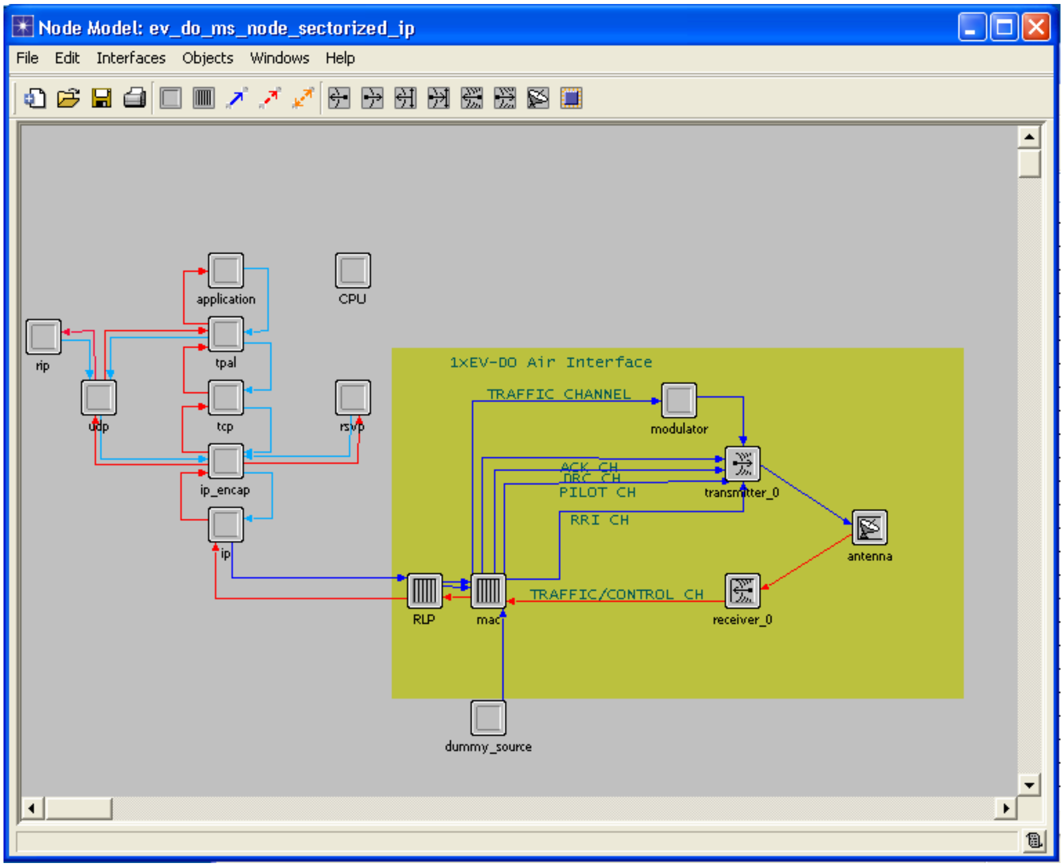

Figure 14: Mobile station components

5.2.3 Mobile Stations

A mobile station receives downlink packets and sends uplink packets to a base station. Uplink packets are composed of data packets, acknowledgement packets, and Data Rate Control packets. A mobile station has its own scheduler to decide which packet will be sent.

The internal components of the mobile station are shown in Figure 14. A MAC module, modulator module, and RLP module are internal components of the mobile station. All other modules in the mobile station are standard OPNET modules. Many modules are reused in the mobile station compared to the base station controller or the base station because the mobile station should simulate packets up to the application layer.

The MAC module in the mobile station can be divided by two main functionalities. In active states, the MAC module maintains Pilot to Interference Noise Ratio (PINR). If inter-cell or intra-cell hand-off is needed, the MAC module initiates a hand-off process. If no hand-off is needed, the MAC module sends data rate control packets to the base station. The MAC module queues all packets from the IP layer. If a current slot is not scheduled yet, the MAC module runs a scheduling algorithm to create a packet. Then the MAC module sends the packet to the modulator module. In sleep states, the MAC module checks only paging control packets. If a paging control packet is received, the MAC module will change its state from sleep to active after the interaction between the base station and the mobile station.

The modulator module in the mobile station is very similar to the modulator module in the base station. Both of them send packets and schedule retransmissions. The only difference between them is how many flows the modulator module can support. The modulator module in the mobile station only processes up to two flows, LoLat and HiCap flows, but the modulator module in the base station processes all flows destined to the mobile stations in three sectors.

The RLP module in the mobile station collects MAC layer packets and sends IP packets to the upper layer. If partial IP packets stay in the buffers too long, the RLP

Table 5: Definition of states in EVRC traffic generator

$s(n)$	$R(n-1)$	$R(n)$
0	1	1
1	1	1/2
2	1	1/4
3	1	1/8
4	1/2	1
5	1/2	1/2
6	1/2	1/4
7	1/2	1/8
8	1/4	1
9	1/4	1/2
10	1/4	1/4
11	1/4	1/8
12	1/8	1
13	1/8	1/2
14	1/8	1/4
15	1/8	1/8

module removes those partial packets from the buffers. The RLP module in the mobile station decides the uplink flows. The LoLat flow is used for real-time traffic and the HiCap flow is used for BE traffic.

5.2.4 Implementation of EVRC

Most of the VoIP traffic generators supported by the OPNET simulator are on-off traffic generators and support only two states. Based on the IS-871 standard [59], the

Table 6: Scaled cumulative transition probabilities

$s(n-1)$	A1	B1	C1
0	0	0	2916
1	0	20906	25264
2	0	0	0
3	0	0	0
4	0	0	4915
5	0	17170	24969
6	21856	25887	27099
7	0	0	0
8	0	0	4522
9	0	5472	16384
10	21856	21856	24576
11	28246	29622	30802
12	0	0	5472
13	0	6554	6554
14	28377	28934	29491
15	29753	32473	32571

EVRC traffic generator needs to produce four rates (full rate, half rate, quarter rate, and 1/8 rate) and support at least four states. We modified the standard OPNET module to implement the EVRC traffic generator. Implementation of the EVRC traffic generator is described in the IS-871 standard [59]. We summarize important implementation details.

First, we define states based on previous rates and current rates. Table 5 shows the definition of states in the EVRC traffic generator.

Algorithm 4 Next transition state in the EVRC traffic generator

if $z_n <$ A1

$s(n) = (4 * s(n-1) + 3)$ mod 16

else if A1 $\leq z_n <$ B1

$s(n) = (4 * s(n-1) + 2)$ mod 16

else if B1 $\leq z_n <$ C1

$s(n) = (4 * s(n-1) + 1)$ mod 16

else

$s(n) = (4 * s(n-1))$ mod 16

end if

When a VoIP call is started, the initial state of $s(n)$ is 0. Random number z_n, whose range is between 0 and 32767, is generated every 20 ms until the end of the VoIP call. Then Table 6 and Algorithm 4 are used for determining the next state. After $s(n)$ is determined, the corresponding rate, $R(n)$, is given by following.

$$R(n) = \frac{1}{2^{(s(n) mod 4)}} \tag{5.3}$$

5.3 Simulation Results

In this section, we first compare bundling algorithms using delay, throughput, and average packet loss rate. We then discuss performance of bundling algorithms with various channel conditions and different cell models. Comparing multi-user packet bundling to single user packet bundling is an next topic. We also compare different maximum bundling delays in the QCB Algorithm. Finally, we compare our algorithms to an existing algorithm.

