제07-001호

2007년도 SWCC
최우수 논문상

제목 : End-to-End QoS Enhancement
in Mobile WiMAX Systems

저자 : 추성민·오성민·김재현(아주대학교),
조성현(스탠포드대학교)

위 논문은 2007년도 SWCC 최우수 논문으로 선정되었기에 이 상장과 상금을 수여함.

2007년 8월 24일

사단법인 한국정보과학회
정보통신연구회 위원장
공학박사 김 종 권

Mobile WiMAX 시스템에서의 종단간 서비스 품질 향상

추상민°[1], 오성민[1], 김재현[1], 조성현[2]
[1]아주대학교 전자공학과
{choosm, smallb01, jkim}@ajou.ac.kr
[2]스탠포드대학교 전자공학과
dr.shcho@gmail.com

End-to-End QoS Enhancement in Mobile WiMAX Systems

Sang-Min Choo°[1], Sung-Min Oh[1], Jae-Hyun Kim[1] and Sunghyun Cho[2]
[1]School of Electrical and Computer Engineering, Ajou University
[2]Department of Electrical Engineering, Stanford University

요 약

본 논문에서는 mobile WiMAX 시스템에서 네트워크 구조에 의해 영향을 받을 수 있는 서비스 품질 기능들에 대해 장단점을 논의하고 계층간 서비스 클래스 매핑 기능 및 스케줄링 알고리즘의 예를 제안한다. 종단간 서비스 품질에 관한 성능을 평가하기 위하여, OPNET을 이용하여 mobile WiMAX의 종단간 시뮬레이터를 구현하였다. 시뮬레이션 결과에 의하면, 현재 mobile WiMAX에서의 종단간 상향링크 처리율은 물리 계층 처리율의 70% 정도 밖에 되지 않으며, VoIP 서비스의 종단간 패킷 전송 지연은 제안한 방식을 적용할 경우 44-67%정도 줄어 들었다.

Abstract

In this paper, we discuss the strong and weak points of QoS related functions, which can be affected by the network architecture in mobile WiMAX system, and propose the design examples of a cross layer service class mapping function and scheduling algorithm. To evaluate the end-to-end QoS, we implemented an end-to-end simulator of mobile WiMAX using OPNET. Simulation results show that the end-to-end uplink throughput of the current mobile WiMAX is about 70 percent of the uplink PHY throughput and the end-to-end packet transmission delay of VoIP services can be decreased by 44–67 percent through the proposed schemes.

1. Introduction

Consumers have grown accustomed to fast Internet accesses. This expectation is driving quality expectations for wireless Internet accesses that rival wired accesses. Based on these needs, the IEEE approved the Project Authorization Request (PAR) 802.16e and started a standardization effort 2002. Recently, IEEE 802.16e-2005 was approved by IEEE-SA in 2005 and published in February of 2006 [1]. IEEE 802.16e specifies the physical and medium access control layers for providing combined fixed and mobile broadband wireless access to subscriber stations in licensed bands. Since IEEE 802.16 specifies many different modes of operation, making eventual equipment compatibility questionable. As a result, industry forums, like WiBro in Korea and WiMAX in North America and Europe, have been developed to specify only a subset of functions in profiles.

• This work was supported by the IT R&D program of MIC/IITA. [2006-S058-01, Development of Network/Service Control Technology in All-IP based Converged network]

• "This research was supported by the MIC(Ministry of Information and Communication), Korea, under the ITRC(Information Technology Research Center) support program supervised by the IITA(Institute of Information Technology Advancement)" (IITA-2006-(C1090-0602-0011))

In recent years, many research results have been published on IEEE 802.16e and mobile WiMAX systems [1-7]. For the most part, a prior research focuses on the physical (PHY) layer and medium access control (MAC) layer. In [2], cross layer protocols between MAC and PHY layers in IEEE 802.16e systems are proposed. A cross-layer adaptation framework and a design example of primitives for cross-layer operation are introduced in [2]. In addition, a MAC and PHY cross-layer considered simulator is introduced in [2]. In [3], QoS issues of IEEE 802.16 networks are investigated. Mechanisms for supporting QoS at the IEEE 802.16 MAC layer are reviewed and analyzed in [3]. These papers, however, focus only MAC and PHY layers of IEEE 802.16 systems. Unfortunately, the upper layer throughput of mobile WiMAX systems is relatively reduced compared with the lower layer throughput in the same way as other mobile communication systems because of the overhead between different layers. Using computer simulation, Kwon et al. show that the MAC layer downlink (DL) and uplink (UL) throughput of IEEE 802.16e systems is only about 75 percent and 90 percent of the PHY layer DL and UL throughput, respectively [2]. Moreover, the end-to-end throughput of mobile WiMAX systems is very likely to be less than the MAC layer

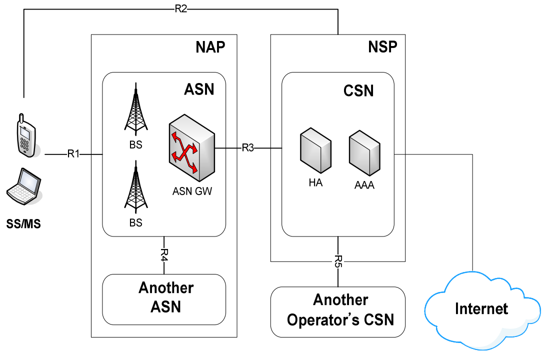


Figure 1. Network reference model in Mobile WiMAX

throughput due to the IP and upper layer overhead. Based on our simulation results, the end-to-end throughput of the mobile WiMAX systems is around 78 percent of the MAC layer throughput. It means that the end-to-end QoS may not be guaranteed even if the MAC layer QoS is definitely guaranteed. Therefore, a cross layer optimized design is necessary not only between MAC and PHY layers but also between the upper layer and MAC layer to guarantee the end-to-end QoS requirements. This topic, however, has not been actively studied until yet.

