

# DiffServ-Aware Pricing for Wireless Internet

Hoon Lee\*

## ABSTRACT

In this work we propose a new pricing scheme for the wireless Internet services over WiMAX system. First, let us review the characteristics of wireless network which is based on multi-hop relay WiMAX system. Next, we show why usage-based and QoS-aware pricing scheme is needed in the wireless Internet. After that, we propose a theoretical model for the price of multimedia services called a DAP (DiffServ-aware pricing) scheme for the WiMAX multimedia network which takes into account the consumed radio resource of WiMAX system as well as the supported QoS in the IP backbone network. Finally, we present explicit formulae for the packet price, price of consumed radio resource, and price of consumed bytes.

**Key words** : Wireless Internet, Quality of Service, DiffServ Internet, DiffServ-Aware Pricing

## I. Introduction

Recently, wireless Internet is popular throughout the world, and WiMAX (world-wide interoperability for microwave access) is considered to be one of the most efficient wireless MAN (metropolitan area network) technologies for the support of a large-scale wireless Internet.

On the other hand, there is a heated debate about the explosive increase in the wireless traffic and conflicting opinions about the price for wireless Internet service. Let us look at an example<sup>[1,2]</sup>. User *A* is subscribed to a wireless Internet with a monthly charge of 40,000Won (Korean currency), and this user consumed 50GB (gigabytes). User *B* is subscribed to the same service with the same charge, but this user consumed only 200MB (megabytes) of data. User *A* consumed 250 times more than that of user *B* at the same charge. Is it fair?

A shift in the usage of wireless Internet from 3G (third generation) phone to 4G (fourth generation) smartphone is popular, which results in explosion of mobile traffic<sup>[3]</sup>. It is known that

Korea is number one in the world concerning the amount of traffic consumed by a smartphone: data traffic per user is 271MB per month, which amounts to 3.2 times greater than that of world average<sup>[4]</sup>.

On the other hand, it is frequent to hear of customer complaints that voice conversations are disconnected while a call is going on in the densely-populated locations such as downtown shopping malls and public places. In [5], we have shown the problem of performance degradation by illustrating a decreased utility as the usage of users increases over the limited wireless resources.

It is known that the number of smartphone customers in Korea had exceeded 20 millions at the 3rd quarter of 2011 (twice the original estimation from [6]). It is also known that 91% of data traffic is generated from smartphones<sup>[7]</sup>. This indicates that data traffic is the main cause of congestion in the wireless network, because almost all the non-voice applications that run over a smartphone are Internet data<sup>[8]</sup>. However, frequency resource in the wireless network is limited and the basic business model for the current wireless Internet is priced at a flat rate

\* 주저자 : Dept. of Information & Communication Engineering, Changwon National University, hoony@cwnu.ac.kr, Lifelong Member  
Paper number : KICS2012-01-001, Manuscript Received : 2012년 1월 2일, Final manuscript received : 2012년 6월 25일

(recently, unlimited usage is changed to limited usage plus extra charge) with best effort (BE) service.

To the best of author's knowledge, we could not find any discussion that mentions about a method to resolve the two problems (unfairness and performance degradation) from pricing scheme. This motivated our work on the proposition of a new pricing scheme that takes into account the attributes of the wireless network itself as well as the volume of the consumed traffic and supported QoS (quality of service) for the services.

The rest of this work is composed as follows: In Section II, we describe related works on the pricing for the WiMAX system. In Section III, we describe attributes and standard of the WiMAX system. In Section IV, we propose a new pricing scheme that reflects volume of the consumed network resource and support of differentiated QoS. In Section V, we illustrate validity of our work by showing result of the numerical experiments. Finally, in Section VI, we summarize this work.

## II. Related Works

There are lots of works concerning the pricing scheme for the WiMAX system<sup>[9-16]</sup>. There are two broad categories for the pricing of the WiMAX system: pricing from the economic and technical points of view.

First let us review works for the pricing based on the economic point of view. Riedel argued that, by using the concept of supply and demand model, when flat rate tariffs are applied to WiMAX system, consumer expands consumption of resource up to maximum in order to increase utility, which is harmful to both the ISP's revenue and QoS to other innocent users<sup>[9]</sup>.

Shu argued that flat pricing encourages waste of resources, increases cost of ISP, and forces light users to subsidize heavy users<sup>[10]</sup>. He argued that the fundamental problem of traffic flood in

the current Internet results from lack of penalties to heavy users, and he proposed that introduction of service differentiation and usage-based pricing is needed in the Internet. Shu proposed an auction-based pricing for the differentiated QoS service by levying high price to users who generate out-profile traffic. This makes sense to both users as well as ISP, because users can receive high-quality service by paying higher fee and ISP can get different prices from users, via which ISP can recover the cost. This is in line with the concept of Paris metro pricing for the DiffServ Internet<sup>[11]</sup>.

Now let us review works for the pricing scheme based on the technical point of view. Ognenoski proposed a utility-based pricing for a WiMAX network<sup>[12]</sup>, where utility of a customer for voice and file-download is defined using ITU-T's E-model and Kelly's nonlinear function, respectively<sup>[13]</sup>. However, this is inappropriate in WiMAX network, because it does not reflect the operational principle (resource allocation method and support of QoS) of WiMAX system.

Belghith et al. proposed pricing for the real-time and non-real-time services over a WiMAX system<sup>[14,15]</sup>. They determined the price of service by the unit of bandwidth that is reserved and used by each class of service (UGS, rtPS, and BE). Lee proposed a very similar pricing scheme in [16], proposing that UGS class is levied by deterministic bandwidth, and rt-PS/nrt-PS classes are levied by statistical bandwidth, so that price is levied by amount of allocated bandwidth.

However, both schemes have two common limits: First, only bandwidth is taken into account in charging. This causes an ambiguity in the cost of network service, because they are levied at the same rate. Second, they levy charge to a connection based on the allocated bandwidth rather than actually consumed bandwidth. This causes overestimation of price, because a WiMAX system operates in a full IP network rather than circuit-switched network, so that a connection does not fully consume the allocated bandwidth

throughout the connection holding time. In IP network, bandwidth that is not actually used by connections is used by other class of connections such as BE.