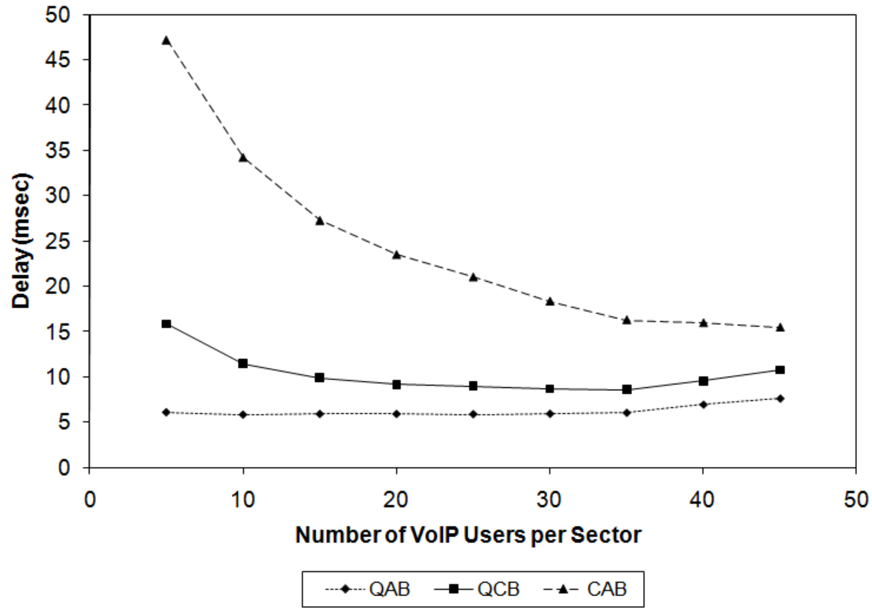

Figure 15: Comparisons of bundling algorithms for VoIP traffic delay

5.3.1 Comparisons of Bundling Algorithms

We first discuss the delay and throughput performances of the bundling algorithms, QAB, CAB, and QCB. Figure 15 compares average delays of VoIP traffic for QAB, CAB, and QCB algorithms. The BE traffic throughput of the three algorithms is shown in Figure 16.

QAB performs the best for VoIP delay as it schedules based on the remaining time to meet the QoS. The BE throughput decreases as the number of VoIP users increases in all cases, because VoIP traffic receives priority over BE traffic. Meanwhile, CAB exhibits the most throughput for BE, maximizing channel utilization.

QCB provides the best of both of the other methods. Notice that despite the extra delay due to the deferred bundling time in QCB (maximum 25 ms), QCB VoIP delay is a

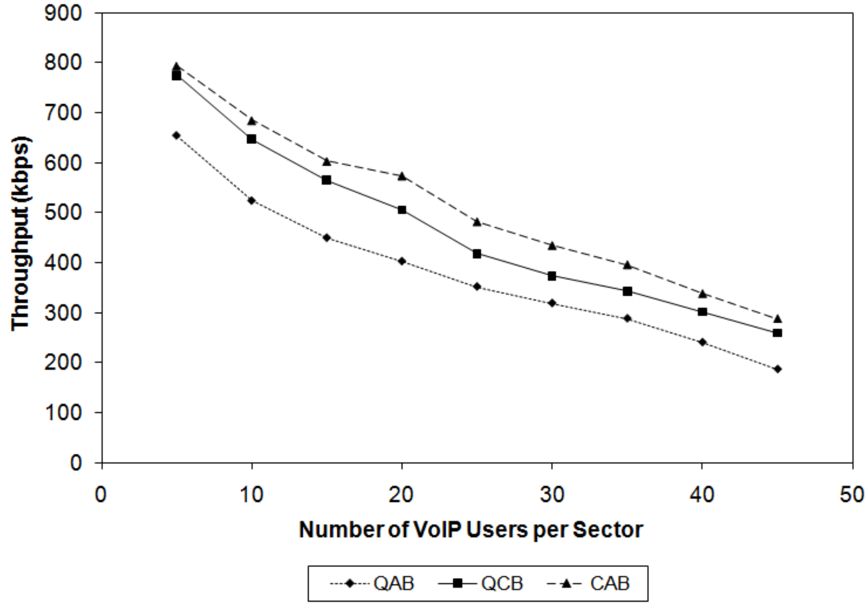

Figure 16: Comparisons of bundling algorithms for BE throughput

lot closer to QAB than CAB. Meanwhile, in terms of BE throughput, our scheme shows high performance close to CAB due to bundling efficiency. Figures 15 and 16 show a good performance tradeoff between the delay and throughput of the QCB algorithm. In fact, if we can allow even more VoIP delay depending on the remaining time to deadline, we can get more BE throughput via exploiting better channel diversity. However, the trade-off between delay and throughput is achieved optimally with around a 25 ms bundling delay, for the given parameters of the traffic load.

Figure 17 compares the average loss rate of VoIP packets for the bundling algorithms. The loss can be due to either channel condition or drops at the queue. In general, the packet loss rate stays very small and insignificant, around the value of 0.1∼0.5%, for all the cases. We find that the impact of the small number of occasional packet losses on

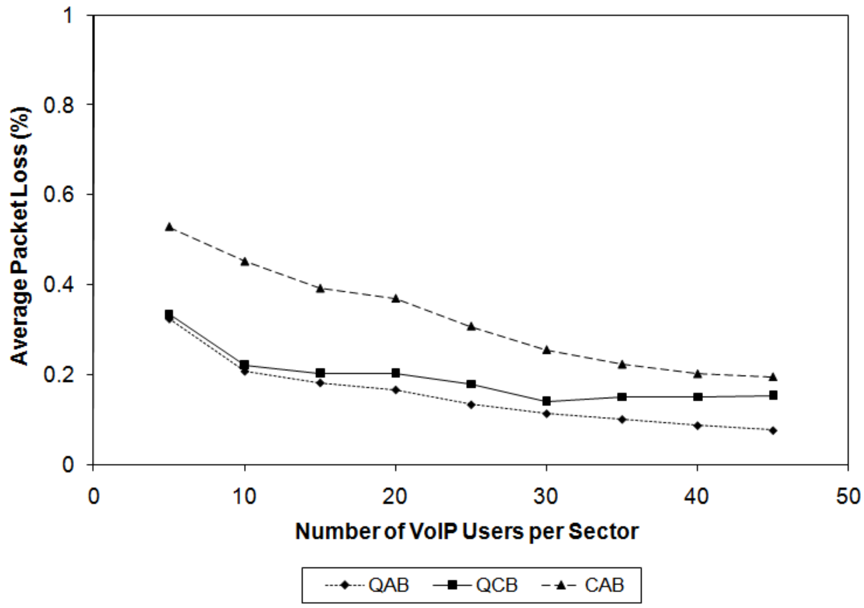

Figure 17: Comparisons of bundling algorithms for packet loss rate

the average loss rate, decreases as the amount of VoIP traffic increases. Also, the more bundling opportunities that are given, the less loss rate is achieved.

