PAPER *Special Section on Communication Quality*

Design of the Cross-Layer QoS Framework for the IEEE 802.16 PMP Networks*

Yi-Ting MAI[†a)], ***Nonmember***, **Chun-Chuan YANG**[†], ***Member***, ***and*** **Yu-Hsuan LIN**[†], ***Nonmember***

SUMMARY As one of the promising techniques in Broadband Wireless Access (BWA), IEEE 802.16 also namely WiMax provides wide-area, high-speed, and non-line-of-sight wireless transmission to support multimedia services. Four service types are defined in the specification of IEEE 802.16 for QoS support. In order to achieve end-to-end multimedia services, 802.16 QoS must be well integrated with IP QoS. In this paper, we propose a framework of cross-layer QoS support in the IEEE 802.16 network. Two novel mechanisms are proposed in the framework for performance improvement: Fragment Control and Remapping. Fragment Control handles the data frames that belong to the same IP datagram in an atomic manner to reduce useless transmission. Remapping is concerned with the mapping rules from IP QoS to 802.16 QoS and is designed to reduce the impact of traffic burstiness on buffer management. Simulation study has shown that the proposed scheme has higher goodput and throughput, and lower delay than the contrast.

key words: *802.16, WiMax, PMP, QoS, cross-layer*

1. Introduction

Broadband Wireless Access (*BWA*) technology provides an easy, time-saving, and low-cost method for deployment of the next generation (beyond 3G) network infrastructure. Since 1998, IEEE 802.16 working group has launched a standardization process called *Wireless Metropolitan Area Network* (*Wireless MAN^{TM}*) for BWA. The most updated specification of 802.16 (IEEE Std 802.16-2004) [1] focuses on fixed location wireless access and supports up to 134 Mbps data rate. Moreover, the standardization of a new 802.16 interface, 802.16e [2], supports wireless access with high mobility, has also been completed recently. The WiMax Forum (Worldwide Interoperability for Microwave Access) [3], [4] a wireless industry consortium with about 100 members including major vendors such as AT&T, Fujitsu, Intel, and Siemens Mobile, is supporting 802.16 technology and promoting its commercial use, which means 802.16 is becoming the most important technology in BWA.

As shown in Fig. 1, the *PMP* (*Point to Multipoint*) configuration of IEEE 802.16 network consists of a *base station* (*BS*) and a couple of *subscriber stations* (*SSs*) that connect to the BS via high-speed wireless link. The BS acts as a gateway to the Internet. Legacy LANs or even more complex subnet systems can connect to the IEEE 802.16 network via SS. An IEEE 802.16 network (including the Legacy LANs that connect to the SS) can cover a large geographical area since the distance between the BS and the SS can be up to 30 miles [1].

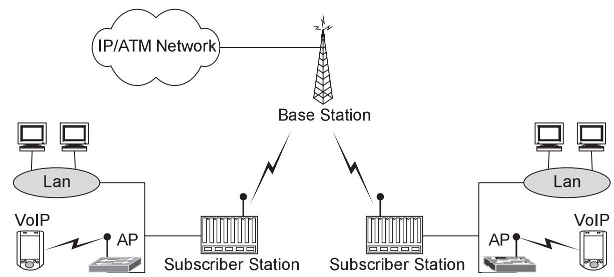


Fig. 1 IEEE 802.16 PMP mode.

IEEE 802.16 was designed to support multimedia service via QoS support of different service types. Mechanisms of QoS support such as admission control and bandwidth allocation in IEEE 802.16 were extensively researched in the literature. Based on the connection-oriented concept, the admission control scheme [5]–[8] must be properly designed to decide whether a new request of traffic flow can be granted or not. The new request is granted only when the bandwidth requirement of the request can be satisfied and none of the quality of the existing traffic flows is violated. On the other hand, some works [9]–[12] in designing an efficient scheduling mechanism for bandwidth allocation of IEEE 802.16 were also researched and proposed. The common idea of these scheduling mechanisms is to dynamically allocate time slots according to the service type of the traffic flows and to achieve higher network utilization. Well-known scheduling algorithms such as *Earliest Deadline First* (*EDF*), *Weighted Fair Queueing* (*WFQ*), *Round Robin* (*RR*), etc. were adopted in the literature. To integrate IP layer scheduling (Layer 3) and IEEE 802.16 scheduling (Layer 2), Chen et al. [13], [14] proposed the idea of multi-layer QoS scheduling support by assigning different scheduling algorithms in Layer 3 and Layer 2 for different combinations of L3 and L2 service types. Cicconetti et al. [15] even proposed the idea of considering the data type (such as web traffic, audio, video, etc.) of the application to assign an IEEE 802.16 service type and adopt an appropriate scheduling mechanism.

Manuscript received August 31, 2007.
Manuscript revised December 18, 2007.
[†]The authors are with the Department of Computer Science and Information Engineering, National Chi Nan University, Puli, Taiwan 545, R.O.C.
*This work was supported in part by the National Science Council, Taiwan, R.O.C., under grant NSC95-2219-E-260-004. A preliminary version of the research in the paper has been published in proceedings the 9th International Conference on Advanced Communication Technology (ICACT), 2007.
a) E-mail: s3321906@ncnu.edu.tw
DOI: 10.1093/ietcom/e91–b.5.1360

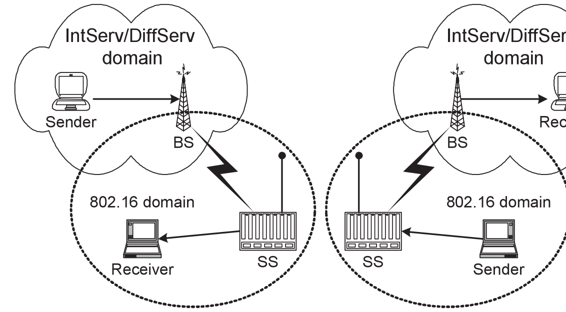
QoS issue for different transmission systems such as Ethernet, WLANs, and 3G cellular, has been addressed in the literature. For Ethernet QoS, Perez et al. [16], [17] proposed the idea of multi-layer infrastructure that classifies and prioritizes the voice and video traffic in order to improve the performance of collaborative system applications. Wang et al. [18], [19] proposed a cross-layer adaptive scheme to improve IEEE 802.11e QoS by concurrently adapting each host's MAC-layer parameters based on their application-layer QoS requirements and physical-layer channel conditions. Some researches [20], [21] proposed the idea of multi-layer QoS scheduling support by assigning different scheduling algorithms in the physical-layer for different upper layer service types in 3G cellular networks. Anastácio et al. [22] presented a description of characterization and classification parameters for beyond 3G mobile systems in the context of cross-layer design. Since the transmission characteristic differs for different transmission systems, QoS design in Ethernet, WLANs, and 3G cannot fit well in IEEE 802.16.