Consequently, we focus on the end-to-end QoS issues of the mobile WiMAX systems in this paper. The rest of the paper is organized as follows. In Section II, we review the end-to-end network architecture of mobile WiMAX systems, which has an effect on the end-to-end QoS. Section III presents the proposed end-to-end QoS model for mobile WiMAX systems. In Section IV, we present the end-to-end simulator and the performance analysis and the concluding remarks are in Section V.

2. Mobile WiMAX End-to-End Network Architecture

2.1. Network Reference Model

In mobile WiMAX, the network architecture is divided into Access Service Network (ASN) and Connectivity Service Network (CSN) [6]. ASN and CSN are defined as functional entities which perform specific functions in a network. An ASN is defined as a set of functions that provides the mobile station (MS) with radio access functions. A CSN consists of network functions to provide IP connectivity services for WiMAX subscribers. Figure 1 shows the Network Reference Model (NRM) of mobile WiMAX, a logical representation of the network architecture. In Figure 1, the NAP and NSP are the network access provider and the network service provider, respectively. The NAP is a unit that provides the radio access infrastructure to NSP. The NSP can be divided into the Home NSP and the Visited NSP from the point-of-view of the WiMAX subscriber. The Home NSP provides service level agreements, authentication, authorization, and billing to its subscriber. On the other hand, the Visited NSP routes data or controls traffic of roaming subscribers to a home NSP. The NAP and NSP consists of one or more ASNs and a CSN, respectively. An ASN defines the functional entities and corresponding message flows associated with the access services. It may consist of one or more base stations (BS) and an ASN gateway (ASN-GW). A CSN may comprise network elements such as routers, AAA proxy/servers, user databases and inter-working gateway devices. The followings are details of the ASN and CSN functions defined by WiMAX Forum [6]:

- ASN Functions
  - Layer 2 connectivity with MS
  - Transfer of AAA messages to WiMAX subscriber's NSP for authentication, authorization and session accounting for subscriber sessions
  - Network discovery and selection of the WiMAX subscriber's preferred NSP
  - Relay functionality for establishing Layer-3 (L3) connectivity with a WiMAX MS
  - Radio Resource Management
  - ASN anchor mobility
  - CSN anchor mobility
  - Paging and Location Management
  - ASN-CSN tunneling

- CSN Functions
  - MS IP address and endpoint parameter allocation for user sessions
  - Internet access
  - AAA proxy or server
  - Policy and Admission Control based on user subscription profiles
  - ASN-CSN tunneling support
  - Inter-CSN tunneling for roaming
  - Inter-ASN mobility
  - Connectivity to IP multimedia services

WiMAX NRM also defines five reference points for interconnection of the logical entities. A reference point defines protocols and procedures between different functional entities. The details of each reference point are defined as follows [6]:

- R1
  - Includes the protocols and procedures between WiMAX subscriber and ASN
  - Depends on IEEE 802.16e-2005 air interface specifications
- R2
  - Consists of protocols and procedures between the MS and CSN associated with Authentication, Services Authorization and IP Host Configuration management
- R3
  - Consists of the set of control plane and protocols between the ASN and the CSN to support AAA, policy enforcement and mobility management capabilities

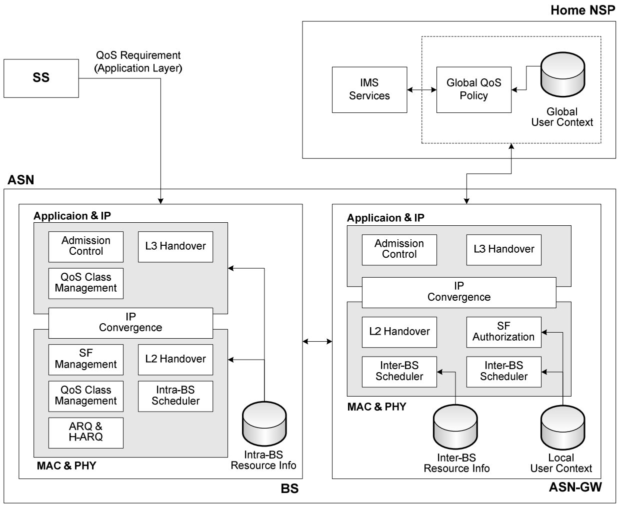


Figure 2. End-to-End QoS model in mobile WiMAX

- Encompasses the bearer plane methods to transfer user data between the ASN and the CSN