5.3.2 Comparisons of Bundling Algorithms with Various Channel Conditions

We consider various channel conditions, and compare the performance of QAB, QCB, and CAB in Figures 18 and 19. As for the normal channel condition, we used the mixture of channel model A∼E as specified in [54] (i.e., Channel model A 30%, model B 30%, model C 20%, model D 10%, and model E 10%). For the good channel condition, we used 100% channel model E, and for the bad condition, we used 10% model A, and 30% for models B, C, and D. In QAB and QCB cases, less average VoIP delays are measured when the channel condition is better. In the CAB case, an average VoIP delay with a bad channel condition is less than an average VoIP delay with a normal

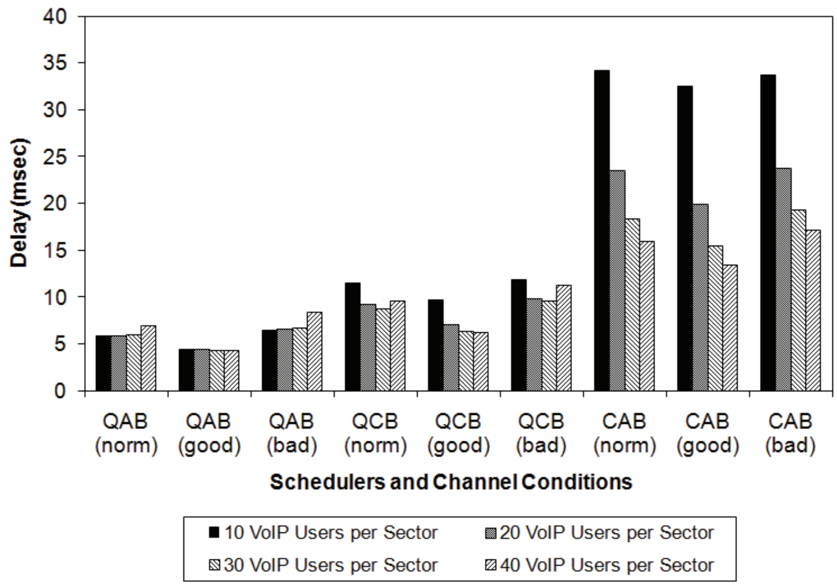

Figure 18: Comparisons of bundling algorithms for average VoIP traffic delay under various channel conditions (normal, good, bad)

channel condition because a bad channel condition case gives more bundling chances than a normal channel condition case. The main factor of an average VoIP delay is deferred bundling time in the CAB case. In general, the performances are better with a good channel condition and worse with a bad channel condition, for all schedulers. We find that regardless of the channel condition, QCB achieves an excellent tradeoff between QAB and CAB, in that a little increase in VoIP delay brings near-CAB BE throughput.

5.3.3 Comparisons of Single Cell Results to 19 Cell Wraparound Results

In this subsection, we compare single cell results to 19 cell wraparound results. Figures 20 shows BE throughput and Figure 21 shows VoIP delay. Single cell results are much better than 19 cell wraparound results because there is no inter-cell interference in

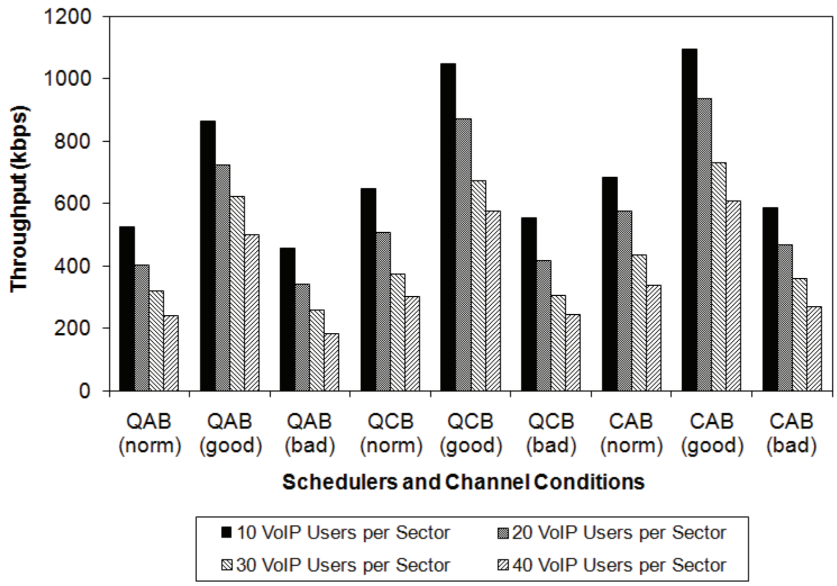

Figure 19: Comparisons of bundling algorithms for BE throughput under various channel conditions (normal, good, bad)

the single cell model. The results will be too optimistic when we use a single cell model. We find that QCB achieves an excellent tradeoff between QAB and CAB in different interference models.

5.3.4 Comparisons of Single User Packet Bundling to Multi-user Packet Bundling

Next, we investigate the variants of QCB and observe the value of a multi-user packet bundling (we name it QCB-MUP or simply MUP) over a single user packet bundling (SUP Multiplex) or no bundling (SUP Simplex). Figure 23 shows interesting behavior for the delay cumulative distribution functions (CDFs) of VoIP packets when using QCB. We compare the single-user packet (SUP) multiplex (i.e., bundled packets from the same user) and MUP schemes. In the SUP multiplex, VoIP delay increases when the number

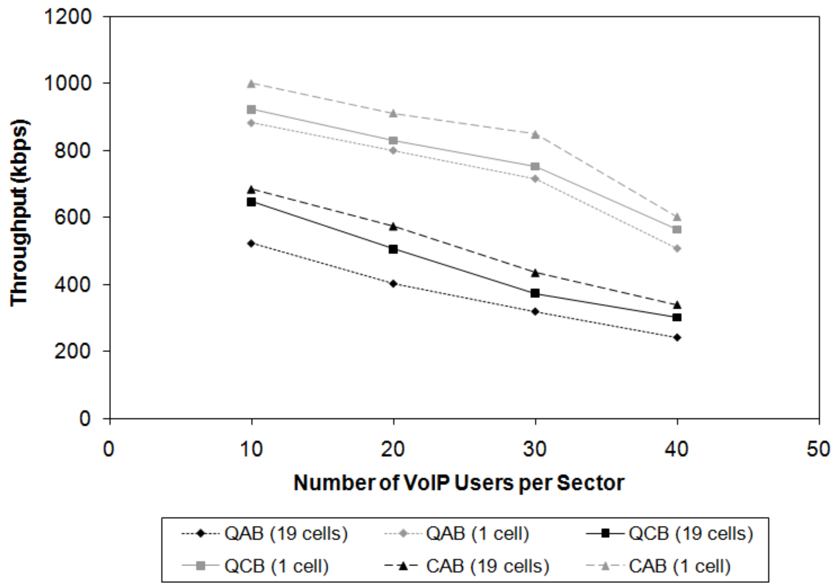

Figure 20: BE throughput of bundling algorithms in single cell and 19 cell wraparound models

of users increase. Meanwhile, with MUP, VoIP delay decreases as the number of users increase. This is because in SUP multiplex, each user takes turns in the use of time slots and the period becomes longer with the increased number of users. On the other hand, in MUP, the more VoIP packets from the increased number of users makes the bundling easier with little need to wait, thus enhancing the multi-user diversity gain. In both QCB-MUP and QCB-SUP multiplex cases, the CDFs show a longer tail for a greater number of VoIP users. VoIP delay generally decreases when the number of VoIP users increase because the chance of bundling is higher. However, some VoIP packets have a higher delay when the number of VoIP users increase because the chance of congestion (many VoIP packets existing in the queue) is higher.