Existing research works as mentioned above presented the requirement of integrating different layers of QoS supporting in IEEE 802.16 or other structures. Their main focus is on multi-layer QoS mapping, and none of them proposed a complete framework of integration. In this paper, a cross-layer QoS framework and associated mechanisms for IEEE 802.16 PMP networks are proposed. In addition to the basic QoS mechanisms such as mapping from IP service types to IEEE 802.16 service types, the admission control for QoS flows, and the scheduling scheme as in other research works, two novel mechanisms namely Fragment Control and Remapping are proposed in the framework to improve network throughput. Fragment Control handles the data frames that belong to the same IP datagram in an atomic manner to reduce useless transmission. Remapping is concerning about dynamically adjusting the mapping rules from IP QoS to 802.16 QoS and is designed to reduce the impact of traffic burstiness on buffer management.

The remainder of the paper is organized as follows. First of all, we present the overall architecture as well as the novel features of the proposed QoS framework in Sect. 2. Key mechanisms in the proposed framework for QoS support in IEEE 802.16 network are presented in Sect. 3. Simulation study for performance evaluation and comparisons is presented in Sect. 4. Finally, Sect. 5 concludes this paper.

2. Cross-Layer QoS Framework

Although the 802.16 standard only defined up to L2 specification for the BS and SS, the proposed framework requires the BS and SS to be equipped with some of the L3 functionalities, such as IP header processing and L3 service class interpretation, for better service support. Since the traffic flows in the 802.16 network are classified as downlink or uplink, we present the framework in the downlink mode and the uplink mode respectively in the following:



(a) Downlink mode (b) Uplink mode

Fig. 2 Downlink and uplink diagram.

2.1 Downlink Mode

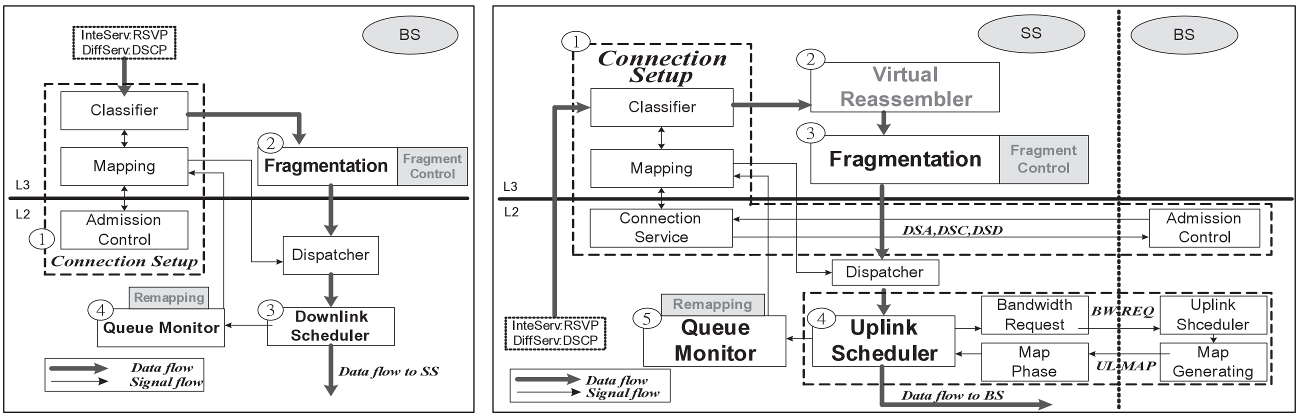
In the downlink mode, we assume the sender is located outside the 802.16 network and the receiver is located within the 802.16 network as displayed in Fig. 2(a). The framework in the downlink mode is illustrated in Fig. 3(a). Main functional blocks in the proposed QoS framework are briefly explained as follows:

(1) Connection Setup: Since there are mainly two categories of Quality-of-Service framework in L3 (IP layer), *Integrated Service* (*IntServ*) [23], [24] and *Differentiated Service* (*DiffServ*) [25], the functional blocks of Classifier and QoS Mapping from L3 to L2 are required at the BS for resource management in L2 admission control. In other words, the BS must be equipped with some of IP layer functionality, such as interpretation of IP header, to have a better support of QoS.

(2) Fragmentation: The size of an IP datagram can be up to 64 kbytes, but the size of each slot (*Maximum Transmission Unit, MTU*) in IEEE 802.16, although depending on the channel rate and the time frame length, is much smaller than the size of IP packets. Thus, fragmentation is a required function at the BS. Moreover, considering that all fragments coming from the same IP datagram must be successfully delivered to the destination for reassembly, these fragments should be treated as a whole in the 802.16 network. Therefore, the mechanism of Fragment Control is proposed in the framework, which maintains the dependency of the fragments from the same IP packet during L2 operations.

(3) Downlink Scheduler: The scheduler at the BS is responsible for dispatching IEEE 802.16 data frames of different service types at proper times (time slots). Since there are four service types, namely *UGS* (*Unsolicited Grant Service*), *rtPS* (*real-time Polling Service*), *nrtPS* (*non-real-time Polling Service*), *BE* (*Best-Effort*), defined in IEEE 802.16 [1], four queues are required in the scheduler.

(4) Queue Monitor: Queue Monitor is used for monitoring the state of each queue in the scheduler and cooperates with the Mapping functional block for better resource management. More specifically, Queue Monitor



(a) Downlink framework (b) Uplink framework

Fig. 3 Cross-layer QoS framework.

can change the mapping rule from L3 service type to L2 service type under certain situations to increase the utilization of the queues in the scheduler. The mechanism of changing the mapping rule is called Remapping in the paper.

2.2 Uplink Mode

We assume the sender is connected to the SS in the uplink mode in Fig. 2(b). As illustrated in Fig. 3(b), the operation of the framework in the uplink mode is more complicated than the downlink mode, since the SS must negotiate with the resource manager BS. Major differences of the uplink mode from the downlink mode are explained in the following:

(1) Cooperation of SS and BS: Although BS is the administrator in IEEE 802.16 network, the SS shall negotiate with the BS in the phase of connection setup and uplink scheduler. For example, message *DSA* (*Dynamic Service Addition*), *DSC* (*Dynamic Service Change*), *DSD* (*Dynamic Service Deletion*) are used in the admission control. Moreover, the SS must send out *BW_REQ* (*Bandwidth Request*) messages to the BS for resource allocation and channel access.

(2) Virtual Reassembler: Since an IP packet received at the SS comes from a subnet system (e.g. a legacy LAN) in which the source host locates, the IP packet is probably merely one of the fragments of its original datagram. To mark the Fragment Control mechanism more effective, a virtual reassembler is added before fragmentation. The virtual reassembler is used of identifying the IP fragments that belong to the same original datagram by virtually reassembling the fragments.

3. Cross-Layer QoS Mechanisms

3.1 QoS Mapping from L3 to L2

There are mainly two QoS frameworks in IP layer: Integrated Service (IntServ) and Differentiated Service (DiffServ), each of them defines different classes of QoS. We adopted a simple mapping rule from IP layer QoS to 802.16 QoS types [13] in our proposed framework as illustrated in Fig. 4.