- R4
  - Consists of the set of Control and Bearer plane protocols originating/terminating in various functional entities of an ASN that coordinate MS mobility between ASNs and ASN-GWs
  - The only interoperable RP between similar or heterogeneous ASNs
- R5
  - Consists of the set of control plane and bearer plane protocols for internetworking between the CSN operated by the home NSP and that operated by a visited NSP

2.2. ASN Profile A,B, and C

As described in the previous section, ASN may comprise one or more BSs and an ASN-GW. We notice that a BS and an ASN-GW can be implemented as one physical node or separate physical nodes. The end-to-end QoS can be highly influenced by the ASN architecture. For this reason, we introduce the specific ASN architecture of mobile WiMAX systems. The Network Working Group (NWG) in WiMAX Forum defines three types of ASN architecture such as ASN Profile A, B, and C in NWG release 1 specification [6]. The Profile A and C are 2-tier network architectures which consist of BS and ASN-GW separately while the Profile B is a 1-tier network architecture which composes a unified BS and ASN-GW node. The 2-tier network architecture is a centralized ASN model in which an ASN-GW can control one or more BSs. On the other hand, the 1-tier network architecture is a distributed ASN model because there is no hierarchy among the unified BS and ASN-GW nodes. The main difference between Profile A and Profile C is the radio resource management (RRM) function. In Profile A, RRM is divided into radio resource authorization (RRA) and radio resource control (RRC) which are located in a BS and an ASN-GW, respectively. On the other hand, the whole RRM function is performed by a BS in Profile C.

3. End-to-End QoS in Mobile WiMAX

3.1. End-to-End QoS Model in Mobile WiMAX

In this section, we propose an end-to-end QoS model for mobile WiMAX systems. An end-to-end

QoS model can be defined in different ways based on the network architecture. For this reason, the network architecture is very important in both research and standardization efforts. Research thus far shows that the 2-tier architecture is typically more profitable than the 1-tier architecture for enhancing the end-to-end QoS performance [4-8]. In case of mobile WiMAX, the network is most likely to be deployed as a 2-tier architecture in the first stage. Therefore, we design an end-to-end QoS model based on the ASN Profile A and C. Figure 2 shows the proposed end-to-end QoS model of the mobile WiMAX systems. As shown in Figure 2 [6, 7], a MS can request various kinds of application level QoS requirements according to the application types and user preferences when it opens a new session. These application level QoS requirements can be interpreted as end-to-end QoS parameters. The representative end-to-end QoS parameters are as follows:

- End-to-End Data Rate
- End-to-End Packet Transmission Delay
- End-to-End Packet Transmission Delay Jitter
- End-to-End Packet Loss Rate
- Call Dropping Rate

Some QoS related functions can be implemented in a BS or an ASN-GW based on the network architecture and an end-to-end QoS policy. For example, service flow related functions, resource management functions, mobility management functions, and IP convergence function can be located in a BS or an ASN-GW according to the network architecture. Notice that the location of the QoS related functions can make a great difference to an end-to-end QoS performance.

An end-to-end QoS policy can be also affected by a cross layer design, especially IP layer and MAC layer. Kwon et al. show cross layer design issues between MAC and PHY layers to boost the performance of the IEEE 802.16e systems in [2]. However, it is mainly focused on the QoS issues of the air link between MS and BS. To improve the end-to-end QoS, cross layer design issues between upper layers including IP layer and MAC layer should be considered as well. Service class mapping methods or QoS parameter mapping schemes between mobile WiMAX MAC layer and IP layer correspond to the issues. The service classes defined in IP layer are in fact very different from the service classes in mobile WiMAX MAC layer. Therefore, the end-to-end performance can be varied according to the service class mapping functions.

3.2. End-to-End QoS variation in 2-tier vs. 1-tier Network Architecture

QoS related functions in the proposed end-to-end QoS model can be separated in a BS and an ASN-GW or located in one unified physical node. Since the location of the QoS functions can have an impact on the end-to-end QoS of mobile WiMAX systems, it is a significant issue to analyze an effect of network architecture for the end-to-end QoS. In this section, we show that the QoS functions and investigate the end-to-end QoS variation by the QoS functions and network architecture. The following functions are good representatives which can be affected by network architecture:

- L3 Handover Function
- L2 Handover Function
- ARQ Function
- Admission Control
- Scheduler
- Inter-cell Interference Mitigation (IIM) Function

An L2 handover function can have an effect on the end-to-end QoS in terms of end-to-end packet transmission delay, delay jitter, and packet loss rate. It is more desirable to locate an L2 handover function in BSs regardless of the network architecture since it performs air link related functions such as periodic measurement of a signal strength and report. On the other hand, it is more advantageous to locate an L3 handover function into a ASN-GW in 2-tier architecture to improve the end-to-end QoS. In 1-tier network architectures, each BS makes its own subnet. In this case, a handover user should request a new IP address when it moves to a target cell. On the other hand, an ASN-GW can make a subnet that includes one or more BSs in 2-tier network architecture. If a MS travels to a target BS which is controlled by the same ASN-GW as a serving BS, then it can use the current IP address in the target BS. Therefore, in case of L3 handover, the 2-tier network architecture is better than the 1-tier architecture for improvement in the end-to-end QoS parameters such as end-to-end packet transmission delay and delay jitter [9].