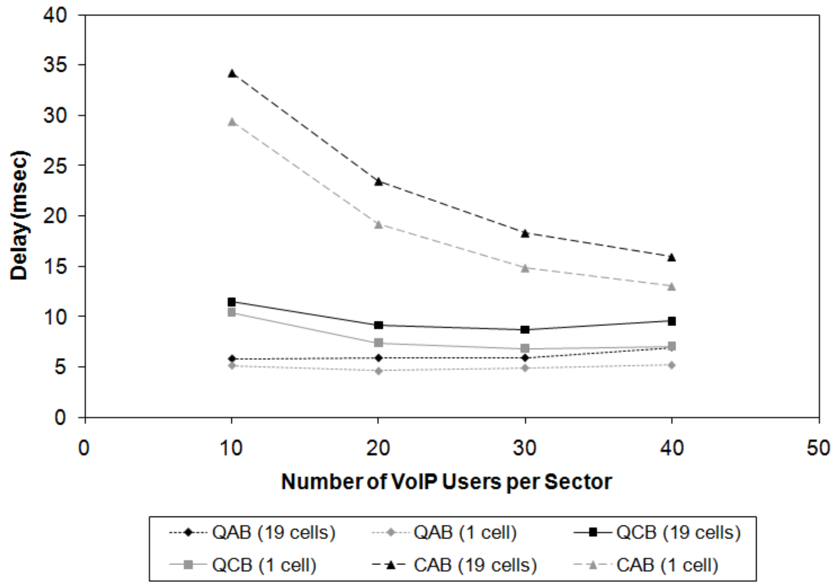

Figure 21: VoIP traffic delay of bundling algorithms in single cell and 19 cell wraparound models

Figure 22 compares the QCB BE throughput of the SUP simplex, SUP multiplex, and MUP. First, the BE throughput decreases as the number of VoIP users increase, since the higher priority is given to VoIP over the BE traffic. The decrease of BE throughput is more prominent in the SUP simplex than in the bundling schemes. The throughput of packet bundling using either the SUP multiplex or the MUP degrades gradually as they attempt to maximize channel utilization with higher rates of bundling. Particularly, the SUP multiplex shows higher BE throughput with a small number of VoIP users, and the MUP wins over the SUP multiplex with a large number of VoIP users. It shows that the efficiency of the MUP increases when the number of users grows, as it takes advantage of multi-user diversity better. As the MUP format is decided by the worst DRC values of

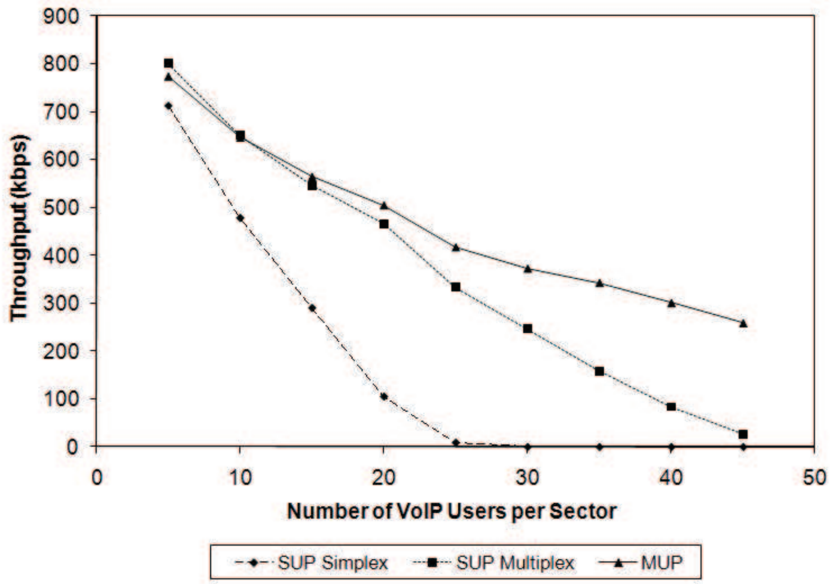

Figure 22: Throughput of BE for SUP-simplex (no-bundling), SUP multiplex and MUP (variants of QCB)

mobile stations whose packets are bundled, it is more likely to find similar DRC users as the number of VoIP users increases.

Table 7 gives the average packet drop rates while changing the number of VoIP users. It clearly shows that the SUP simplex cannot handle much VoIP traffic from around 25 users, since it has very high packet loss rates over 10%. This table shows that bundling is required if we want to handle VoIP traffic. The SUP multiplex and the MUP both have low drop rates.

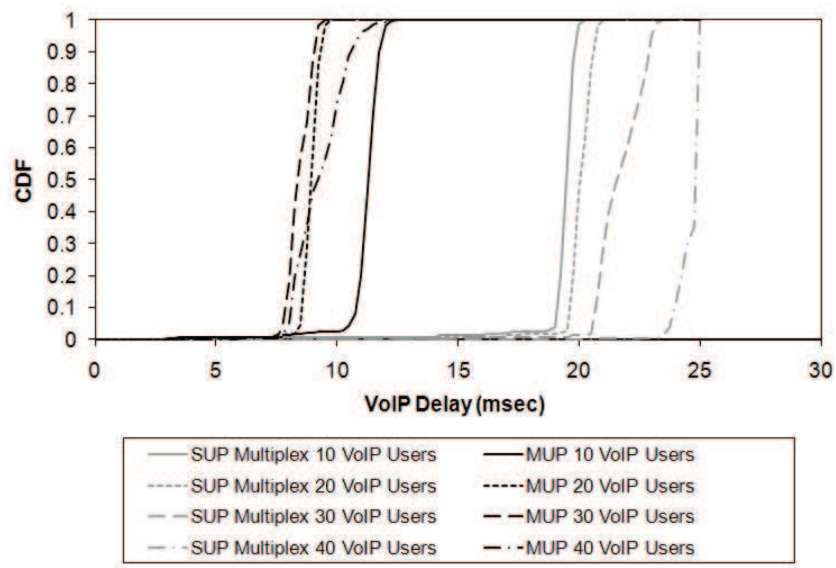

Figure 23: Empirical cumulative density functions of VoIP packet delays for SUP multiplex and MUP (variants of QCB)

5.3.5 Comparisons of Different Maximum Bundling Delays in the QCB Algorithm

In this subsection, we compare different maximum bundling delays in the QCB algorithm. Figures 24 shows BE throughput and Figure 25 shows VoIP delay. The average BE throughput of the maximum bundling delay with 15 ms is less than the average BE throughput of other maximum bundling delays. The average VoIP delays of the maximum bundling delay with 15 ms and 25 ms is better than the average VoIP delays of the maximum bundling delay with 45 ms and 65 ms. In general, the maximum bundling delay in the QCB algorithm is not the significant factor but the maximum bundling delay with 25 ms is the best tradeoff value in those results.

Table 7: Average packet loss (%)(packet error + drop)

No. VoIP users/sector	SUP Simplex	SUP Multiplex	MUP
5	0.23	0.42	0.33
10	0.19	0.31	0.22
15	0.17	0.29	0.2
20	0.74	0.26	0.2
25	10.9	0.26	0.18
30	24.62	0.19	0.14
35	35.46	0.18	0.15
40	43.62	0.22	0.15
45	49.65	0.25	0.15

5.3.6 Comparisons of Our Algorithms to an Existing Algorithm

We first compare the delay and throughput performances of our bundling algorithms to an existing scheme, PF-MUP that selects a user's packet whose priority is the highest (longest delay in our case) according to the PF algorithm, then adds other packets only with the same channel quality. Figure 26 compares average delays of VoIP traffic for PF-MUP, QAB, CAB, and QCB schemes. The BE traffic throughput of the four schemes is shown in Figure 27.