	IP QoS →	802.16 QoS
IntServ	*Guarantee Service (GS)*	*Unsolicited Grant Service (UGS)*
IntServ	*Controlled Load (CL)*	*Real-time Polling Service (rtPS)*
DiffServ	*Expedited Forwarding (EF)*	*Real-time Polling Service (rtPS)*
DiffServ	*Assured Forwarding (AF)*	*Non- real-time Polling Service (nrtPS)*
IntServ, DiffServ	*Best Effort (BE)*	*Best Effort (BE)*

Fig. 4 Mapping rule from IP QoS to 802.16 QoS.

3.2 Admission Control

We adopt a simple rate-based admission control scheme, in which the new QoS flow must provide the required bandwidth and the BS check if there is enough capacity for the new flow. The algorithm of the admission control is displayed in Fig. 5. For example, a new UGS flow with bandwidth requirement b_{UGS} is accepted when the remaining capacity (i.e. the total capacity of the link B — the current load b_C) is larger than b_{UGS}. Moreover, since the characteristic of the flow in each service type varies, the required bandwidth defined for each service type should be different. More specifically, the peak rate for an UGS flow, the average rate for an rtPS flow, and the minimum rate of an nrtPS flow are used in the admission control respectively.

3.3 Fragment Control

As mentioned in Sect. 2.1, since fragmentation is always necessary for an IP packet to be transmitted via the 802.16 link, the objective of proposed Fragment Control is to provide a grouping mechanism so that the fragments of the same IP packet are treated as a whole during L2 processing. We assume that all fragments of the same IP packet are

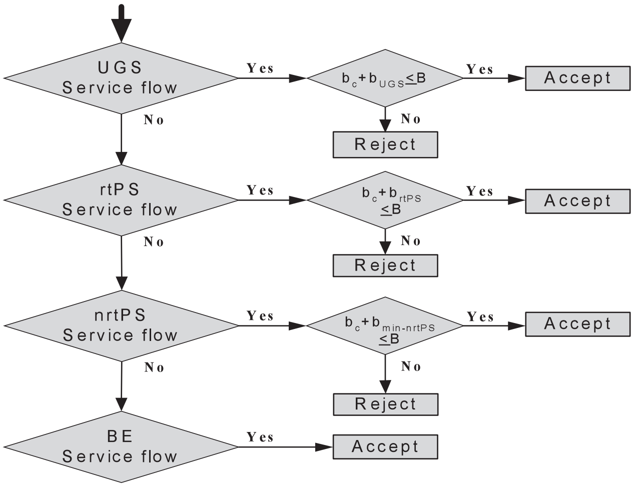


Fig. 5 Admission control rule.

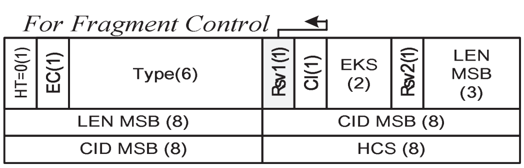


Fig. 6 802.16 MAC header format.

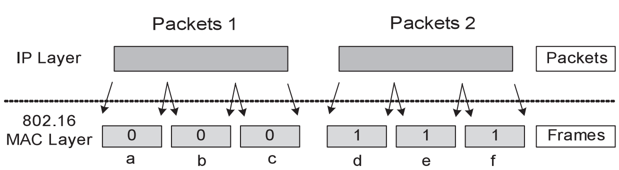


Fig. 7 Marking L2 frames according to L3 packets.

put into the L2 buffer in an atomic manner such that fragments coming different IP packets are not interleaved in the buffer. Therefore, one bit of a flag field in the header of the 802.16 MAC frame is enough for grouping the fragments. The reserved bit (*Rsv 1*) in the header of the 802.16 MAC frame (as shown in Fig. 6) is used for fragment grouping.

The fragments coming from the same IP packet are marked with the same value ('0' or '1') alternately in the flag field of the MAC frames and put into the 802.16 queue as illustrated in Fig. 7. L2 buffer operations are designed to treat the fragments with the same marking as a group. Therefore, in the case of congestion, the fragments of the same group should be removed all together for saving unnecessary frame transmissions in the congestion control mechanism such as *Drop Tail* (*DT*) or *Random Early Detection* (*RED*). The DT scheme drops data at end of the buffer when buffer overflows. The RED scheme adopts early detection of congestion by using two threshold values min_{th} and max_{th} on buffer utilization. Drop probability of the data frames located in between min_{th} and max_{th} is increasing linearly from zero to the maximum probability max_p as illustrated in Fig. 8.

Given that we assume the sender is connected to the SS

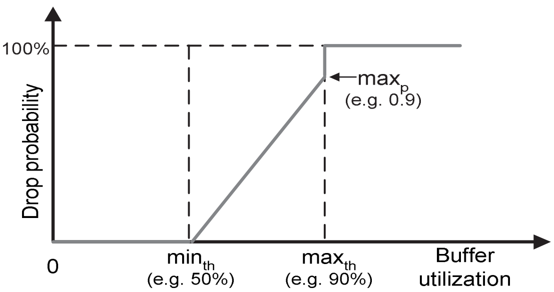
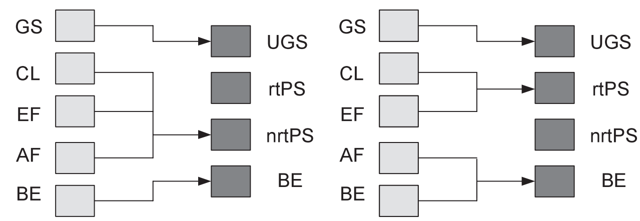


Fig. 8 The buffer state and drop rate in RED rule.



(a) New rtPS buffer mapping rule (b) New nrtPS buffer mapping rule

Fig. 9 The new buffer mapping rule of rtPS and nrtPS.

directly in the uplink mode, all fragments of an original IP datagram sent by the sender are expected to arrive at the SS within a short time. Therefore, the Virtual Reassembler is designed at the SS for grouping all fragments of the same IP datagram and marking the value of Rsv bit in the corresponding data frames with the same value of '0' or '1.'

3.4 Remapping

The proposed Remapping scheme is concerning with integrated buffer management of rtPS and nrtPS queues to achieve better buffer utilization and reduce frame dropping. Since the framework adopts static mapping rules from L3 QoS classes to 802.16 service types, there are cases that the rtPS queue overflows due to bursty traffic condition while the nrtPS queue still can accept more data frames. To better utilize buffers in the queues, a remapping rule is designed for L3 higher priority CL and EF packets to use nrtPS buffers when the rtPS queue is going to be full. Note that our proposed Remapping rule only focuses on VBR traffic sources of rtPS and nrtPS, since (1) UGS is for CBR traffic that requires dedicated resource, and (2) BE is the lowest priority service that does not require much attention.