Summarizing the above discussion, we argue that the best policy for the price may be made if one takes into account both points of view (economic and technical). However, we argue that the basic principle for the pricing has to take into account the technical aspect of the WiMAX system, after that economic aspects such as an auction or marketing policy have to be added, otherwise the fundamental facts about the characteristics of the system *per se* is blurred, and this causes a misinformed bias about the real price of WiMAX service.

Therefore, let us review the basic principle for the operation of the WiMAX system, which is given in Section III.

### III. Wireless Internet over WiMAX

Let us describe the principle of the operation of Internet over the WiMAX system in two areas: the radio access network (RAN) and the IP backhaul network (BhN).

#### 3.1. Radio access network

Customers can exchange voice conversation, browse web pages, download audio/video files, join on-line game, and enjoy SNS (social networking service) by accessing to the WiMAX RAN composed of BS (base station) and RS (relay station), from where the connection is extended to wired IP BhN where servers are located.

First, note that IEEE 802.16 standard has three different versions, 802.16d, 802.16e, and 802.16j, where the former is called fixed WiMAX and the latter two are called mobile WiMAX.

This work focuses on the 802.16j network, which is an extension of 802.16e for enhancing throughput and coverage<sup>[17]</sup>.

In 802.16j WiMAX system, a number of SSs (subscriber stations) connect to a network directly

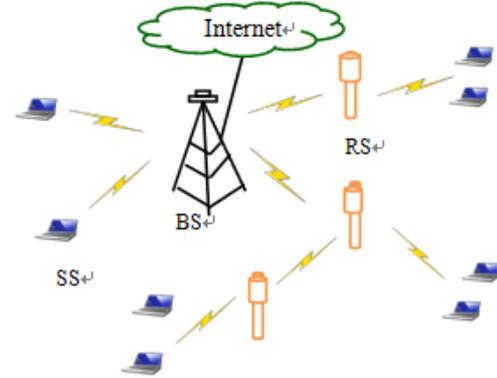


Fig.1. Architecture of MHR WiMAX network

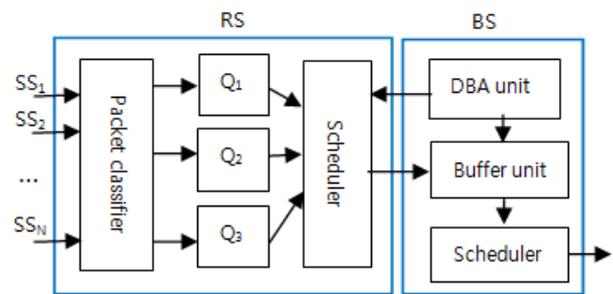


Fig. 2. Block diagram of the RS and BS

to a BS or indirectly to a BS via an RS. BS is a centralized server for the WiMAX system which multiplexes and demultiplexes traffic in a cell to and from the Internet, whereas RS is an agent acting as a mediator between BS and SSs, via which the range of WiMAX cell extends to large-scale. Thus, the WiMAX system operates in multi-hop relay (MHR) mode. Fig.1 illustrates the architecture of MHR WiMAX network.

Let us describe the operation of the WiMAX network. A number of users access to an SS (SS is an aggregation system of user's terminal) via wireless resource. A BS or RS accommodates multiple SSs, which is connected to an ISP via a switching center. Noting that the access point of the ordinary customers to the WiMAX network is an RS or BS, the former being dominant, let us focus on RS. The functional block diagram of the BS and RS in the WiMAX network is given in Fig.2, where the left and right box represents RS and BS, respectively<sup>[18]</sup>.

WiMAX is based on a connection-oriented MAC (media access control) protocol, so that resource is allocated on a connection basis, the identification of which is given by source and destination addresses.

The resource allocation is based on reservation or contention. The operation of BS and RS are as follows<sup>[18]</sup>: When end users need bandwidth for uploading data, each user reports bandwidth request to RS. RS sends its aggregated bandwidth requirement to BS, which is based on the queue status and QoS requirements. BS runs a DBA (dynamic bandwidth allocation) algorithm, via which it creates a bandwidth grant, which is given to RS. Upon receiving the grant from BS, RS distributes the bandwidth to the corresponding SSs.

Upon receiving packets from SSs, RS classifies them into a number of QoS classes, which are buffered in the corresponding queue (in Fig.2,  $Q_1$ ,  $Q_2$ , and  $Q_3$  is for high, medium, and low priority, respectively) in the RS. RS's final job is to send the packets to BS, which is done by a packet scheduler. The scheduling algorithm is proprietary of the vendors, operation of which is discussed in detail in Section IV. A series of above-mentioned processes such as packet classification, buffering, and scheduling are also done at the BS, and packets will be transferred to the Internet.

On the other hand, WiMAX assumes full IP network, where data is transferred in a unit of packet. The packet-oriented operation of network requires queue to adapt the mismatch between the rate between source traffic and channel service<sup>[19]</sup>. Computer networks have at least four elements of end-to-end delay: signal processing delay, queuing delay, propagation delay, and packet transfer delay. Among them, queuing delay is the most variable factor and it plays a key role in the degradation of total delay performance in the network. This mandates the queuing analysis in the modeling, design, and performance evaluation of WiMAX system such as RS and BS.

In fact there are buffers at every node in the end-to-end path of a WiMAX connection {SS →

RS (more than 1) → BS → gateway router → interior router (more than 1) → gateway router → server} for uplink and vice versa for downlink. This means that the queuing delay at each node has to be kept below the delay budget of each node, otherwise the end-to-end delay target of a flow is not guaranteed. One can find the delay budget of 3G network in [20], and the delay budget of the WiMAX system can be estimated in the same way.

In packet-switched wireless networks, guarantee of QoS to a packet is a challenging problem, because there are so many factors that determine the transfer rate of the signal such as interference between nodes, time-varying channel capacities, mobility of users, etc. Therefore, BS and RS have to prepare sufficient buffers and an appropriate packet scheduling scheme for each SS or each type of service class, which is discussed in detail later.