When the number of VoIP users is small, VoIP delay of PF-MUP is closer to the VoIP delay of QAB. When the number of VoIP users increases, the VoIP delay of PF-MUP is closer to VoIP delay of QCB. PF-MUP is a good scheme for VoIP delay. However, the BE throughput of PF-MUP is the worst among the four schemes. Because the PF-MUP scheme only adds packets with the same channel quality, the average number of bundled

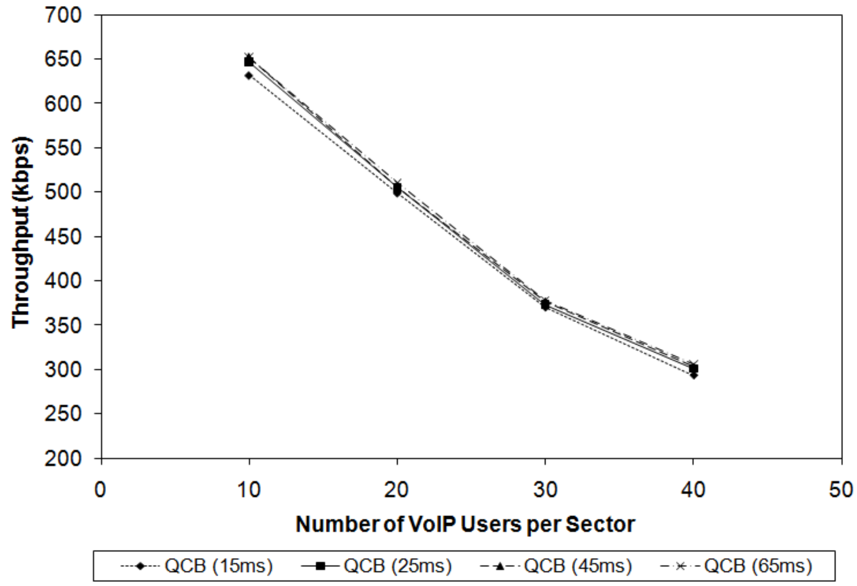

Figure 24: Comparisons of different maximum bundling delays in the QCB algorithm for BE throughput

VoIP packets of the PF-MUP scheme is much less than the average number of bundled VoIP packets of other schemes. Then PF-MUP should use more multi-user packets to send VoIP packets. The more the multi-user packet is used, the less BE throughput is achieved.

Figure 28 compares the average VoIP packet loss rate of the PF-MUP, QAB, CAB, and QCB schemes. The average VoIP packet loss rate of the PF-MUP scheme is very close to the average VoIP packet loss rate of the QCB scheme.

In general, the QCB scheme performs better than the existing scheme, PF-MUP.

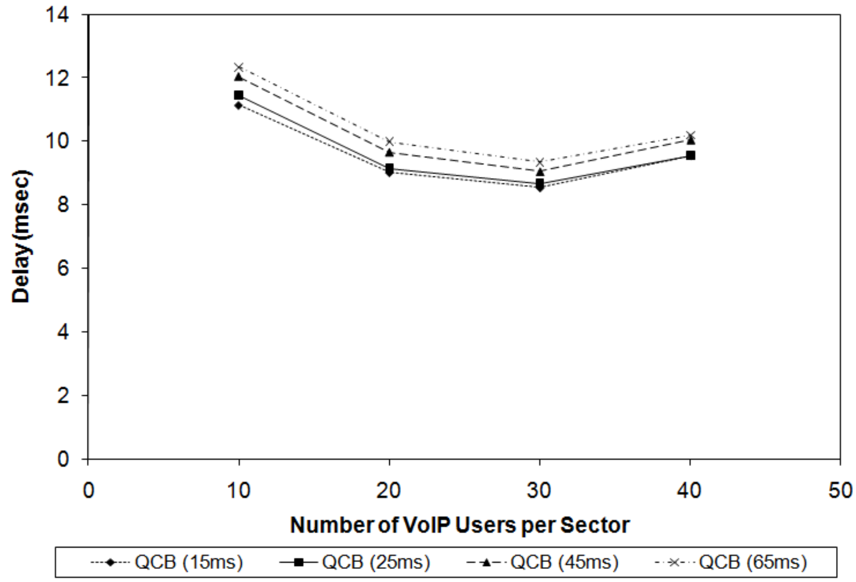

Figure 25: Comparisons of different maximum bundling delays in the QCB algorithm for VoIP traffic delay