To support Remapping, buffer utilization of rtPS and nrtPS queues must be continuously monitored. Moreover, two threshold parameters, *Upper-Bound* and *Lower-Bound*, are defined for the queues. Rules in the Remapping scheme are explained as follows:

(1) When buffer utilization of the rtPS queue exceeds its Upper-Bound, the queue monitor notifies the Mapping module in the framework triggering new remapping rules that map CL, EF, and AF packets to nrtPS as illustrated in Fig. 9(a).
(2) In the case of remapping being operated, if buffer uti-

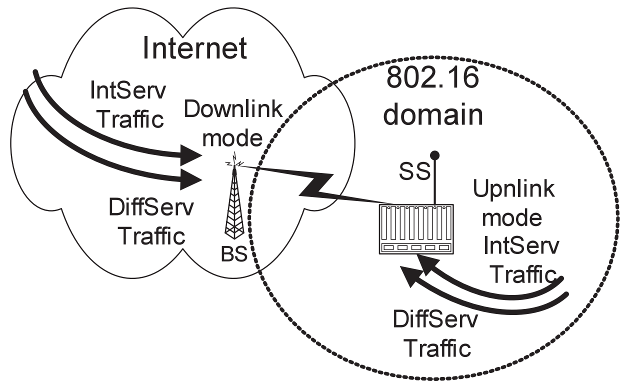


Fig. 10 Cross-layer QoS architecture of simulation.

lization of rtPS queue is lower than Lower-Bound, the mapping rules are restored back to the original ones as shown in Fig. 4.

(3) In the case of the nrtPS queue exceeds its Upper-Bound (i.e. the nrtPS is going to be full soon), the original mapping rules are restored only when buffer utilization of rtPS queue is lower than the middle line of Upper-Bound and Lower-Bound to reduce oscillations of rule application.

(4) Similarly, in nrtPS case, when its buffer is getting saturated, the mapping rule is revised so that L3 AF packets can use the L2 buffer of BE to reduce packet loss. The new mapping is shown in Fig. 9(b), and the Remapping scheme likes the rtPS state.

4. Performance Evaluation

4.1 Simulation Environment

The network environment used in the simulation is displayed in Fig. 10. Six types of traffic flow each for different IP class (*IntServ-GS, IntServ-CL, IntServ-BE, DiffServ-EF, DiffServ-AF, DiffServ-BE*) are generated in the simulated network. The inter-arrival time and the duration of each traffic flow are exponentially distributed. We varied the size of IP packets from 704 bytes to 1728 bytes in the simulation. Detail parameters of the simulation for Fragment Control are shown in Table 1, and parameters for Remapping are displayed in Table 2. The difference between Table 1 and Table 2 is that the total rate of the QoS flows ($\rho_{\rm QoS}$) can be up to 100% of the network capacity to emphasize the performance impact of Fragment Control. Moreover, we will also compare the performance of Fragment Control with two different congestion control schemes namely DT (Drop Tail) and RED (Random Early Detection) in the simulation. Note that the Remapping scheme is designed for congestion avoidance, and the parameter *Upper_Bound* is used as the congestion threshold in Table 2. According to the queuing theory as well as network practice, utilization above 80% is typically treated as an indication for near-congested situation. Thus, the value of *Upper_Bound* is set

Table 1 Simulation parameters for fragment control.

Input	*Service type*	*Mean*	*Variation*
UGS=3Mbps rtPS=4Mbps nrtPS=3Mbps BE=2Mbps ($\rho_{\rm QoS}$ = 1)	GS (UGS)	3072 Kbps	
	CL (rtPS)	2048 Kbps	512 Kbps
	BE (BE)	1024 Kbps	256 Kbps
	EF (rtPS)	2048 Kbps	512 Kbps
	AF (nrtPS)	3072 Kbps	768 Kbps
	BE (BE)	1024 Kbps	256 Kbps
Total Bandwidth		10 Mbps	
QoS Traffic Load ($\rho_{\rm QoS}$)		0.65~1	
MAC Frame Size		5 ms, 100 slots	
Simulation Time		500 sec	
L2 Buffer Size (each service type)		50 Kb = 100 slots	
IP Packet Size (byte)		704, 960, 1216, 1472, 1728	
Original Datagram Size (byte)		Mean 3072, Variation 50%	
RED Max, Min Threshold		90%, 50%	
Processing Delay		1ms	
Propagation Delay		1ms	
Spectrum		5.0 GHz	
Bandwidth		20 MHz	
Modulation		16-QAM	

Table 2 Simulation parameters for remapping.

Input	*Service type*	*Mean*	*Variation*
UGS=3Mbps rtPS=3Mbps nrtPS=2Mbps BE=2Mbps ($\rho_{\rm all}$ = 1)	GS (UGS)	3072 Kbps	
	CL (rtPS)	1024 Kbps	256 Kbps
	BE (BE)	1024 Kbps	256 Kbps
	EF (rtPS)	2048 Kbps	512 Kbps
	AF (nrtPS)	2048 Kbps	512 Kbps
	BE (BE)	1024 Kbps	256 Kbps
Total Bandwidth		10 Mbps	
Traffic Load ($\rho_{\rm all}$)		0.65~1	
MAC Frame Size		5 ms, 100 slots	
Simulation Time		500 sec	
L2 Buffer Size (each service type)		50 Kb = 100 slots	
IP Packet Size (byte)		704, 960, 1216, 1472, 1728	
Original Datagram Size (byte)		Mean 3072, Variation 50%	
Upper-Bound		80%	
Lower-Bound		40%	
Spectrum		5.0 GHz	
Bandwidth		20 MHz	
Modulation		16-QAM	

as 80%. Furthermore, as mentioned in Sect. 3.4, another parameter *Lower_Bound* is used to get rid of the oscillation in applying Remapping rules. We simply set the value of *Lower_Bound* as the half of *Upper_Bound* (i.e. 40%) in the simulation.

4.2 Simulation Results

4.2.1 Fragment Control

The average goodput of the case with Fragment Control (FC) and the case without FC under packet size 1728 bytes in the downlink mode is displayed in Fig. 11. As mentioned above, either DT or RED is used for congestion control in the simulation. Figure 12 displays the average goodput in the uplink mode. The two figures demonstrate that the FC scheme combined with either DT or RED can achieve higher goodput rates in all service types. The figures also show that as the total load of QoS flows goes up, the goodput gain of FC also increases. The reason behind the poor goodput per-

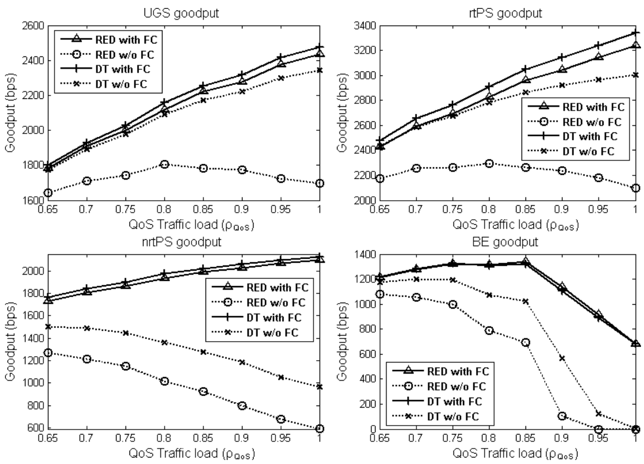


Fig. 11 Goodput with FC vs. w/o FC in downlink (packet size = 1728 bytes).