Now let us investigate the generic operation of WiMAX system, which is based on OFDMA (orthogonal frequency division multiple access). In OFDMA, a frame is composed of time and frequency axis. Subcarrier and symbol is the minimum unit of frequency and time resource, respectively. In the frequency domain, a bandwidth of a frame is divided into a number of subcarriers. Time is divided into a number of symbols. Combining the frequency-time domain, a slot is made, which is composed of one subcarrier and one, two or three symbols (e.g., 1 subcarrier × 3 symbols = 1 slot). A continuous group of slots is called a burst.

RS allocates a number of bursts to each user, amount of which is determined by several factors such as demand, channel conditions, and QoS requirements. Therefore, OFDMA is able to support different bandwidths to different SSs by allocating different number of burst to different users. However, OFDMA only does not guarantee end-to-end QoS such as delay and loss of packet, it only provides each SS with an appropriate bandwidth at each time of transfer opportunity.

There are two modes of operation for RS in the

WiMAX system: transparent mode and nontransparent mode, each being used for

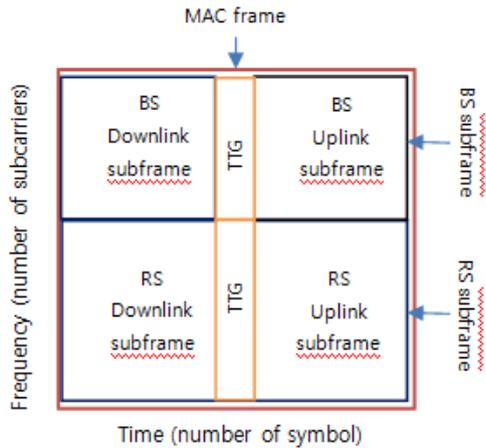


Fig. 3. Frame format of MHR WiMAX system

throughput enhancement and coverage extension, respectively. This work assumes a nontransparent mode in which the radio resource is shared by BS and RS. Because RS acts as an access node to users as well as a relay node to BS, the structure of a frame has to incorporate this fact. There exist three types of frame in RS: fragmented relay (FR), dynamically controlled relay (DCR), and out of band relay (OoBR). Among them, it is known that OoBR mode has the best performance in the throughput for a system with large bandwidth (say 20MHz)<sup>[21]</sup>.

Fig.3 illustrates a brief sketch of a format for a single frame of the WiMAX system with nontransparent and OoBR modes. The x-axis is time resource and y-axis is frequency resource.

First, frequency resource of a frame is divided into BS subframe and RS subframe, proportion of which depends on the topology and traffic of the system. A BS subframe is further divided into downlink and uplink subframe. The downlink subframe is subdivided into downlink access zone (data bound for SSs) and relay zone (data bound for RS), and uplink subframe is subdivided into uplink access zone and uplink relay zone, each of which acts as an access for SSs and relay of uplink traffic from RS to BS.

An RS subframe is also composed of downlink and uplink subframe. Downlink subframe is further subdivided into downlink access zone

(downlink from RS to SSs) and relay zone (receiving data from BS to RS), and uplink subframe is also subdivided into uplink access zone (transmitting data from SSs to RS) and uplink relay zone (transmitting data from RS to BS), the former is used for an access of data from SSs and the latter is used for relaying the data to BS.

The downlink MAP (DL-MAP) is located at the head of a frame, and it indicates the mobile device which frequency resource is allocated to it. The uplink MAP (UL-MAP) can be included as an option. FCH (frame control header) is used for frame control. The remaining field in a frame is allocated for the transmission of user's data. A detailed structure of the frame and subframe are given in<sup>[22]</sup>.

When it comes to resource allocation to SS, there exist two types: grant per connection (GPC) and grant per SS (GPSS). In GPC, SS handles bandwidth request on a connection basis, whereas in GPSS, SS aggregates bandwidth request for all of its connections. GPSS is preferred by ISPs due to simplicity, and it complies well with the philosophy of the DiffServ architecture of the IP network.

WiMAX system has been developed under the assumption of provision of the QoS to a user. WiMAX system defines four service categories: UGS (unsolicited grant service), RT-PS (real-time polling service), NRT-PS (non-real-time polling service), and BE (best effort).

UGS requires RS to allocate a fixed-size grant to a flow at the initiation of a connection. Therefore, a flow can be provided with fixed amount (peak rate) of bandwidth, which is reserved and maintained throughout the connection holding time. So contention for the bandwidth is not needed to this category. UGS is used for real-time streams such as VoIP without silence suppression.

RT-PS requires RS to allocate a minimum rate to a connection periodically (in a frame basis). SS has no right to use a contention REQ. RT-PS is used for real-time streams with variable data

rates such as MPEG video.

NRT-PS requires RS to allocate a minimum rate to a connection by regularly allowing a unicast REQ opportunity to a connection even during congestion. When congestion occurs, the required bandwidth may not be allocated. NRT-PS is used for delay-tolerant data such as web browsing. REQs for the RT-PS and NRT-PS categories are made on a per-flow basis, whereas grant of resource is issued on a per-SS basis.

Finally, RS does not guarantee bandwidth to SSs that belongs to a BE class, so that BE uses a residual bandwidth in the frame that is not used by other classes. BE is used for e-mail or ftp.

Summarizing the resource allocation of the WiMAX system we conclude that usage of radio resource has to be taken into account pricing, via which fairness between users can be secured such that heavy users pay more money compared to light users.

### 3.2. IP backhaul network

Now let us turn to the BhN of WiMAX network. In between the RAN and BhN lies a router (service router, say SR) which acts as a gateway between two different technologies. This corresponds to serving GPRS support node (SGSN) in 3G UMTS network<sup>[19]</sup>. Since SR serves multiple BSs, the SR is a hot spot for the traffic between RAN and BhN, so that it has to prevent overload of RAN as well as support of QoS to the users connected to RAN. To realize that, packet differentiation scheme has to be incorporated in SR.