An ARQ function is closely related with the handover function in consideration of the end-to-end QoS or network efficiency. In case where handovers seldom occur, QoS is not affected by the network architecture because a BS can efficiently control an ARQ process regardless of the network architecture. However, 2-tier network architecture is more desirable to support an ARQ process in consideration of handover [10, 11]. If an ASN-GW controls an ARQ function in 2-tier network architecture, then the handover can be performed without an additional delay caused by ARQ processes. An additional handover delay, however, can be required in 1-tier network architecture when the packet transmission error is occurred. In 1-tier network architecture, the handover process is delayed until the retransmission process is completed because only the serving BS knows the current ARQ state and buffers the required packets in its queue. The packet error rate of a handover user is relatively high and the retransmission process can be frequently required. Consequently, the end-to-end packet transmission delay and delay jitter can be increased in 1-tier network architecture because of an additional handover delay. Recently, many research results for this issue were contributed to 3GPP LTE RAN1 [10, 11]

An admission control function also has an effect on the end-to-end QoS in terms of call dropping rate. To reduce call dropping rate, it is more favorable to locate an admission control function in an ASN-GW. In this case, an ASN-GW can make use of load balancing schemes to reduce the call dropping rate [12]. The 1-tier network architecture, however, is not efficient for load balancing because the loading information of the neighbor cells may not be easily shared.

A scheduler is one of the most important functions to improve the end-to-end QoS. A scheduler can be located in a BS or an ASN-GW, or both of them. It is easy to use multiple input and multiple output (MIMO) antenna technologies or adaptive modulation and coding (AMC) when a BS has scheduling function. The reason is that the BS can reflect the real time channel conditions of each user to the scheduling functions without an additional backhaul delay. In this case, we can expect an enhancement not only in throughput but also in end-to-end data rates. On the other hand, it is not easy to use real time channel information in a scheduler when located in an ASN-GW. The scheduler, however, can increase the overall system throughput by resource allocation with inter-cell coordination. This effect grows larger when every BS uses the same frequency. Therefore, the location and types of scheduler which are desirable to guarantee the end-to-end QoS can be different by the system and applications.

An Intercell interference mitigation (IIM) function can be considered as a physical layer issue. An IIM has an effect on the link quality of cell edge users and the MAC layer cell throughput. We also figure out that an end-to-end data rate is highly affected by the MAC layer throughput. An IIM function can have an impact on the performance variation of end-to-end QoS parameters such as end-to-end data rate and packet loss ratio even if it is commonly implemented in MAC and PHY layers. Therefore, cross layer optimized design from PHY layer to application layer is required to guarantee the end-to-end QoS. The existing IIM schemes can be classified into distributed algorithms and centralized algorithms [11]. In the distributed algorithms, an IIM is usually located in a BS. The distributed algorithms are performed under the predetermined assumption. Therefore, it has relatively low flexibility than the centralized algorithms. The centralized IIM is implemented in a higher level node such as an ASN-GW and suitable for the 2-tier network architecture. It can cope efficiently with the variation of the system parameters or cell loading. The performance of an IIM functions can be varied due to the system parameters such as fre-

Table 1. Comparisons of strengths and weakness of end-to-end QoS functions according to the ASN models

QoS Functions	2-Tier ASN Model (ASN Profile A, C)	1-Tier ASN Model (ASN Profile B)	Related End-to-End QoS Parameters
L3 Handover Function	Positive if it is operated in ASN-GW	Negative	Transmission Delay Delay Jitter Packet Loss Rate
L2 Handover Function	Positive if it is operated in BS	Positive if it is operated in BS	Transmission Delay Delay Jitter Packet Loss Rate
ARQ Function	Positive for handover process if it is operated in ASN-GW	Negative for handover Process	Transmission Delay Delay Jitter Packet Loss Rate
Admission Control	Positive if it is operated in ASN-GW	Negative	Call Dropping Rate
Scheduler	Positive if intercell coordination is required	Positive if fast feedback is required	Data Rate Transmission Delay
Intercell Interference Mitigation (IIM) Function	Positive to Centralized Scheme	Positive to Distributed Scheme	Data Rate Packet Loss Rate

quency reuse factor, cell structures, cell loading, etc [11]. To improve an end-to-end QoS, we should carefully select an appropriate IIM algorithm with regard to network architectures.

In Table 1, the correlation among network architecture, the QoS related functions, and the-end-to-end QoS are investigated. The 2-tier network architecture can be defined as the centralized network architecture because hierarchically upper node such as an ASN-GW can control one or more lower nodes such as BS. Therefore, WiMAX Profile A and Profile C are advantageous to enhance the end-to-end QoS in case where the inter-cell coordination is required. The overhead in ASN, however, such as backhaul delay and unnecessary control packet transmission can be increased in the 2-tier architecture. On the other hand, the 1-tier network architecture can be defined as distributed network architecture or flat network architecture because there is no hierarchy among different BSs. Thus, the 1-tier architecture can reduce the overhead in ASN since the QoS related user context, parameters, and functions are saved and controlled by a unified node such as BS with ASN-GW functions. Table 1 compares strength and weakness of the end-to-end QoS functions according to the ASN Profiles and lists up the related end-to-end QoS parameters.