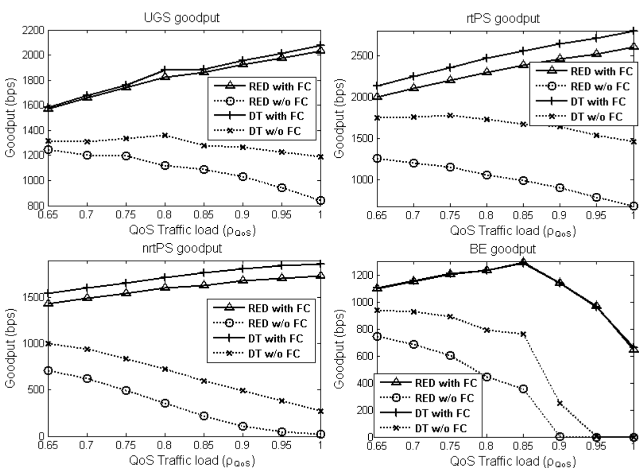


Fig. 12 Goodput with FC vs. w/o FC in uplink (packet size = 1728 bytes).

formance of the DT scheme without the aid of Fragment Control is twofold: (1) Some scheduling buffers in 802.16 are wasted in buffering those fragments that cannot be successfully reassembled due to part of the original IP packet is dropped. (2) Those fragments that cannot be successfully reassembled also compete the network capacity with other fragments. In the case of RED congestion control, the goodput is even worse than that of DT since random dropping in RED does not consider the coherence of fragments and creates more fragments that cannot be successfully reassembled.

The average delay of the case with FC and the case without FC under packet size 1728 bytes is displayed in Fig. 13 and Fig. 14. As in the case of goodput, the figures demonstrate proposed FC is helpful for delay improvement either with DT or RED congestion control. It is worth mentioning that the buffering delay of Virtual Reassembler in the uplink mode does not affect too much the delay advantage of FC over the counterparts as demonstrated in Fig. 14.

For more investigation of the performance improvement of the proposed FC scheme, Figs. 15–18 display the performance gain of FC under different packet sizes and QoS loads in terms of the average goodput and the average delay. In the case the uplink mode (Fig. 16), the goodput

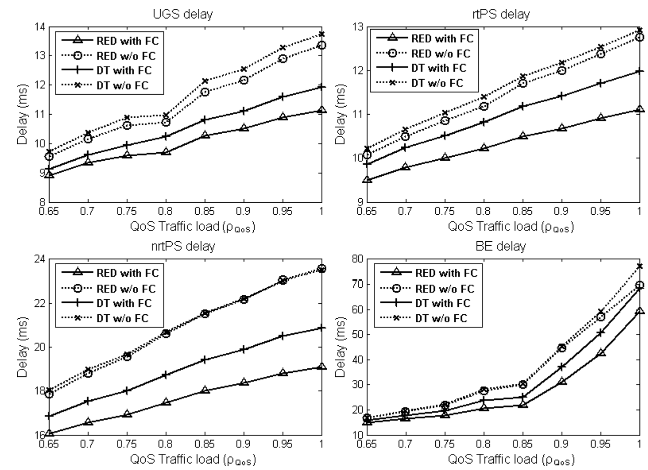


Fig. 13 Delay with FC vs. w/o FC in downlink (packet size = 1728 bytes).

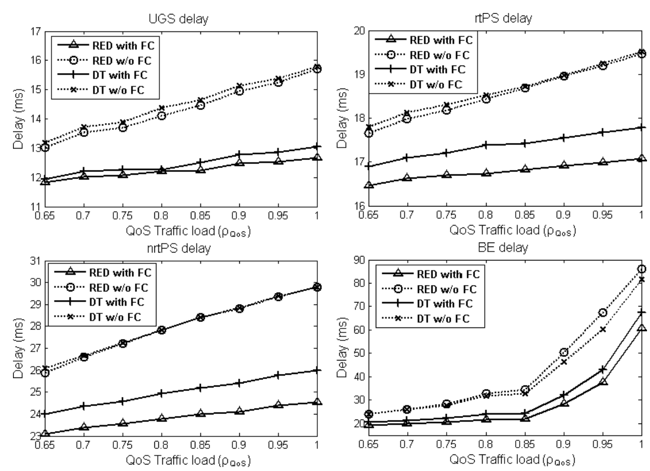


Fig. 14 Delay with FC vs. w/o FC in uplink (packet size = 1728 bytes).

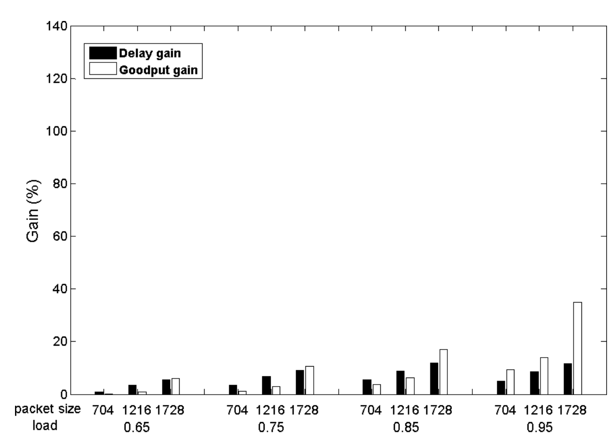


Fig. 15 Gain of DT with FC over DT w/o FC in downlink.

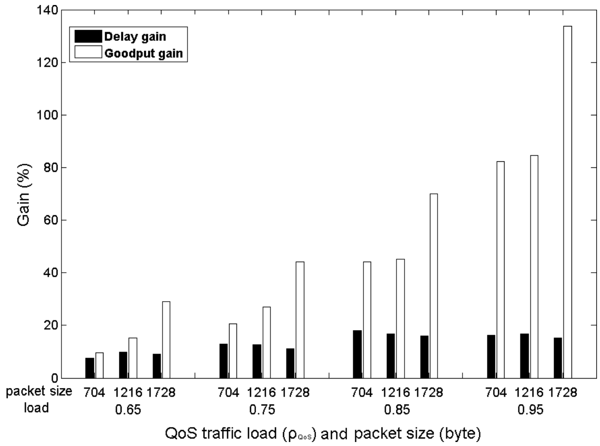


Fig. 16 Gain of DT with FC over DT w/o FC in uplink.