Inside the BhN is IP cloud, which is composed of concatenation of routers equipped with DiffServ (differentiated service) function. Note that the operation of IP network is relatively well-known and it is not a major issue in this work, so let us briefly introduce the IP DiffServ technology. In DiffServ IP networks, packets are treated in a priority-based service inside the BhN. DiffServ defines three PHBs (per hop behaviors): expedited forwarding (EF), assured forwarding (AF), and BE. Packets of EF class receive absolute

guarantee of delay under any network condition. Packets of AF class receive minimum and statistical guarantee of delay, whereas no guarantee is given to BE class.

DiffServ has two features. First, service differentiation is given to a PHB basis rather than individual flows. Second, classification of packets is carried out at the edge of the network. At once a packet enters the IP network, no other process is needed other than distribution of packets to different PHBs. These two features give us an intuition about support of QoS and pricing for the wireless Internet that is transported by an IP backhaul network. First, pricing for the wireless Internet has to be based on the DiffServ such that packet of different QoS classes is charged differently, which complies with the former feature of DiffServ. Second, the charging function is located at the edge of the network, which complies with the latter feature of DiffServ.

If we summarize the characteristics about support of QoS in WiMAX RAN and IP BhN we can draw a conclusion that differentiated treatment of packets is needed for voice over data and it has to be reflected in the price, via which performance of voice traffic can be guaranteed even at congestion in the network. Finally, we found that support of QoS differentiation and corresponding pricing in the WiMAX network can resolve the problems of unfairness and performance degradation at the same time.

## IV. DiffServ-aware Pricing

Before proposing our new pricing scheme for the WiMAX service, let us review the current pricing scheme.

### 4.1. Review of current pricing scheme

First, let us summarize the basic type of pricing for the multimedia services. The first contender is voice call service provider, where charge is levied by minutes (In practice, it is charged by a unit of second). This idea originated

from the nature of circuit-switched wired and wireless telephone network, where a circuit is established between the end-to-end path of a voice connection. A voice connection generates a periodic stream of constant bits of data over a dedicated 64Kbps channel for a G.711 codec or 15.2Kbps channel for an iLBC (internet low bit rate codec)<sup>[23]</sup>. Therefore, the only factor that determines the charge of the voice call is the consumed seconds of each call. This is called a time pricing (TP).

When it comes to Internet services such as web browsing, streaming, and ftp, it does not make sense if we adopt TP because the amount of data is very different for different applications. For example, the traffic from streaming video is enormously larger than that from web browsing, even though they have the same connection holding time. This requires volume pricing (VP). VP levies charge based on the volume of data consumed by the connection. However, the pure VP is not widely spread to the commercial network. Instead, when ADSL (asymmetric digital subscriber line) appeared in the market, a flat pricing (FP) package with unlimited usage of data has become the norm of the business model. This was possible because ISPs have sufficient amount of bandwidth at the wired network.

The FP rule is also adopted by the wireless Internet. As the mobile network is evolved to 3G network, where voice and data service coexist, a new scheme for the charging was required. So born is the hybrid pricing (HP), which adopts a dual strategy for the charging: unlimited usage for data and limited usage for voice. A typical charging plan for 3G mobile service is shown in Table 1<sup>[24]</sup>.

As we can find from Table 1, there is no limit in the usage of data. It is known that this had caused the explosion of traffic in the wireless network. This became a major threat to the quality of the mobile voice service. In March 2011, a 4G mobile Internet service called WiBro (wireless broadband) has been launched in Korea<sup>[25]</sup>. The main reason for the wide-scale launching of a WiBro service is to cope

with the explosion of mobile data traffic over the 3G network that has been triggered by the smartphone.

Table 1. 3G service charge

Service plan	AIO54	AIO64	AIO79	AIO94
Monthly charge (Won)	54,000	64,000	79,000	94,000
Voice (minutes)	300	400	600	1,000
Text ( unit)	250	450	650	1,050
Data (bytes)	Unlimited			

Table 2. WiBro service charge

Service [Allowance/month]	Basic charge [Won/month]	Extra charge [Won/MB]
1GB	10,000	10
10GB	20,000	
30GB	30,000	
50GB	40,000	

The charge of the WiBro service (as of 4th quarter, 2011) [unit: Korean Won] is based on a multi-step flat pricing with additional charge for extra consumption, which is shown in Table 2<sup>[26]</sup>.

The current charging scheme is flat pricing if users consume data within allowance. As we can see from Table 2, the higher service plan a user subscribes, the cheaper the per-byte price. This encourages users to consume large amount of traffic at a small marginal cost of usage, which eventually causes network congestion.

Here, let us think about dilemma of current flat rate pricing. For example, what happens when an ISP levies 30,000Won per month for 30GB of data, whereas other ISP levies the same monthly fee for 40GB? Many users will switch to the latter ISP. Then, the former ISP has to increase the allowance or lower the charge, and the latter ISP has to change rate following the change of the former, and so on. This is just a chicken game, and it is known that this is one of the reasons for an unprofitable business model of the current wireless ISP.

Note that voice is not included in the current WiBro service, only data is supported, which means that current WiBro service does not

support full feature of QoS to WiMAX system. This may be the main reason for the adoption of flat pricing in the current WiBro service. However, this policy has a few limits. First, if WiBro service wants to have a competing power in the market, it has to support a full feature of multimedia service including voice, otherwise it has no competence over 4G LTE (long term evolution) service which supports voice. Second, the BE service model with data only does not resolve the problem of explosion of traffic when the number of customer increases. To overcome these limits, WiMAX system needs the provision of prioritized multimedia service and corresponding charging scheme.

As we have shown in [5], the current FP scheme encourages overconsumption of wireless resource. Here, we can find a paradox: ISPs worry about the explosion of mobile data, whereas they promote heavy consumption of scarce wireless bandwidth. If this situation continues, overload of wireless network is inevitable and user's perceived QoE (quality of experience) will also deteriorate.

#### 4.2. DiffServ-aware pricing

As we have described in Section III, a series of operations such as QoS class mapping and scheduling of packets have to be carried out by RS and BS as well as the nodes of IP backhaul network, which is described below.