3.3. The Proposed Schemes for the End-to-End QoS Enhancement

3.3.1. Service Class Mapping Functions

In ASN and IP Network, QoS should be guaranteed for the end-to-end QoS provisioning, respectively. For the QoS guaranteed service, WiMAX system divides a traffic into several service flows in order to improve the end-to-end QoS performance. When a service flow is generated, the resource negotiation between MS and BS or ASN-GW is performed to guarantee the QoS of service flow. After the resource negotiation procedure per service flow is established, a packet is transmitted from MS to ASN-GW. When the packet is transmitted through ASN, the packet is distinguished by five service classes defined in WiMAX. However, a problem can be occurred when ASN-GW forwards the packet to IP network, because service classes defined in IP network are different from those in WiMAX. In IP network, the QoS provisioning mechanism as differentiated service (DiffServ) is generally considered. DiffServ uses thirteen service classes based on the per-hop behavior as well as the priority and drop precedence of service classes. If service classes are not mapped to the corresponding service classes between DiffServ and WiMAX, the QoS service level of the packet can be changed. For example, if a voice packet is mapped to the lowest priority packet such as web browsing or FTP in ASN-GW, the QoS of the voice packet can not be guaranteed in IP network. Hence, it is necessary to apply the service class mapping function to ASN-GW for end-to-end QoS provisioning.

WiMAX and DiffServ define each service class considering QoS provisioning mechanism. WiMAX systems support the five service classes such as unsolicited grant service (UGS), real-time polling service (rtPS), extended real-time polling service (ertPS), non-real-time polling service (nrtPS), and best effort (BE). First, the UGS is designed to support real-time service flows that generate fixed size data packets on a periodic basis, such as T1/E1 and voice over IP (VoIP) without silence suppression. Second, the rtPS is designed to support real-time service flows that generate variable size data packets on a periodic basis, such as moving pictures experts group (MPEG). Third, the ertPS is designed to support real-time service flows that generate variable size data packets on a periodic basis, such as VoIP service with silence suppression. Fourth, the nrtPS is designed to support delay-tolerant service flows consisting of variable size which a minimum data rate is required, such as file transfer protocol (FTP). Fifth, the BE is designed to support service flows for which no minimum service level is required and therefore may be handled on a space-available basis [1].

In IP network, two QoS provisioning mechanisms such as integrated service (IntServ) and DiffServ are considered. However, IntServ mechanism has a scalability problem, because all nodes in end-to-end network should support end-to-end resource reservation protocol. For this reason, DiffServ mechanism is considered as the primary QoS provisioning mechanism in IP network. DiffServ uses the type of service (ToS) field of IP packet header for classifying the service classes. The DiffServ codepoint (DSCP) consists of six bits in ToS field. Codepoint is defined as a specific value of the DSCP and per-hop behavior (PHB) is combined to form a specified set of characteristics for handling different kinds of traffic, depending on the needs of the application. The standardized PHBs include expedited forwarding (EF) and assured forwarding (AF). The EF is defined for low-loss, low-delay, and low-jitter service class. The AF is defined for an enhanced best-effort service: traffic is expected to be elastic in nature. The receiver will detect loss or variation in delay in the network and provide feedback such that the sender adjusts its transmission rate to approximate available capacity. The AF allows four service classes of burst traffic for a router queue assignment and three drop precedence levels. In case of AF41, the four indicates the highest priority service class and the one presents the lowest drop precedence [13].

Based on the service classes mentioned above, we propose the service class mapping between DiffServ and WiMAX systems. The EF is a good candidate to be mapped to the UGS and ertPS for the voice traffic because of the similarity of QoS requirements. AF service classes can be mapped to rtPS and nrtPS. rtPS is defined to support delay sensitive services such as a video streaming services. Therefore, it should be mapped to AF41 to maintain the minimum delay in IP network. nrtPS can also be mapped to AF. However, nrtPS is well matched for AF31, AF21, and AF11 since it supports delay insensitive services such as FTP. The followings show the proposed service class mapping pairs between DiffServ and WiMAX systems with representative applications:

- EF ↔ UGS/ertPS (Voice Services)
- AF41 ↔ rtPS (Video Streaming)
- AF31, AF21, AF11 ↔ nrtPS (FTP)
- BE ↔ BE (HTTP)

3.3.2. IP Packet Scheduler

As mentioned previously, an IP packet scheduler can be applied to a BS or an ASN-GW. In case of an IP packet scheduler in BS, it may not be useful because the optical cable or Gigabit Ethernet is considered as the link between a BS and an ASN-GW. However, the IP packet scheduler is required for the QoS provisioning in an ASN-GW because the ASN-GW is directly connected to IP network and IP network can be congested with burst traffic.

We consider several IP packet schedulers including first-in first-out (FIFO), weighted round robin (WRR), and dynamic processor sharing (DPS). Since the voice traffic is very sensitive to delay and jitter, DPS allocates the resource to voice traffic above all in order to guarantee the QoS. When a voice packet arrives at DPS, DPS gives the highest priority to the voice packet. Therefore, the voice packet can be transmitted without a queuing delay. To improve the QoS performance, we set the weight factor using the delay budget which can be obtained from QoS requirement in [14]. To avoid the unfairness among voice services and other service classes, CAC is applied to DPS. In section 4, we show that DPS is more suitable for the packet scheduler in an ASN-GW than FIFO or WRR in order to guarantee the required QoS.