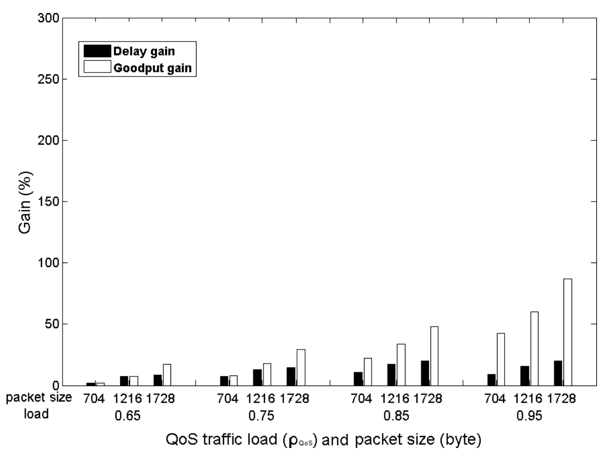


Fig. 17 Gain of RED with FC over RED w/o FC in downlink.

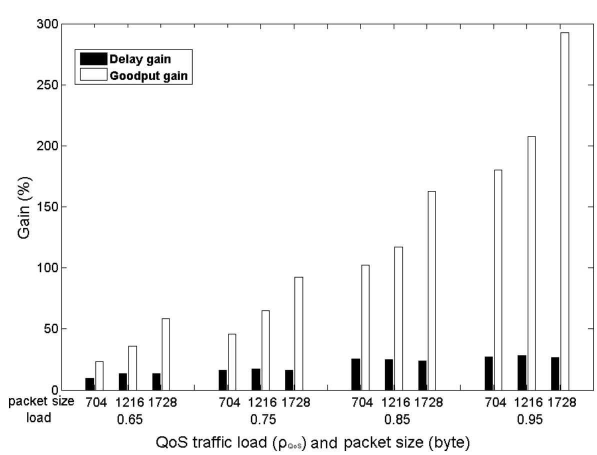


Fig. 18 Gain of RED with FC over RED w/o FC in uplink.

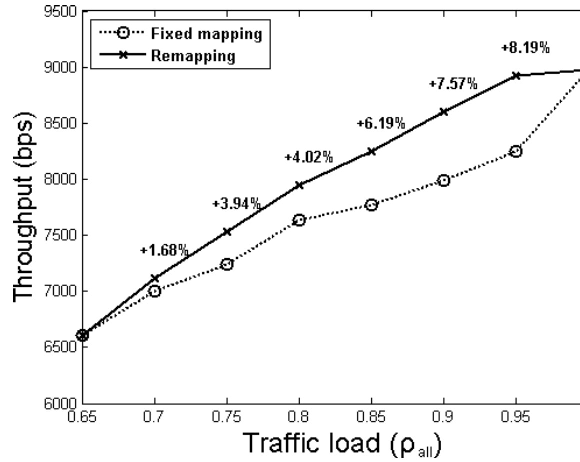


Fig. 19 Throughput: Remapping vs. Fixed mapping (downlink packet size = 1728 bytes).

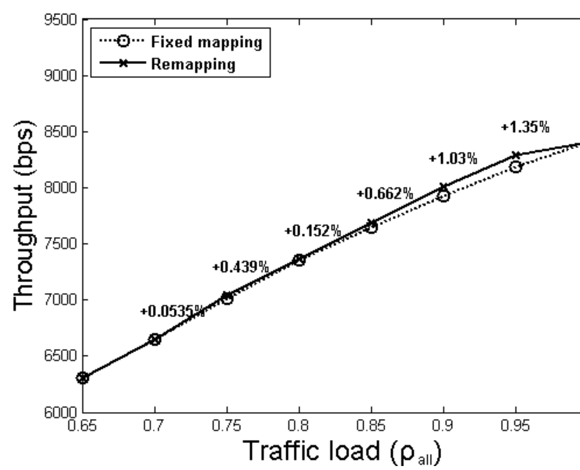


Fig. 20 Throughput: Remapping vs. Fixed mapping (uplink packet size = 1728 bytes).

gain of FC combined with DT can be up to 30% under QoS load = 0.65 and 70% under QoS load = 0.85. The delay gain of FC with DT is mostly around 15%. On the other hand, as demonstrated in Fig. 17 and Fig. 18, the gain of FC with RED in terms of the goodput and delay is even greater than that of FC with DT. For example, the goodput gain of FC with RED can be up to 50% under QoS load = 0.65 and 150% under QoS load = 0.85 in the uplink mode.

4.2.2 Remapping

Figure 19 and Fig. 20 display the throughput of the Remapping scheme and the contrast Fixed mapping scheme. Since Remapping provides the mechanism to integrate buffer management of L2 scheduling queues, it can achieve higher throughput under heavy input loads. However, as the total input load is reaching the saturation point of the network capacity (i.e. 100%), the Remapping scheme does not make any good since all scheduling queues are almost full at most of the time. As shown in Fig. 21, the best gain of Remapping in terms of throughput is about 8%. Note that the gain of Remapping in the uplink mode is much smaller than that of the downlink mode since the operation of Virtual Reassembler in the uplink mode creates more traffic burstiness and weakens the benefit of Remapping. As shown in Figs. 22–24, with the help of Remapping, the goodput of Remapping

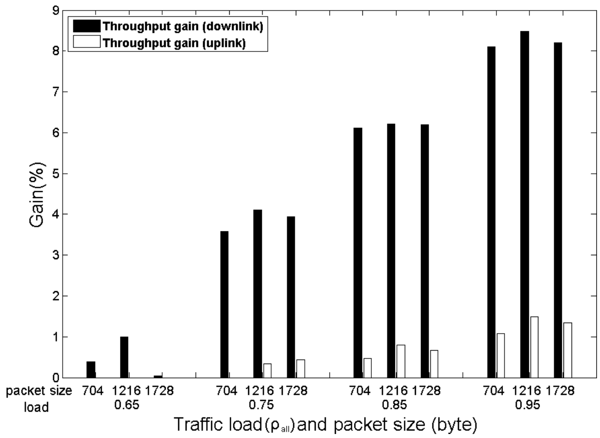


Fig. 21 Throughput gain of Remapping over Fixed mapping.

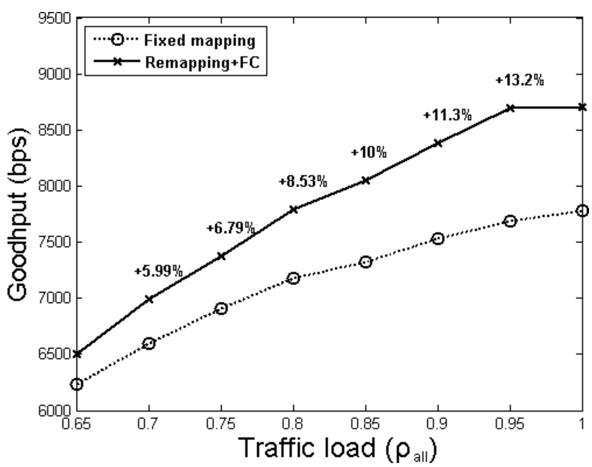


Fig. 22 Goodput: Remapping + FC (DT) vs. Fixed mapping (downlink packet size = 1728 bytes).

and Fragment Control is much better than the Fixed mapping mechanism. Moreover, the gain of Remapping and Fragment Control can be up to 71% under the total traffic load = 0.95.