##### 4.2.1. QoS mapping

If we map the QoS classes between WiMAX and IP backhaul networks, it can be given as shown in Table 3, and this mapping is applied throughout the remaining discussion. Note that RT-PS and NRT-PS is aggregated into a single PS class, because the type of resource allocation is similar.

##### 4.2.2. Scheduling algorithm

Table 3. QoS mapping between WiMAX and IP

WiMAX	IP	Class id	Application
UGS	EF	1	VoIP, video phone
PS	AS	2	streaming video real-time gaming web browsing, chat
BE	BE	3	ftp, e-mail

As we have presented in Section III, definition of QoS class and resource allocation in the WiMAX and IP networks are recommended by standardization bodies, whereas scheduling of packets depends on vendor's implementation. So, there exist various types of scheduling schemes in wired and wireless packet networks.

The typical scheduling schemes in a wired network are strict priority (SP), weighted round robin (WRR), deficit round robin (DRR), weighted deficit round robin (WDRR), and priority-based WDRR (PB-WDRR)<sup>[27]</sup>. Among them, it is shown in [27] that PB-WDRR, which is a mix of SP and WDRR, is the best solution in supporting differentiated QoS to different service classes as well as respecting the fairness among flows of the same class.

When it comes to wireless networks, performance is determined by diverse factors such as channel states, interference between users, locations of the mobile devices as well as capacity of radio. Typical methods for scheduling packets in wireless networks include opportunistic scheduling (OS) and fair scheduling (FS)<sup>[28]</sup>. OS maximizes the throughput of the network such that radio resource is allocated to users with best channel conditions. However, OS does not guarantee QoS as well as fairness between different connections. Only users of good SNR (signal to noise ratio) can achieve maximum throughput. FS tries to guarantee fairness to each user by giving a chance to transmit packets fairly to every user. Therefore, a real-time application such as voice can be fitted to this scheme. However, FS does not support QoS differentiation.

Rodrigues proposed a scheduling scheme for achieving the trade-off between resource efficiency

and user fairness in wireless networks using utility-based adaptive resource allocation scheme, where utility for real-time and nonreal-time traffic is delay and throughput, respectively<sup>[29]</sup>. Nevertheless, we could find no causality between scheduling and achieved QoS, which results from two inherent characteristics of the wireless network: opportunistic resource allocation and frame-based operation of OFDMA. OFDMA operates in a frame basis, so that different packets from different flows can achieve the same level of QoS (delay and throughput) at once they ride over the same frame.

Therefore, true differentiation of QoS in wireless network can be achieved by giving a chance of service to higher priority packets by a packet scheduling scheme at the output of RS. To do that, we decouple the operation of radio resource allocation (RSA) and packet scheduling (PS) at the entrance of radio network, where RAS allocates radio resource to a connection and PS provides differentiated QoS to different classes of packets. In this work we assume that RAS is achieved by the generic scheme of wireless system, and we only focus on the packet scheduling scheme.

In [30] we have shown that SP can be applied to high-speed packet switched network with two QoS levels by illustrating that the delay performance of high-priority packet is comparable to a single class queue whereas low-priority packet is not severely hurt throughout a wide spectrum of offered load. This was possible by the increased speed of link (e.g., in the range of 100Mbps) as well as sufficient capacity of memories in the network equipment such as routers and multiplexers. Therefore, we apply the SP scheduling scheme in this work by extending the number of QoS class into three, which operates as follows: At the input side of an RS, packets are classified into three classes: class 1, 2, and 3. Packets of the same class are stored at the same queue and served by FIFO scheme. Packets of different classes are served by SP scheduler, via which real-time applications do not

suffer from variable delay and jitter.

#### 4.2.3. Pricing scheme

Let us propose how price is levied to each class of service in the WiMAX system, which is divided into two phases: resource consumption and packet scheduling. Because the resource of wireless network is scarce, consumption of the radio resource has to be taken into account in pricing. As we have shown in the frame structure of OFDMA, the amount of radio resource consumed by a user at each frame is computed by counting the number of bursts. This is feasible because RS is aware of the amount of bursts that it had issued to an SS at every service cycle (frame).

When packets are classified into a QoS class at the RS, the RS, BS, and routers inside the BhN schedule packets by SP scheduler, based on which differentiated pricing is applied to each class.

##### (A) Phase 1: Usage-based pricing

RS computes the number of actually consumed burst for each SS at each frame. The number  $a_i$  of burst consumed by class 1 service at frame  $i$  is fixed to be  $a$ . The basic number  $\chi_i$  of burst for class 2 service at frame  $i$  is also fixed at every frame  $\chi$ , whereas extra number of burst ( $a_i$ ) consumed by a frame is variable at each service cycle. The volume of burst ( $\beta$ ) for class 3 service varies at every frame. Finally, the unit price of a burst is given by  $\kappa_j$ ,  $j=[1\sim 3]$ . Then we can compute the price of burst by collecting the number of burst consumed by each connection for each type of class.

##### (B) Phase 2: DiffServ-aware packet pricing

Because RAN has a hub and spoke topology, and the packets traverse a number of internal routers in the BhN, the performance metrics may have additivity or concavity. For example, let  $m(i,j)$  be a QoS metric for a link  $(i,j)$ . For a path  $P=(s,i,j,\dots,l,t)$ , where  $s$  is source and  $t$  is destination, a metric  $m(i,j)$  is concave if

$m(P)=\min\{m(s,i),m(i,j),\dots,m(l,t)\}$  and additive if  $m(P)=m(s,i)+m(i,j)+\dots+m(l,t)$ .

Note that delay is additive and bandwidth is concave. Therefore, a decision problem for the optimal packet price based on the delay QoS is NP-complete. To resolve this problem, let us introduce the concept of partitioned delay budget. Gopalan proposed a scheme for delay budget partitioning to maximize the efficiency of the usage of the network resource<sup>[31]</sup>. The delay budget at the RAN has to be kept strictly (usually below 5ms), because the end-to-end transfer delay of wireless network is more than twice that of pure wire-line network<sup>[19]</sup>. Therefore, the delay budget for the RS where network resource is mostly scarce is the main point of concern, and the remaining budget is relatively stable. From now on let us focus on RS.