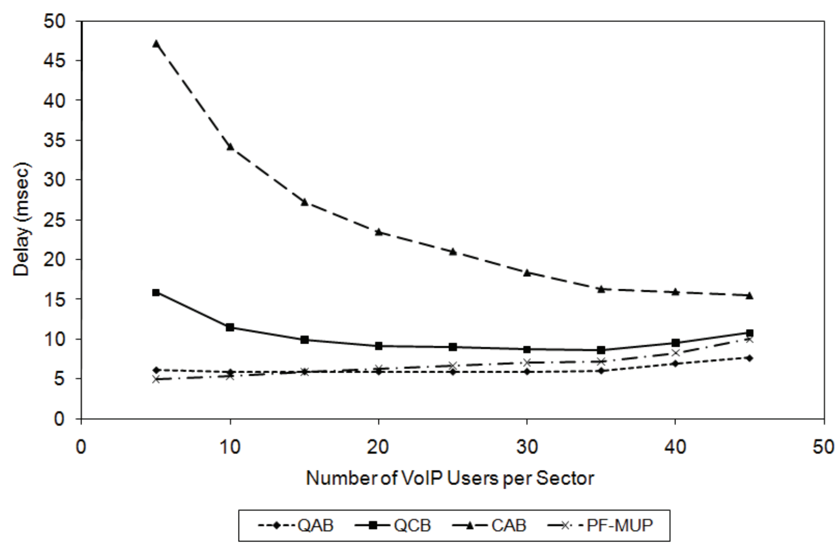

Figure 26: Comparisons of four schemes for VoIP traffic delay

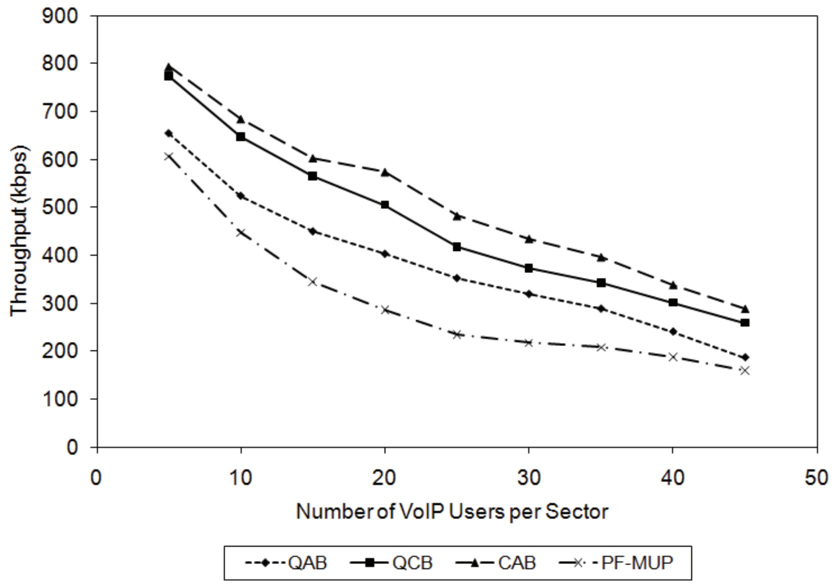

Figure 27: Comparisons of four schemes for BE throughput

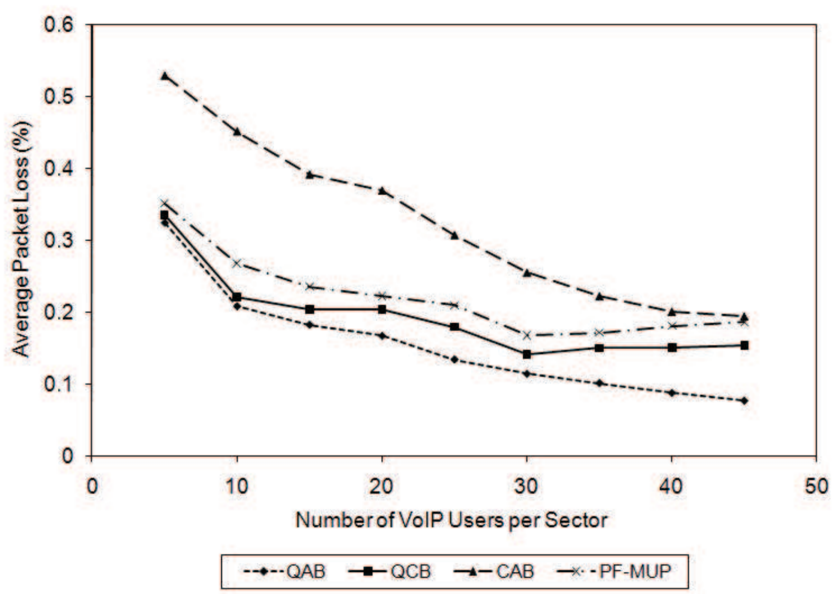

Figure 28: Comparisons of four schemes for packet loss rate

5.3.7 Summary

For various operating conditions, CAB has the best BE throughput and the worst VoIP delay and loss, while QAB has the best VoIP delay and loss, and the worst BE throughput. The existing PF-MUP has the performance that is close to but poorer than QAB. The proposed QCB achieves the best of the QAB and CAB, in that with a slight increase in VoIP delay and loss than with QAB, it provides BE throughput close to CAB. As for variants of QCB, a multiple-user packet bundling is more effective than single user multiple packet bundling, especially with the greater number of users.

Finally, let us consider the overhead of extra packet headers incurred by our proposed packet bundling, QCB. When single user packet bundling is used (See Figure 7, (b)) for n packets, the excess header size is $8 \times n$ bits. With multi-user packet bundling, it is $16 \times n$ bits. With our simulation using 30 VoIP users and 10 BE users per sector and 25 ms delay max allowance, the average number of bundled packets in an SUP packet was 1.9, and the average size of the bundled VoIP packets was 486 bits. Meanwhile, the average number of bundled packets in an MUP packet was 4.3, and the average size of the bundled VoIP packets was 1429 bits. Therefore, the overheads of SUP and MUP are $1.9 \times 8/486 = 3.1\%$ and $4.3 \times 16/1429 = 4.8\%$ respectively, which is a negligible increase compared to the huge utility gain.

CHAPTER 6

CONCLUSIONS

We have proposed a joint QoS and Channel Aware Packet Bundling (QCB) algorithm for VoIP packets to improve spectral efficiency in cellular networks. The packet size of real-time data such as VoIP is often very small, leaving channels underutilized in TDM cellular systems. Packet bundling could improve the channel utilization in such networks. However, a careful treatment should be paid due to location dependent and time varying channel characteristics of wireless networks. Since the packet bundling algorithm is an NP-complete problem, we introduce approximation algorithms, namely QoS Aware Packet Bundling (QAB), Channel Aware Packet Bundling (CAB), and QCB. We have validated the efficacy of the approximation algorithms through extensive simulations of a complete EV-DO implementation, the first of its kind to the best of our knowledge. We have shown that the QCB scheme out-performs QAB and CAB as well as an existing bundling algorithm, thus truly maximizing a multi-user/traffic diversity gain, as it achieves a high throughput for BE traffic while keeping a low delay. We have further investigated the behavior of QCB variants, and found that the QCB-Multi-User-Packet (QCB-MUP) is more effective when there are larger numbers of VoIP users and the QCB-Single-User-Packet-multiplex (QCB-SUP-multiplex) demonstrates more BE-throughput and a lower overhead with small numbers of VoIP users.

As for future work, we plan to investigate the performance of QCB when multiple flows per node are allowed. With multiple flows per node, we expect the BE throughput of the QCB-SUP multiplex will be improved more than the current results show. With our current work, when VoIP packets are sent in the SUP multiplex case, only VoIP packets are sent because the node doesn't have any BE traffic. When multiple flows are permitted, VoIP and BE traffic may be sent together leading to a better channel utilization in the QCB-SUP multiplex. Multiple flows, however, are not expected to make a difference in the performance of the QCB-MUP scheme.

The possible future extensions of the current work can be found in multi-carrier wireless environments [3, 5]. The EV-DO revision B and WiMAX systems are good candidates for future research areas.

In EV-DO revision B, we need to revise our QAB scheme because multiple carriers give more choices. In single carrier cases, we should change the MUP format to send more VoIP packets if the DRC value of the second VoIP packet is smaller than the DRC value of the first VoIP packet. In multi-carrier cases, we can assign the second VoIP packet to a different carrier if the DRC value of the second VoIP packet is smaller than the DRC value of the first VoIP packet.

In the WiMAX system, small real-time packets like VoIP make the channels underutilized because of scheduling constraints. The main constraint is the DL-MAP and UL-MAP. Those maps are coded using the lowest coding method because all mobile stations need to decode the map correctly. The size of the map is directly proportional to the number of scheduled mobile stations. The bandwidth spent by the map is comparable with the bandwidth spent by the downlink traffic if too many mobile stations are scheduled in one frame. The other constraint in downlink is burst allocation. The burst allocation should be a rectangular shape in a two dimensional space (symbol time and subcarrier). The frequency reuse is another big factor to design a scheduling algorithm. The partial usage of subchannels (PUSC) can reduce inter-cell or intra-cell interferences. Our bundling algorithm can be applied to reduce the size of the map.

REFERENCE LIST

[1] Alouini,M.-S., and Goldsmith, A. J. Adaptive modulation over Nakagami fading channels. *Kluwer Journal of Wireless Communication 13*, 1 (2000), 119-143.

[2] Andrews, M., Kumaran, K., Ramanan, K., Stolyar, S., Vijayakumar, R., and Whiting P. CDMA data QoS scheduling on the forward link with variable channel conditions. *Bell Labs Technical Memorandum*, 4 (2000), 1-45.

[3] Andrews, M. and Zhang, L. Scheduling Algorithms for MultiCarrier Wireless Data Systems. In *Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking (MobiCom '07)* (Montréal, Québec, Canada, Sep. 9 - 14). ACM, New York, 2007, pp. 3-14.

[4] Attar, R., Ghosh, D., Lott, C., Fan, M., Black, P., Rezaiifar, R., and Agashe, P. Evolution of cdma2000 Cellular Networks: Multicarrier EV-DO. *IEEE Communications Magazine 44*, 3 (2006), 4653.