4.2.3 Discussion

To compare the performance improvement by Fragment Control for the two congestion control schemes DT and RED, Fig. 25 and Fig. 26 display the performance gain of DT with FC over RED with FC in the downlink and uplink modes respectively. The figures demonstrate that the average goodput in the case of DT with FC is slightly greater (up to 7%) than that of RED with FC, but the average delay presents the opposite result. The reason is because in RED the random early dropping of data frames in the scheduling buffers decreases the queue length and thus reduces the throughput, the goodput, and the buffering time of each data frame. Moreover, the phenomenon implies that the selection of the congestion control scheme can be based on the performance demand of the application data.

By considering the delay gain and goodput gain of DT

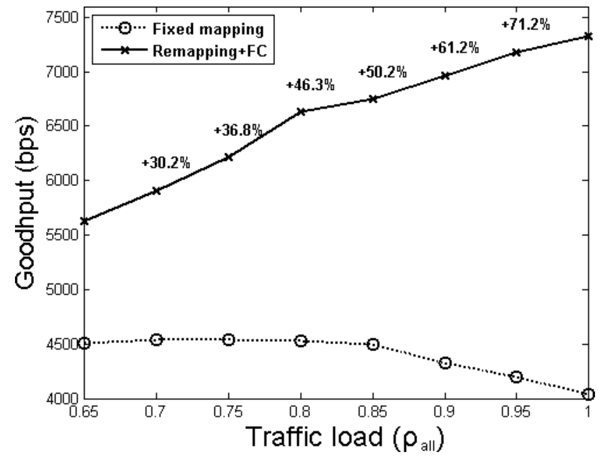


Fig. 23 Goodput: Remapping + FC (DT) vs. Fixed mapping (uplink packet size = 1728 bytes).

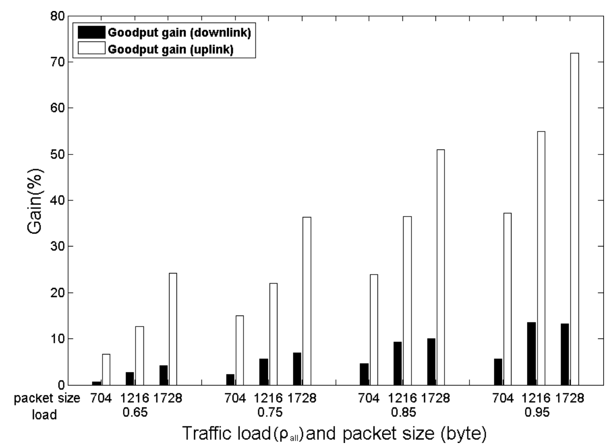


Fig. 24 Goodput gain of Remapping + FC (DT) over Fixed mapping.

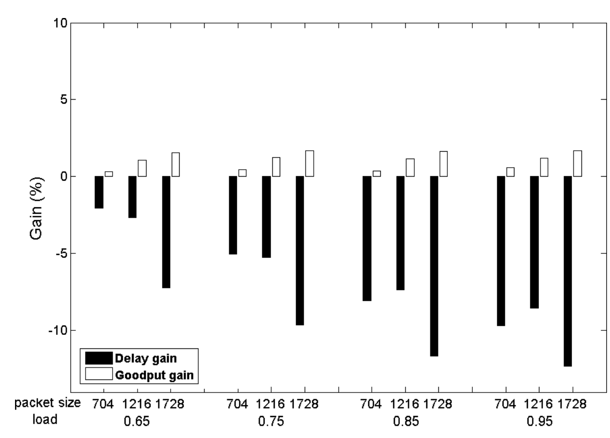


Fig. 25 Gain of FC + DT over FC + RED in downlink.

and RED in Fig. 25 and Fig. 26, we summarize the suggested congestion control scheme for different traffic types and load in Table 3. Since the goodput gain of DT over RED for downlink is insignificant as shown in Fig. 25, RED is suggested as the congestion control scheme to achieve higher

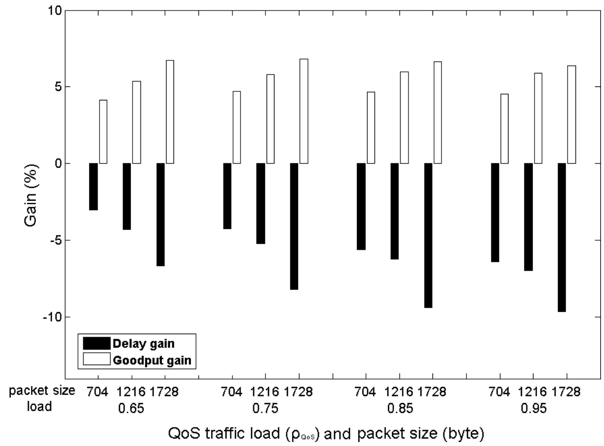


Fig. 26 Gain of FC + DT over FC + RED in uplink.

Table 3 Suggested congestion schemes for traffic type and load.

Traffic Type	***Downlink***		***Uplink***	
	Real-time	*File transfer*	*Real-time*	*File transfer*
Middle load (e.g. ρ_{QoS}=0.6)	RED		RED	DT
Heavy load (e.g. ρ_{QoS}=0.9)	RED		RED	DT

delay gain. For the uplink traffic, since the delay gain of RED and the goodput gain of DT are nearly equal, the traffic type that reflects the performance demand plays the role for selecting the proper scheme. For real-time traffic which is more delay-sensitive, RED is suggested to achieve high delay performance. DT, on the other hand, is used to achieve high goodput performance for application like file transfer. As also shown in Table 3, the traffic load does not have impact on the selection of the congestion control scheme.

5. Conclusion

As the most promising Wireless-MAN technology, IEEE 802.16 provides broadband, wide coverage, and QoS support to meet the demand of the next generation BWA (Broadband Wireless Access) network. To achieve the better QoS service in the IEEE 802.16 network, we proposed a cross-layer QoS framework integrating L3 and L2 QoS in the IEEE 802.16 network. Main functional blocks in the framework include: QoS mapping from L3 to L2, Admission control, Fragment Control, and Remapping. Fragment Control handles the data frames from the same IP datagram as a group in L2 operations to reduce useless transmission and delay time. Remapping is designed for more flexible use of L2 buffers by changing the mapping rules from IP QoS to L2 service type under congested situation of the rtPS and nrtPS queues. Extensive simulation results have demonstrated that the proposed framework as well as the associated mechanisms can achieve the better performance in terms of the delay, goodput and throughput in the heavy input load.