Let us evaluate delay of packets in an RS by using M/G/1 queue, where link is modeled as a single server, the arrival process is Poisson (this is sufficient to model aggregated traffic at RS) with mean rate  $\lambda_i$  for class  $i$ ,  $i=[1\sim3]$ , packets, the service rate is generally distributed with mean  $\mu_i$  and variance  $\sigma_i^2$  for class  $i$ ,  $i=[1\sim3]$ , packets, and the buffer capacity is sufficiently large. Furthermore, let us assume that the server is non-preemptive. The mean offered load of the class  $i$  packet is  $\rho_i = \lambda_i / \mu_i$  and the total offered load  $\rho$  to the system is given by  $\rho = \sum_{i=1}^3 \rho_i$ , where  $\rho < 1$  for stability.

Because users of each class of service are generating packets independently from each other,  $\lambda_i$ ,  $i=[1\sim3]$ , is independent from each other. The mean waiting time  $W_i$  of class  $i$  ( $i=[1\sim3]$ ) packet is given in (1)<sup>[32]</sup>.

$$W_1 = \frac{\sum_{k=1}^3 \lambda_k \tau_k}{2(1-\rho_1)}, \tag{1}$$

$$W_2 = \frac{\sum_{k=1}^3 \lambda_k \tau_k}{2(1-\rho_1)(1-\rho_1-\rho_2)},$$

$$W_3 = \frac{\sum_{k=1}^3 \lambda_k \tau_k}{2(1-\rho_1-\rho_2)(1-\rho)}$$

where  $\tau_k = \sigma_k^2 + \frac{1}{\mu_k^2}$ , for  $[i=1\sim3]$ .

Note that there exist the following relationships between  $W_i$  and  $W_j$  ( $i,j=[1\sim3]$ ):  $W_3 > W_2 > W_1$ , where

$$W_2 = W_1 \frac{1}{(1-\rho_1-\rho_2)}, \tag{2}$$

$$W_3 = W_2 \frac{1-\rho_1}{(1-\rho)}$$

Therefore, the price of serving class 1 packet has to be higher than that of class 2 and 3 packet, which is incentive-compatible.

When a prioritized packet treatment is introduced into the price, charge is levied to each class such that the total revenue of the network service is maximized. Let us denote the unit price of transmission of a class  $i$  packet to be  $p_i$ ,  $i=[1\sim3]$ . Then, the price  $P_i$ ,  $i=[1\sim3]$ , per unit time for each type of packet is given as follows:

$$P_i = p_i \lambda_i, i = [1, 2, 3] \tag{3}$$

Note that, even though class 1 packet is treated with the highest priority, upon arrival to the buffer, class 1 packet has to wait until the end of service for the packets from other classes (classes 2 and 3) that is in progress as well as the packets of class 1 that have arrived before it, so that it is inevitable that class 1 packet has to wait in the buffer before its turn for the service comes. Therefore, let us assume that the server redeems a fee to class 1 packet if there is any waiting in the buffer, where redemption is proportional to the mean packet delay encountered by class 1 packet in the buffer.

On the other hand, no redemption is given to class 2 and 3 packets, because they are not delay-sensitive. As we have shown in our previous work in [30], the mean delay of class 2 packet is sufficiently small due to the increased speed of the link in the current wireless and wired network (It is given in [30] that when the link speed is in the range of 100Mbps, the delay for the class 2 packets is in the order of a few milliseconds even in the very high offered load).

When it comes to class 3 packets, no QoS demand is given, so that packets are not redeemed for any reason.

Let us define the redemption fee for class 1 packet as follows:  $\xi$  Won per backlogged packet in the buffer. From the Little's law, the amount of packet waiting in the class 1 queue is  $\lambda_1 W_1$ , and the redemption fee is equal to  $\xi \lambda_1 W_1$ .

Finally, we obtain the total revenue  $R_T$  by adding the prices for packets from every class and subtracting the redemption fee for class 1 packets, which is given in (4).

$$R_T = \sum_{i=1}^3 P_i - \xi \lambda_1 W_1 \tag{4}$$

The purpose of the network operator is to maximize  $R_T$ , which is stated as follows:

$$\begin{aligned} \text{Objective : } & \text{Max}[R_T = \sum_{i=1}^3 P_i - \xi \lambda_1 W_1] \\ \text{Condition : } & R_T \geq 0, \rho < 1. \end{aligned} \tag{5}$$

The necessary condition for the existence of maximum in eq.(5) with respect to  $\lambda_i, i=[1\sim3]$ , gives us the following result:

$$\frac{\partial}{\partial \lambda_i} R_T = 0, i = [1, 2, 3] \tag{6}$$

From (6) we obtain (7).

$$p_i = \xi \frac{\partial(\lambda_1 W_1)}{\partial \lambda_i}, i = [1, 2, 3] \tag{7}$$

Solving eq.(7), we obtain an explicit formula for  $p_i, i=[1\sim3]$ , which is given in (8).

$$\begin{aligned} p_1 &= \xi \frac{\partial}{\partial \lambda_1} \left\{ \frac{\lambda_1 \sum_{i=1}^3 \lambda_i \tau_i}{2(1-\rho_1)} \right\} = \xi \left\{ \frac{2\lambda_1 \tau_1 + \lambda_2 \tau_2}{2(1-\rho_1)} + \frac{\lambda_1^2 \tau_1 + \lambda_1 \lambda_2 \tau_2}{2\mu_1(1-\rho_1)^2} \right\}, \\ p_2 &= \xi \lambda_1 \frac{\partial}{\partial \lambda_2} \left\{ \frac{\sum_{i=1}^3 \lambda_i \tau_i}{2(1-\rho_1)} \right\} = \frac{\xi \lambda_1 \tau_2}{2(1-\rho_1)}, \\ p_3 &= \xi \lambda_1 \frac{\partial}{\partial \lambda_3} \left\{ \frac{\sum_{i=1}^3 \lambda_i \tau_i}{2(1-\rho_1)} \right\} = \frac{\xi \lambda_1 \tau_3}{2(1-\rho_1)} \end{aligned} \tag{8}$$

(C) Price of a burst

Now let us relate the mapping between the price of consumed burst and price of packet with differentiated scheduling. If the size of a class  $i, i=[1\sim3]$ , packet is given, we can compute the number of bursts  $B_i, i=[1\sim3]$ , that is required to pack the packet into a WiMAX frame.