4. Simulation and Performance Evaluation

4.1. Implementation of an End-to-End Simulator

In this paper, we modeled an end-to-end simulator to evaluate the end-to-end QoS performance of mobile WiMAX system using OPNET. Figure 3 represents the network model and node models of the end-to-end simulator. We assumed that an ASN network architecture is 2-tier type defined in mobile WiMAX ASN profile A/C. The ASN network architecture consists of MS, BS, and ASN-GW, and the functions of BS and ASN-GW is based on ASN profile C.

We modeled PHY and MAC layer based on IEEE 802.16e-2005 specification [1,2], and built MS and BS node models using the IP, TCP/UDP, and application node modules provided by OPNET as well as PHY node module and MAC node module. As shown in Figure 3b, various protocols defined in application layer, TCP/UDP, and IP layer are applied to this simulator. When data are generated in application layer, these are transmitted to the lower layer as a form of packet. While a packet is delivered through each layer, it is encapsulated. When a packet arrives at MAC layer, a packet is fragmented according to modulation and coding scheme (MCS) level and scheduling information. By implementing these packet transmission procedures through all layers, we can obtain the performance analysis results for the overhead of each layer.

We simply implemented the connection identifier (CID) mapping, because the detailed functions related to the CID mapping have not been defined in IEEE 802.16e standard. When a packet is transmitted, the classifier allocates the packet to the corresponding CID in convergence sublayer (CS). In this case, the classifier should classify the packet using information only in the IP packet header such as source address, destination address, and type of service. We modeled that an IP packet can be mapped to a CID using destination address in the IP packet header, when an IP packet arrives at a BS or an ASN-GW. However, the classifier can not correctly sort a packet with a destination address when a MS has different CIDs. To resolve this problem, the classification mechanism should be investigated considering the relation between information of the IP packet header and CID characteristics. For the end-to-end QoS performance evaluation, we implemented the service class mapping be-

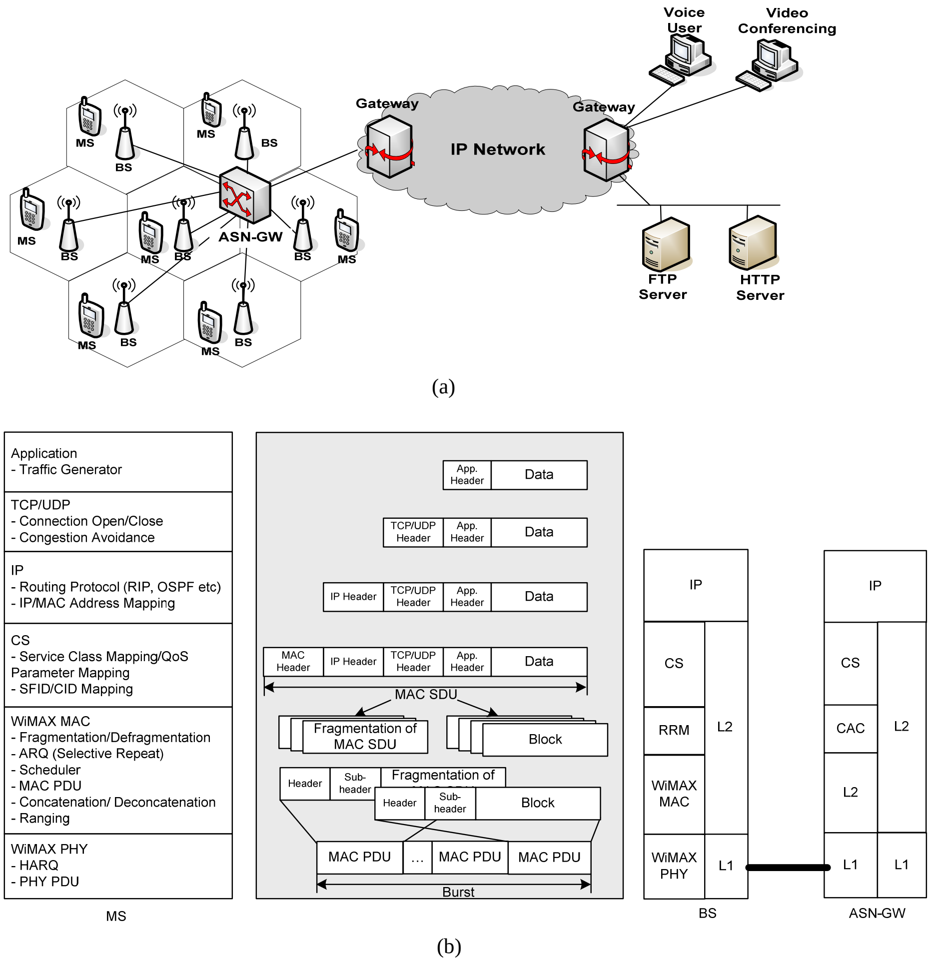


Figure 3. End-to-end Simulator: (a) Networks model of the end-to-end simulator and (b) Node models of MS, BS and ASN-GW

tween IP network and WiMAX in CS mentioned in the previous section, and applied FIFO, WRR, and DPS to ASN-GW node.