[5] Beard, C., Choi, B.-Y., and Park, H. QoS and Channel Aware Packet Bundling for VoIP and Data Traffic in Multi-Carrier Cellular Networks. In *Proceedings of the 22nd International Teletraffic Congress (ITC 22)* (Amsterdam, Netherlands, Sep. 7 - 9). IEEE, New York, 2010, pp. 1-8.

[6] Bhushan, N., Lott, C., Black, P., Attar, R., Jou, Y.-C., Fan, M., Ghosh, D., and Au, J. CDMA2000 1xEV-DO Revision A: A Physical Layer and MAC Layer Overview. *IEEE Communications Magazine 44*, 2 (2006), 7587.

[7] Bi, Q., Chen, P.-C., Yang, Y., and Zhang, Q. An Analysis of VoIP Service Using 1xEV-DO Revision A System. *IEEE Journal On Selected Areas in Commmunications 24*, 1 (2006), 3645.

[8] Borin, J. F. and da Fonseca, N. L. S. Scheduler for IEEE 802.16 Networks. *IEEE Communications Letters 12*, 4 (2008), 274276.

[9] Cao, Y. and Li, V. O. K. Scheduling Algorithms in Broadband Wireless Networks. *Proceedings of the IEEE 89*, 1 (2001), 7687.

[10] Chang, P.-R. and Lin, C.-F. Wireless ATM-based Multicode CDMA Transport Architecture for MPEG-2 Video Transmission. *Proceedings of the IEEE 87*, 10 (2001), 1807-1824.

[11] Chen, P.-C., Da, R., Mooney, C., Yang, Y., Zhang, Q., Zhu, L. H., and Zou, J. Quality of Service Support in 1x EV-DO Revision A Systems. *Bell Labs Technical Journal 11*, 4 (2007), 169184.

[12] Choi, B.-Y., Kim, J. H., and Beard, C. QoS and Channel-Aware Packet Bundling for Capacity Improvement in Cellular Networks. *IEEE Transactions on Vehicular Technology 59*, 8 (2010), 3886-3901.

[13] Choi, Y.-J. and Bahk, S. Channel-aware VoIP Packet Scheduling in cdma2000 1xEV-DO Networks. *Elsevier Journal of Computer Communications 30* (2007), 2284-2290.

[14] Chuah, M. C., Doshi, B., Dravida, S., Ejzak, R., and Nanda, S. Link Layer Retransmission Schemes for Circuit-mode Data over the CDMA Physical Channel. *Mobile Networks and Applications 2*, 2 (1997), 195-211.

[15] Cicconetti, C., Lenzini, L., and Mingozzi, E. Quality of Service Support in IEEE 802.16 Networks. *IEEE Network 20*, 2 (2006), 50-55.

[16] Dahlman, E., and Parkvall, S., and Sköld, J., and Beming, P. *3G Evolution: HSPA and LTE for Mobile Broadband*. Academic Press, Oxford, U.K., 2007.

[17] de Bruin, I., Heijenk, G., El Zarki, M., and Lei Zan, J. L. Fair Channel-dependent Scheduling in CDMA Systems. In *Proceedings of IST Mobile and Wireless Communications Summit 2003* (Aveiro, Portugal, Jun. 15 - 18). IST, Brussels, Belgium, 2003, pp. 737-741.

[18] Degermark, M., Nordgren, B., and Pink, S. IP Header Compression (IPHC). *RFC2507*. IEFT, 1999. from IEFT: http://tools.ietf.org/html/rfc2507.

[19] Fan, M., Ghosh, D., Bhushan, N., Attar, R., Lott, C., and Au, J. On the Reverse Link Performance of cdma2000 1xEV-DO Revision A System. In *Proceedings of 2005 IEEE International Conference on Communications (ICC '05)* (Seoul, Korea, May 16 - 20). IEEE, New York, 2005, pp. 2244 - 2250.

[20] Ferrari, D. and Verma, D. C. A Scheme for Real-time Channel Establishment in Wide-area Networks. *IEEE Journal of Selected Areas in Communications 8*, 3 (1990), 368-379.

[21] Garey, M. R. and Johnson, D. S. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, New York, 1979.

[22] Garg, V. K. *CDMA IS-95 and cdma2000*. Prentice Hall, New Jersey, 2000.

[23] Garg, V. K. *Wireless Communications and Networking*. Morgan Kaufmann Publishers, Massachusetts, 2007.

[24] Haleem, M. A. and Chandramouli, R. Adaptive Downlink Scheduling and Rate Selection: a Cross Layer Design. *IEEE Journal on Selected Areas in Communications 23*, 6 (2005), 1287-1297.

[25] Ho, Q.-D., Ashour, M., and Le-Ngoc, T. Channel and Delay Margin Aware Bandwidth Allocation for Future Generation Wireless Networks. In *Proceedings of IEEE Global Telecommunications Conference, 2008 (GLOBECOM '08)* (New Orleans, LA, Nov. 30 - Dec. 4). IEEE, New York, 2008, pp. 1-5.

[26] Hueda, M. R., Rodríguez, C., and Marqués, C. Enhanced-performance Video Transmission in Multicode CDMA Wireless Systems using a Feedback Error Control Scheme. In *Proceedings of IEEE Global Telecommunications Conference, 2001 (GLOBECOM '01)* (San Antonio, TX, Nov. 25 - 29). IEEE, New York, 2008, pp. 619-626.

[27] Iera, A., Molinaro, A., and Pizzi, S. Channel-Aware Scheduling for QoS and Fairness Provisioning in IEEE 802.16/WiMAX Broadband Wireless Access Systems. *IEEE Network 21*, 5 (2007), 34-41.

[28] Johnson, D. S., Demers, A. J., Ullman, J. D., Garey, M. R., and Graham, R. L. Worst-Case Performance Bounds for Simple One-Dimensional Packing Algorithms. *SIAM Journal on Computing 3*, 4 (1974), 299-325.

[29] Joshi, N., Kadaba, S. R., Patel, S., and Sundaram, G. S. Downlink Scheduling in CDMA Data Networks. In *Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking (MobiCom '00)* (Boston, MA, Aug. 6 - 11). ACM, New York, 2000, pp. 179-190.

[30] Khattab, A. K. F. and Elsayed, K. M. F. Opportunistic Scheduling of Delay Sensitive Traffic in OFDMA-basedWireless Networks. In *Proceedings of 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM '06)* (Buffalo, NY, Jun. 26 - 29). IEEE, New York, 2006, pp. 10-19.

[31] Kim, D. H. and Kang, C. G. Delay Threshold-based Priority Queueing Packet Scheduling for Integrated Services in Mobile Broadband Wireless Access System. In *Proceedings of the First International Conference on High Performance Computing and Communications (HPCC '05)* (Sorrento, Italy, Sep. 21 - 23). Springer, New York, 2005, pp. 305-314.

[32] Kim, J. G. and Krunz, M. M. Bandwidth Allocation in Wireless Networks with Guaranteed Packet-loss Performance. *IEEE/ACM Transactions on Networking 8*, 3 (2000), 337-349.

[33] Kim, J. H., Choi, B.-Y., and Beard, C. QoS and Channel Aware Packet Bundling for VoIP Traffic in Cellular Networks. In *Proceedings of the 21st International Teletraffic Congress (ITC 21)* (Paris, France, Sep. 15 - 17). IEEE, New York, 2009, pp. 1-8.

[34] Krishnam, M., Reisslein, M., and Fitzek, F. An Analytical Framework for Simultaneous MAC Packet Transmission (SMPT) in a Multi-Code CDMA Wireless System. *IEEE Transactions on Vehicular Technology 53*, 1 (2004), 223-242.