Because we know the price  $p_i$  of class  $i$  packet, we can obtain the price of a burst, which is given in (9).

$$\kappa_i = \frac{p_i}{B_i}, i = [1, 2, 3] \tag{9}$$

Note that (9) is the price of a burst consumed for the service class  $i, i=[1\sim3]$ , from which we can compute the price of data in whatever unit we want to know such as per megabyte price, which is illustrated in Section V.

### V. Numerical Results

In order to compute the price of packet for each class let us assume the system and source traffic parameters, which are shown in Table 4 and 5, respectively. The system parameters are based on [33], whereas a part of the source traffic parameters are based on [30].

In order to estimate the total capacity (unit: Mbps) of a frame for RS, let us perform a brief calculation. The number of subcarriers in a 20MHz frequency band is 1,828, among which the number of subcarriers that are used for the transmission of data is assumed 1,684 after assigning a number of subcarriers for the control data (including guard signals). And, let us assume that the highest modulation order of 64 QAM (6bit per symbol) is used. Then we have a maximum theoretical bandwidth  $C$  for the WiMAX system, which is given as follows:  $C=6 \times 1,684 \text{bits}/108.8 \mu\text{s}=98,287,937$  bps.

Note that  $C$  is almost 100Mbps. However, it is usually known that the effective throughput of the system that is used for the transmission of data is

50% if we assume various overhead as well as the signal condition.

In this work we assume that the packet size of voice follows G.711 codec with fixed length 200bytes, packet size of streaming video and file sharing is variable that follows a normal distribution with the same mean and standard deviation of 500bytes and 125bytes, respectively.

In [34], it is known that about 5% of the traffic from the mobile service (at the 4th quarter,

Table 4. System parameters

Parameters	Value
Channel bandwidth	20MHz
Subcarrier spacing	10.94KHz
OFDMA symbol time	108.8ms
Number of symbols/frame	48
Frame duration	5ms
Burst size (subcarrier×symbol)	1×3
Bandwidth efficiency	50%
Bandwidth ratio of BS:RS	1:1
Bandwidth ratio of DL:UL for RS	1:1

Table 5. Source traffic parameters

Class id	Packet size [bytes]		Mixing ratio [%]	Application
	Mean	STD		
1	200	0	10	Voice
2	500	125	40	Streaming video
3	500	125	50	file sharing

2011) is generated by a voice call, 40% from streaming video such as YouTube, and 55% from web browsing and file sharing. Based on this fact, we assumed the mixing ratio (offered load) of the traffic for voice, streaming video, and file transfer as shown in Table 5.

In the numerical experiments, we have assumed that the total offered load is 0.8, among which each class occupies the load based on the mixing ratio. We also assume that the redemption fee is determined by the current WiBro service plan with the largest allowance of 50GB, from which we obtained that  $\xi=10.8$  Won per backlogged packet of class 1.

The per-packet price for each class is given in Table 6. Note that voice packets are levied by a

Table 6. Per-packet price

Class	Per-packet price [Won/packet]
1	$0.816 \times 10^{-3}$
2	$0.4 \times 10^{-3}$
3	$0.4 \times 10^{-3}$

Table 7. Per-burst price

Class	Per-burst price [Won/burst]
1	$48 \times 10^{-6}$
2	$9.5 \times 10^{-6}$
3	$9.5 \times 10^{-6}$

Table 8. Per-MB price

Class	Per-MB price [Won/MB]
1	4.08
2	0.8
3	0.8

relatively high price for each transmission of packet, which is the cost of receiving the highest priority service in the scheduler. On the other hand, the price of video and data packet is the same, which results from the incentive incompatibility of the packet.

If we map the per-packet price into per-burst price, we obtain the result shown in Table 7. Note that the per-burst price is the cost of the usage of a burst, the minimum allowance of frequency-time domain, which represents the price of unit resource in the network. However, it is dubious for the users to understand the physical meaning of it. So, let us finally translate the above-mentioned per-burst price into per-MB (megabyte) price, which is given in Table 8.

From Table 8, we can find that the per-MB price of voice is about five times that of data by providing high-quality service. This may be one of the reasons why mobile ISPs allow small amount of voice minutes as compared to unlimited allowance of data in the 3G network (See Table 1).

Therefore, if ISPs introduce a VoIP over the WiMAX system with QoS support, they can get higher revenue as compared to a pure data only service. Note also that, if ISP levies a charge for the usage of wireless Internet in the unit of MB, it is easy for the users to understand their usage

of the network resource from the charge levied in the unit of consumed data.

## VI. Conclusion

This work proposed a new pricing scheme for the wireless Internet over WiMAX network. We showed why introduction of service differentiation in the wireless access and IP backhaul networks is needed by summarizing the generic operation principle of WiMAX system. We also presented a chain of relationships between resource allocation, QoS differentiation, mapping of QoS between WiMAX system and IP network, and scheduling the packets in the RS, based on which we proposed a DiffServ-aware pricing scheme for wireless multimedia services. Finally we obtained an explicit formula for a per-packet price for each class. Via numerical experiments, we computed per-packet price, per-burst price, and per-MB price for the voice, data and video services, and illustrated the validity of the proposition.

Summarizing our work, the novelty of this work lies in two points. First, we built a theoretical framework for the resource consumption and QoS guarantee for the WiMAX system and pricing. Second, we presented explicit formulae for the price of the voice, video and data packets based upon the usage and supported QoS from the WiMAX system, which have never been proposed up to now.