4.2. Simulation Environment

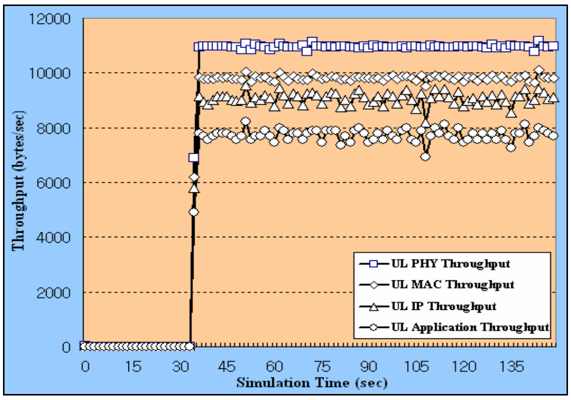
In order to build up the WiMAX simulation environment, we considered path loss, log-normal shadowing, and frequency-selective Rayleigh fading according to user's mobility. Also, we considered WiMAX system as IEEE 802.16e OFDMA system that uses 5msec TDD frame size. For modeling the multi-cell interference, we built seven hexagonal cells. We modeled three types of user's mobility such as stationary, pedestrian, and vehicular [2].

In order to implement the congestion environment in IP network, we applied a PPP-E1 link. If the IP network congests with burst traffic, the data transmission rate may decrease in IP network. For this reason, we simplified this phenomenon using a PPP-E1 link to connect between ASN-GW and gateway.

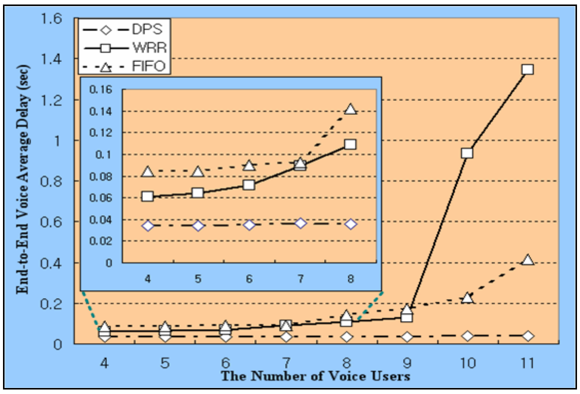
We used the following traffic models for evaluating the end-to-end QoS performance. First, we modeled the silence and talk spurt duration of VoIP using exponential distribution with mean 0.65 sec and 0.35 sec, respectively. We also applied G.711 (data rate: 64 Kbps) to the encoder scheme of VoIP. Second, we assumed that the frame size of video streaming is 18.2 Kbytes and frame inter-arrival time is 10 frame/sec. Third, we modeled that the file size of FTP follows the truncated log normal distribution with mean 2 Mbytes, and reading time is exponentially distributed with mean 180 sec. Fourth, we applied HTTP 1.1 protocol, and we modeled that page inter-arrival time is exponentially distributed with mean 30 sec. Object size follows the lognormal distribution with mean 10 Kbytes [15]

4.3. Performance Analysis and Discussion

In Figure 4a, we show the uplink (UL) throughput of PHY, MAC, IP, and application layers for mobile WiMAX system. We compare the UL throughput among different layers since the downlink (DL) throughput can be highly affected by the control information such as MAP messages. From the simulation results, we can obtain that the UL MAC layer throughput is about 12 percent lower than the UL PHY layer throughput. The UL IP layer throughput is only about 79 percent and 92 percent of the UL PHY and MAC layers throughput, respectively. Moreover, the simulation results indicate that the UL application layer throughput is approximately 30 and 22 percent lower than the UL PHY and



(a)



(b)

Figure 4. Simulation results: (a) The throughput of PHY, MAC, IP, and application layer and (b) End-to-end VoIP packet average transmission delay according to the IP packet schedulers

MAC layers throughput, respectively. From the results, we can figure out the capacity of application and IP layers from the given MAC and PHY throughput. For example, the user-perceived data rate in UL is only 0.7Mbps even if the UL data rate supported in PHY layer is 1Mbps. It means that the end-to-end QoS may not be guaranteed even if the MAC and PHY layer QoS is fully guaranteed due to the overhead of each layer. In addition, there exists protocol duplication between different layers. Automatic Repeat reQuest (ARQ) protocol can be a typical example of the protocol duplication. Each PHY, MAC, and TCP layer defines its own ARQ protocols. The purposes and algorithms of each ARQ protocol are diverse. However, we can reduce the signaling overhead between different layers through the cross-layer optimized ARQ protocol design. For the improvement of the end-to-end QoS in mobile WiMAX, we should take differences of the QoS classes and QoS parameters between different layers into consideration as well. As we described earlier, the kinds of IP service classes and MAC layer service classes are quiet different. Furthermore, the QoS parameters in IP layer can be different from the QoS parameters in mobile WiMAX MAC layer. The QoS functions, which can efficiently support the QoS class mapping and parameter mapping between different layers, should be defined to guarantee the required end-to-end QoS level.