References

[1] IEEE Std 802.16-2004, "IEEE standard for local and metropolitan area networks — Part 16: Air interface for fixed broadband wireless access systems," Oct. 2004.

[2] IEEE Std 802.16e-2005, "IEEE standard for local and metropolitan area networks — Part 16: Air interface for fixed broadband wireless access systems — Amendment 2: Physical and medium access control layers for combined fixed and mobile operation in licensed bands," Feb. 2006.

[3] "Business case models for fixed broadband wireless access based on WiMax technology and the 802.16 standard," WiMax Forum, 10 Oct. 2004.

[4] S.J. Vaughan-Nichols, "Achieving wireless broadband with WiMax," Computer, vol.7, no.6, pp.10–13, June 2004.

[5] K. Wongthavarwat and A. Ganz, "Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems," International Journal of Communication Systems, vol.16, pp.81–96, Feb. 2003.

[6] H.S. Alavi, M. Mojdeh, and N. Yazdani, "A quality of service architecture for IEEE 802.16 standards," Proc. 2005 Asia-Pacific Conference on Communications, pp.249–253, Oct. 2005.

[7] H. Wang, B. He, and D.P. Agrawal, "Admission control and bandwidth allocation above packet level for IEEE 802.16 wireless MAN," Proc. 12th International Conference on Parallel and Distributed Systems (ICPADS 2006), vol.1, pp.1–6, July 2006.

[8] L. Wang, F. Liu, Y. Ji, and N. Ruangchaijatupon, "Admission control for non-preprovisioned service flow in wireless metropolitan area networks," Proc. 4th European Conference on Universal Multiservice Networks (ECUMN 2007), pp.243–249, Feb. 2007.

[9] D.H. Cho, J.H. Song, M.S. Kim, and K.J. Han, "Performance analysis of the IEEE 802.16 wireless metropolitan area network," Proc. First International Conference on Distributed Frameworks for Multimedia Applications (DFMA 2005), pp.130–136, Feb. 2005.

[10] H. Wang, W. Li, and D.P. Agrawal, "Dynamic admission control and QoS for 802.16 wireless MAN," Proc. Wireless Telecommunications Symposium (WTS 2005), pp.60–66, April 2005.

[11] R. Mukul, P. Singh, D. Jayaram, D. Das, N. Sreenivasulu, K. Vinay, and A. Ramamoorthy, "An adaptive bandwidth request mechanism for QoS enhancement in WiMax real time communication," Proc. IFIP International Conference on Wireless and Optical Communications Networks, pp.1–5, April 2006.

[12] A. Sayenko, O. Alanen, J. Karhula, and T. Hamalainen, "Ensuring the QoS requirements in 802.16 scheduling," Proc. 9th ACM International Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2006), pp.108–117, Oct. 2006.

[13] J. Chen, W. Jiao, and Q. Guo, "Providing integrated QoS control for IEEE 802.16 broadband wireless access systems," Proc. IEEE 62nd Vehicular Technology Conference (VTC 2005-Fall), vol.2, pp.1254–1258, Sept. 2005.

[14] J. Chen, W. Jiao, and H. Wang, "A service flow management strategy for IEEE 802.16 broadband wireless access systems in TDD mode," Proc. IEEE International Conference on Communications (ICC 2005), vol.5, pp.3422–3426, May 2005.

[15] C. Cicconetti, L. Lenzini, E. Mingozzi, and C. Eklund, "Quality of service support in IEEE 802.16 networks," IEEE Netw., vol.20, no.2, pp.50–55, March-April 2006.

[16] J.A. Perez, V.H. Zarate, C. Cabrera, and J. Janecek, "A network and data link layer infrastructure design to improve QoS for real time collaborative systems," Proc. Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services (AICT/ICIW 2006), p.19, Feb. 2006.

[17] J.A. Perez, V.H. Zarate, and C. Cabrera, "A network and data link layer QoS model to improve traffic performance," Lecture Notes in Computer Science (LNCS), vol.4097, pp.224–233, Springer-Verlag, Aug. 2006.

[18] C. Wang, P.C. Lin, and T. Lin, "A cross-layer adaptation scheme

for improving IEEE 802.11e QoS by learning," IEEE Trans. Neural Netw., vol.17, no.6, pp.1661–1665, Nov. 2006.

[19] C. Wang, T. Lin, and J.L. Chen, "A cross-layer adaptive algorithm for multimedia QoS fairness in WLAN environments using neural networks," IET Communications, vol.1, no.5, pp.858–865, Oct. 2007.

[20] H. Jiang, W. Zhuang, and X. Shen, "Cross-layer design for resource allocation in 3G wireless networks and beyond," IEEE Commun. Mag., vol.43, no.12, pp.120–126, Dec. 2005.

[21] I. Gutierrez, F. Bader, J. Pijoan, and S.B. Slimane, "Adaptive resource management for a MC-CDMA system with mixed QoS classes using a cross layer strategy," Proc. IEEE 65th Vehicular Technology Conference (VTC 2007-Spring), pp.3036–3040, April 2007.

[22] N. Anastácio, F. Merca, O. Cabral, and F.J. Velez, "QoS metrics for cross-layer design and network planning for B3G systems," Proc. IEEE International Symposium on Wireless Communications Systems (ISWCS 2006), pp.592–596, Sept. 2006.

[23] R. Braden, D. Clark, and S. Shenker, "Integrated services in the Internet architecture: An overview," IETF RFC 1633, June 1994.

[24] J. Wroclawski, "The use of RSVP with IETF integrated services," IETF RFC 2210, Sept. 1997.

[25] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," IETF RFC 2475, Dec. 1998.

Yi-Ting Mai received his B.S. degree in Department of Science Education from National Hualien University of Education, Hualien, Taiwan, in 1998 and M.S. degrees in Department of Information Management from Chaoyang University of Technology, Taichung, Taiwan, in 2004. He is currently a Ph.D. student in the Department of Computer Science and Information Engineering, National Chi Nan University (NCNU), Puli, Taiwan. His current research topics include wireless networks, QoS, and multimedia network protocols.

Chun-Chuan Yang received his B.S. degree in computer and information science from National Chiao-Tung University, Taiwan, in 1990 and Ph.D. degree in computer science from National Taiwan University in 1996. He joined the Department of Computer Science and Information Engineering, National Chi-Nan University (NCNU), Puli, Taiwan, as an Assistant Professor in 1998. Since February 2008, he has been a Full Professor. His research area of interests includes multimedia network protocols, multimedia synchronization control, and multimedia applications.

Yu-Hsuan Lin received his B.S. degree in Department of Information Management from Nation Taichung Instituted of Technology, Taiwan, in 2004 and M.S. degree in computer science from National Chi Nan University, Puli, Taiwan, in 2006. His current research topic includes 802.16 MAC QoS protocols.