[35] Kwon, H., Kim, Y., Han, J.-K., and Kim, D. An Efficient Radio Resource Management Technique for the Reverse Link in cdma2000 1xEV-DV. In *Proceedings of 2005 IEEE Wireless Communications and Networking Conference (WCNC '05)*

(New Orleans, LA, Mar. 13 - 17). IEEE, New York, 2005, pp. 364-368.

[36] Liang, Z., Chew, Y. H., and Ko, C. C. A Linear Programming Solution to Subcarrier, Bit and Power Allocation for Multicell OFDMA Systems. In *Proceedings of 2008 IEEE Wireless Communications and Networking Conference (WCNC '08)* (Las Vegas, NV, Mar. 31 - Apr. 3). IEEE, New York, 2008, pp. 1273-1278.

[37] Lin, S. and Yu, P. S. A Hybrid ARQ Scheme with Parity Retransmission for Error Control of Satellite Channels. *IEEE Transactions on Communications 30*, 7 (1982), 1701-1719.

[38] Lin, X., Shroff, N. B., and Srikant, R. A Tutorial on Cross-layer Optimization in Wireless Networks. *IEEE Journal on Selected Areas in Communications 24*, 8 (2006), 1452-1463.

[39] Lott, C., Bhushan, N., Ghosh, D., Attar, R., Au, J., and Fan, M. Reverse Traffic Channel MAC Design of cdma2000 1xEV-DO Revision A System. In *Proceedings of IEEE 61st Semiannual Vehicular Technology Conference (VTC '05 Spring)* (Stockholm, Sweden, May 30 - Jun. 1). IEEE, New York, 2005, pp. 1416-1421.

[40] Nagle, J. On Packet Switches with Infinite Storage. *IEEE Transactions on Communications 35*, 4 (1987), 435-438.

[41] Shakkottai, S., Srikant, R., and Stolyar, A. Pathwise Optimality and State Space Collapse for the Exponential Rule. In *Proceedings of 2002 IEEE International Symposium on Information Theory (ISIT '02)* (Lausanne, Switzerland, Jun. 30 - Jul. 5). IEEE, New York, 2002, pp. 379.

[42] Shakkottai, S. and Stolyar, A. L. Scheduling for Multiple Flows Sharing a Time-varying Channel: The Exponential Rule. *American Mathematical Society Translations, Series 2, 227* (2002), 185-202.

[43] So-In, C., Jain, R., and Tamimi, A.-K. Scheduling in IEEE 802.16e Mobile WiMAX Networks: Key Issues and a Survey. *IEEE Journal on Selected Areas in Communications 27*, 2 (2009), 156-171.

[44] Srinivasan, R. *Scheduling in Packet Switched Cellular Wireless Systems*. Ph.D Thesis, University of Maryland, College Park, 2004.

[45] Viswanath, P., Tse, D. N. C., and Laroia, R. Opportunistic Beamforming using Dumb Antennas. *IEEE Transactions on Information Theory 48*, 6 (2002), 1277-1294.

[46] Walke, B. H. *Mobile Radio Networks: Networking, Protocols and Traffic Performance*. John Wiley, West Sussex, U.K., 2002.

[47] Wang, F., Ghosh, A., Sankaran, C., Fleming, P. J., Hsieh, F., and Benes, S. J. Mobile WiMAX Systems: Performance and Evolution. *IEEE Communications Magazine 46*, 10 (2008), 41-49.

[48] Wang, Y., Chan, S., Zukerman, M., and Harris, R. J. Priority-Based Fair Scheduling for Multimedia WiMAX Uplink Traffic. In *Proceedings of 2008 IEEE International Conference on Communications (ICC '08)* (Beijing, China, May 19 - 23). IEEE, New York, 2008, pp. 301-305.

[49] Wu, Q. and Esteves, E. *Advances in 3G enhanced technologies for wireless communications*. edited by Wang, J. and Ng, T.-S., Artech House, Boston, 2002, Chapter

4.

[50] Xiaoyi, W. and Yajiang, Q. Enhanced Proportional Fair Scheduling for cdma2000 1x EV-DO Reverse Link. In *Proceedings of 2006 IFIP International Conference on Wireless and Optical Communications Networks (WOCN '06)* (Bangalore, India, Apr. 11 - 13). IEEE, New York, 2006, pp. 29-34.

[51] Yajiang, Q., Chunye, W., and Xiaoyi, W. Scheduling for Multi-user Packet in CDMA2000 1x EV-DO. In *Proceedings of the 2nd International Conference on Mobile Technology, Applications and Systems* (Guangzhou , China, Nov. 15 - 17). IEEE, New York, 2005, pp. 23-28.

[52] Yavuz, M., Diaz, S., Kapoor, R., Grob, M., Black, P., Tokgoz, Y., and Lott, C. VoIP over cdma2000 1xEV-DO revision A. *IEEE Communications Magazine 44*, 2 (2006), 50-57.

[53] Zhang, L. Virtual Clock: a New Traffic Control Algorithm for Packet Switching Networks. *ACM SIGCOMM Computer Communication Review 20*, 4 (1990), 19-29.

[54] 3GPP2 C.R1002-0 v1.0, cdma2000 Evaluation Methodology. Dec., 2004. from 3GPP2: http://www.3gpp2.org/public_html/specs/C.R1002-0_v1.0_041221.pdf.

[55] 3GPP2 C.S0024-0 v2.0, cdma2000 High Rate Packet Data Air Interface Specification. Oct., 2000. from 3GPP2: http://www.3gpp2.org/public_html/specs/C.S0024_v2.0.pdf.

[56] IEEE 802.16-2004, Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems. Oct., 2004. from IEEE: http://standards.ieee.org/getieee802/download/802.16-2004.pdf.

[57] IEEE 802.16-2005, 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. Feb., 2006. from IEEE: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1603394.

[58] TIA 45.5, The cdma2000 ITU-R RTT Candidate Submission. Apr., 1998. from TIA: http://www.tiaonline.org/standards/technology/documents/TR455_RTT_V18.pdf.

[59] TIA/EIA IS-871, Markov Service Option (MSO) for cdma2000 Spread Spectrum Systems. Apr., 2004. from TIA: http://www.tiaonline.org/standards/technology/cdma2000/documents/TIA-EIA-IS-871.pdf.

[60] TIA IS-127, Enhance Variable Rate Codec (EVRC). Dec., 1999. from 3GPP2: http://www.3gpp2.org/public_html/specs/C.S0014-0_v1.0_revised.pdf.

VITA

Jung Hwan Kim was born on April 12, 1970 at Seoul in Korea. He graduated from Kyeonggi Science High School in 1986, at Kyeonggi-Do in Korea. He received a Bachelor of Science degree in electrical engineering in 1993 from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea.

He researched for Computer Engineering and Research Lab in KAIST in 1993 and worked for Apex Systems from 1994 to 1996. He won Jang Young-Sil Prize awarded by Department of Science and Technology, Korea in 1996. He continued his career at Oracle, Korea from 1997 to 1999.

In 1999, Mr. Kim joined in Computer Science at University of Missouri-Kansas City (UMKC) and received a Master degree in 2001. He continued to work toward the Ph.D. degree in computer science at UMKC.

He researched WiMAX capacity planning for Sprint in 2008. He is currently a Lead Engineer with Openet, Reston, Virginia since 2009. His research interests include scheduling algorithms and simulation techniques for wireless telecommunication systems.