To overcome this problem, we propose the service class mapping between IP and MAC layers and IP packet scheduler which can be applied to convergence sublayer in an ASN-GW. The proposed service class mapping function and DPS in an ASN-GW can improve the end-to-end QoS performance of VoIP services. To evaluate the performance of the proposed schemes, we assume that the available UL data rate for VoIP services is 1Mbps and the other UL resources are allocated to the background burst traffic.

Figure 4b indicates the end-to-end VoIP packet transmission delay according to kinds of IP packet scheduler. The required data rate of a VoIP session is 75 Kbps in PHY layer. Figure 4b shows that the end-to-end VoIP packet transmission delay in WRR and FIFO schedulers is rapidly increased when the number of VoIP users is more than nine users. On the other hand, the end-to-end VoIP packet transmission delay in DPS can be maintained stable because DPS allocates the highest priority to the VoIP packets. We can obtain that DPS can reduce the end-to-end VoIP packet transmission delay by 44 percent to 67 percent compared with WRR. Consequently, we can increase the VoIP service capacity through applying DPS in an ASN-GW. Many other QoS functions can influence the specific QoS parameters and the performance of the end-to-end QoS.

All the issues mentioned above are very important for improving the end-to-end QoS performance. The simulation results show that the cross layer optimization not only between MAC and PHY layers but also between upper layer and MAC layer is necessary in improving the end-to-end QoS performance. Moreover, schedulers and QoS mapping functions can highly affect the end-to-end QoS performance. Therefore, a careful and thorough design of QoS related functions is significant. In the future, we will study on other QoS related functions with network architecture and the protocol design methodology for these cross-layer issues in mobile WiMAX systems.

5. Conclusion

In this paper, we present the issues to improve the end-to-end QoS in mobile WiMAX systems. We review and analyze the mobile WiMAX network architecture since it can affect the QoS control functions and the performance of the end-to-end QoS. We also investigate the correlation between network architecture and QoS related functions, and compare the strength and weakness of the end-to-end QoS functions according to the WiMAX ASN Profiles. In addition, we propose a cross layer design example of the QoS related functions such as QoS class mapping between IP and MAC layers and IP scheduler for improving the quality of voice services. The analysis results and cross layer design examples can be utilized for the end-to-end QoS enhancement in mobile

WiMAX evolution systems. To evaluate the end-to-end QoS performance, we design and implement an end-to-end simulator of mobile WiMAX. The simulator is based on the WiMAX Profile C and implemented from PHY layer to application layer. We compare the throughput among PHY, MAC, IP, and application layer, and evaluate the effect of the proposed QoS related functions on the end-to-end QoS. The design and implementation methods of an end-to-end simulator can also be utilized for an efficient design of QoS functions in mobile WiMAX evolution systems.

References

[1] *IEEE Std 802.16e-2005, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, Feb. 2006.

[2] T. Kwon *et al*., "Design and implementation of a simulator based on a cross-layer protocol between MAC and PHY layers in a WiBro Compatible IEEE 802.16e OFDMA system," *IEEE Commun. Mag*., Dec. 2005, pp. 136-146.

[3] C. Cicconetti, L. Lenzini, E. Mingozzi, and C. Eklund, "Quality of Service Support in IEEE 802.16 Networks," *IEEE Network*, Mar. 2006, pp. 50-55.

[4] A. Ghosh, D. Wolter, J. Andrews, and R. Chen, "Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential," *IEEE Commun. Mag*., Feb. 2005, pp. 129- 136.

[5] C. Hoymann, "Analysis and Performance Evaluation of the OFDM-based Metropolitan Area Network IEEE 802.16," *Comp. Net*., vol. 49, no. 3, Oct. 2005, pp. 341-363.

[6] *WiMAX Forum NWG, WiMAX End-to-End Network System Architecture – Stage 2: Architecture Tenets, Reference Model and Reference Points,* Dec. 2005.

[7] *WiMAX Forum, Mobile WiMAX – Part 1: A Technical Overview and Performance Evaluation,* Aug. 2006.

[8] *3GPP TR 23.882 v1.4.2, 3GPP System Architecture Evolution: Report on Technical Options and Conclusions (Release 7),* Oct. 2006.

[9] J. Hu and C. Yang, "On the design of mobility management scheme for 802.16-based network environment," *Proc. IEEE VTC 2005 Fall*, vol. 2, Sep. 2005, pp. 720-724.

[10] *3GPP TR 25.813 v7.0.0, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN) – Radio interface protocol aspects (Release 7),* Jun. 2006.

[11] *3GPP TR 25.814 v7.1.0, Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA) (Release 7),* Sep. 2006.

[12] H. Wang, B. He, and D. Agrawal, "Admission Control and Bandwidth Allocation above Packet Level for IEEE 802.16 Wireless MAN," *Proc. IEEE ICPADS 2006*, vol. 1, Jul. 2006, pp. 599 – 604.

[13] *J. Babiarz, K. Chan, and F. Baker, Configuration Guidelines for DiffServ Service Classes,* RFC 4594, Aug. 2006.

[14] *3GPP TR 25.853 v0.4.0, Delay Budget within the Access Stratum (Release 4)*, Mar. 2006.

[15] *C50-20010802-02, 1xEv-DV Evaluation Methodology – Addendum,* Jul. 2001