The result in this work can be also applied to pricing multimedia service over the LTE network, because the physical and MAC layer functions as well as the capacity of frequency resource of WiMAX and LTE is almost the same except that the QoS class is defined slightly differently.

From this work, we conclude that ISPs can levy more charge to a voice service as compared to the conventional data service by providing a better quality of service over the same amount of radio resource, which gives us an intuition that they can get more revenues and users can receive better quality of services if ISPs introduce diverse service plans such as voice and video in the

WiMAX system.

In the future we will extend this work by collecting data on the usage of mobile Internet for different types of applications, via which we can find more concrete relationship between the pricing and performance/QoE of the wireless Internet.

## Acknowledgments

The author wishes to thank the anonymous reviewers for their constructive comments on this work. This work is supported by 2011-2012 overseas dispatch program of Changwon national university.

## References

- [1] H. Kim and I. Lee, "Unlimited usage of data service kills ordinary users", ChosunBiz.com, Vol.28082, April 13, 2011.
- [2] J. Yoo, "Investment cost has to be reflected in the telecommunication charge", <http://economy.hankooki.com>, April 27, 2011.
- [3] 'HetNet' a means to cope with traffic explosion, Nikkei Communications, April 2011.
- [4] H.J. Ryu, "If it were not for smart phone", Chosun Economy-I, December 1, 2010.
- [5] Hoon Lee, "The other side of flat pricing in wireless Internet", Proceeding of APCC, Jeju, Oct. 2012.
- [6] H.J. Kim, "Analysis on communication service", Daishin Investment forum, Dec. 16, 2010.
- [7] C. Lee, "Internet congestion in our neighborhood", Chosunilbo, no.28096, April 29, 2011.
- [8] S.M. Ha, "From July 2001 on, the speed of mobile Internet will be 7 times faster than current by the introduction of LTE", Chosunilbo, Jan. 13, 2011.

- [9] A. Riedel, T. Fischer, and D. Hogrefe, "Optimising WiMAX business models using incentive engineering", Proc. of ICON, 2006.
- [10] Jun Shu and Pravin Varaiya, "Pricing network services", Proc. of INFOCOM 2003.
- [11] Andrew Odlyzko, "Paris metro pricing: The minimalist differentiated services solution", <http://www.uop.edu.jo/download/PdfCourses/PM/paris.metro.minimal.pdf>.
- [12] O. Ognenoski et al., "User perception of QoS and Economics for a WiMAX network in a backup scenario", Wireless VITAE 2009.
- [13] F. P. Kelly, "Charging and rate control for elastic traffic", European journal on telecommunications, vol.8, no.1, 1997.
- [14] A. Belghith et al., "Pricing of real-time applications in WiMAX system", IEEE 68th Vehicular technology conference, VTC2008-Fall, Calgary, Canada, Sep.2008.
- [15] A. Belghith et al., "Pricing of differentiated-QoS services WiMAX networks", IEEE Globecom 2008, New Orleans, LA, USA, Nov. 2008.
- [16] Hoon Lee, "Charging and Revenue Estimation for the WiMAX System", Journal of KICS, Vol.34, No.3, March 2009.
- [17] Y. Xiao and Y. Pan, Emerging wireless LANs, wirelessPANs, and wireless MANs, Wiley, 2009.
- [18] R. Fei et al., "A utility-based dynamic bandwidth allocation algorithm with QoS guarantee for IEEE 802.16j-enabled vehicular networks", The 8th International conference on embedded computing, 2009.
- [19] D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of service", IEEE Tr. On Wireless Communications, vol.2, no.4, July 2003.
- [20] M.C. Chuah and Q. Zhang, Design and performance of 3G wireless networks and wireless LANs, Springer, 2006.
- [21] F. Gordejuela-Sanchez et al., "Efficient mobile WiMAX capacity estimation in a multihop environment", ICCS 2008.
- [22] D. Wong et al., Wireless broadband networks, Wiley2009.
- [23] Lin Cai, Xuemin Shen, and Jon W. Mark, Multimedia services in wireless Internet, Wiley, 2009.
- [24] <http://www.sktelecom.com>.
- [25] K. Ho, "Can KT's WiBro strategy be accepted by customers?" Chosun Economy, vol. 28047, March 3, 2011.
- [26] <http://www.wibro100.com>.
- [27] Miguel Barreiros and Peter Lundqvist, QoS-enablednetworks,Wiley,2011.
- [28] Stefania Sesia, Issam Toufik, and Matthew Baker edition, LTE-The UMTS Long Term Evolution, Wiley, 2009.
- [29] E.B. Rodrigues and F. Casadevall, "Control of the trade-off between resource efficiency and user fairness in wireless networks using utility-based adaptive resource allocation", IEEE Communications Magazine September 2011.
- [30] Hoon Lee, "Anatomy of delay performance for the strict priority scheduling scheme in multi-service Internet", Computer Communications, Vol.29, Issue 1, December 1, 2005.
- [31] K. Gopalan, T.-C. Chiueh, Y.-J. Lin, "Delay budget partitioning to maximize network resource usage efficiency", INFOCOM 2004, March 7-11, Hongkong.
- [32] Hoon Lee, Fundamentals on Queuing Systems, Bumhan Publishing, 2010.
- [33] P. Kolomitro, A.-E. M. Taha, and H. Hassanein, "A performance comparison of frame structures in WiMAX relay networks", Proc. P2MNET 2010, Denver, Colorado, USA.

- [34] K. Park and J. Choi, "Status, problems, and prospects of mobile service", Information and communication, KICS, December 2011.

이 훈 (Hoon Lee)



Feb.1984 B. E. from  
Kyungpook Nat'l Univ.

Feb.1986 M. E. from  
Kyungpook Nat'l Univ.

March 1996 Ph.D. from  
Tohoku Univ., Japan.

Feb.1986~Feb. 2001: KT

R&D Center

March 2001~Changwon Nat'l Univ.

<Research fields> Design and performance  
evaluation of computer network, Traffic  
engineering, QoS and charging