A Multi-hop Polling Service with Bandwidth Request Aggregation in IEEE 802.16j Networks

Chun Nie Thanasis Korakis Shivendra Panwar
Department of Electrical and Computer Engineering, Polytechnic University, Brooklyn 11201
Email: {cnie01@students.poly.edu, korakis@poly.edu, panwar@catt.poly.edu}

Abstract— **The IEEE 802.16j protocol for a multi-hop relay (MMR) WiMAX network is being developed to increase data rates and extend service coverage as an enhancement of existing WiMAX standards. The IEEE 802.16j protocol supports transparent and non-transparent modes. In the *transparent mode*, only data traffic is relayed by an intermediate relay station (RS) between a mobile station (MS) and the base station (BS), while in the *non-transparent mode*, both signaling and data traffic are forwarded by RSs. Furthermore, non-transparent mode is either distributed or centralized with regard to scheduling. The difference between them resides in that distributed scheduling enables RSs to participate in bandwidth allocation (*BWAlloc*), while centralized scheduling leaves all *BWAlloc* coordinated by the BS. In this paper, we propose a novel multi-hop polling service (*mPS*) for non-transparent centralized scheduling in a multi-hop 802.16j environment. Our model is adaptive to the traffic pattern so as to provide bandwidth efficiency over access and relay links. Besides, aggregation of bandwidth requests (*BWReq*) from MSs is conducted at the RS to further save bandwidth. The performances of mPS with *BWReq* aggregation is evaluated via simulations which demonstrate our approach outperforms the current multi-hop bandwidth request mechanism in terms of overall spectrum efficiency.**

I. INTRODUCTION

WiMAX has emerged as an advanced broadband wireless access technology and has attracted a lot of attention. While prior WiMAX standards, such as IEEE 802.16d/e [1] [2], have well defined specifications for the legacy single-hop network, a multi-hop scenario is now being deployed. Recently the 802.16j Relay Task group was formed to standardize a WiMAX multi-hop relay (MMR) system. In an MMR system, MSs are allowed to route through intermediate RSs to reach the BS, which differs from the single-hop WiMAX topology. While the 802.16j standards are yet to be discussed and finalized, the basic hierarchy of an MMR WiMAX network has already been proposed in [3]. In [3], three network elements, BS, RS and MS, are defined in an MMR WiMAX network, (see Fig. 2). These three elements establish the hierarchical topology of a MMR network. Unlike a single-hop WiMAX system, RSs work as intermediate nodes between the BS and the MSs, and forward signals between the two ends. Based on the functionality of an RS, IEEE 802.16j has classified MMR systems into a *transparent mode* and a *non-transparent mode* [3]. The RS in *transparent mode* does not forward management signaling but merely forwards data traffic, while an RS in *non-transparent mode* forwards both management signaling and data traffic. Specifically, the *non-transparent mode* is either distributed or centralized. In the former case, the RS participates in *BWAlloc* along with the BS, while the latter mode only allows the BS to schedule *BWAlloc*. In an MMR network, the radio link from MSs and their superordinate RS is defined as an access link (AL), while the radio link from the RS to the BS is defined as the relay link (RL). The centralized *non-transparent mode* is examined by our study because the non-transparent mode can extend cell coverage without installing more BSs and centralized *BWAlloc* is relatively simpler than distributed approach.

As defined in [1] [2], IEEE 802.16d/e networks have a centralized medium access control (MAC) layer. That is, all required bandwidth for the uplink (UL) applications has to be scheduled and granted by the BS. When an MS needs to transmit to the BS in the UL, *BWAlloc* is conducted via a bandwidth request/grant process between the MS and the BS. Corresponding to the traffic characteristics of different services, five types of scheduling services have been defined[1]: unsolicited grant service (UGS), real-time polling service (rtPS), non-real-time polling service (nrtPS), extended real-time polling service (ertPS) and best effort (BE) service. Among them, UGS, rtPS and ertPS are mainly used for real-time (RT) traffic, while nrtPS and BE are usually utilized for non-real-time (NRT) traffic. While several research efforts have been devoted to improving *BWAlloc* mechanism in the UL of WiMAX network [4]–[6], they are all concentrated on a single-hop scenario. In [7], [8], a number of *BWAlloc* mechanisms are proposed for a multi-hop WiMAX system. Despite their improved performance bandwidth efficiency, these methods are only focused on distributed scheduling while centralized scheduling is unexplored. A bandwidth allocation method was studied in [9] for a IEEE 802.16j system. However, it does not consider the operation of any polling service.

In this paper, we propose a novel framework for a multi-hop polling service (mPS) to facilitate efficient UL bandwidth allocation in a centralized non-transparent MMR WiMAX, suitable for many RT applications. We aim to design a polling

This work is supported by the NSF Grant CNS-0435303, and also by the New York State Center for Advanced Technology in Telecommunications (CATT) and the Wireless Internet Center for Advanced Technology (WICAT).

[1]Scheduling service refers to the data handling scheme for data transport for the MAC layer.

service in a multi-hop topology catering to bursty applications and achieving optimal bandwidth efficiency without compromising delay performance. Compared to existing *BWAlloc* methods, our contributions are summarized as follows.

- mPS is appropriate for most data applications which exhibit bursty ON/OFF traffic patterns and achieves a good tradeoff between delay and signaling overheads.
- mPS is BS/RS initiated to simplify the design of the MS.
- mPS easily manages delay performance for RT applications and substantially decreases signaling overhead in a MMR environment.
- mPS can incorporate piggybacking and bandwidth stealing for further performance enhancement.
- mPS is combined with *BWReq* aggregation to reduce signaling cost and delays over the RL.

The rest of the paper is organized as follows. In Section II, we examine existing scheduling services in single-hop and MMR WiMAX system. In Section III, we propose mPS as a new scheduling service and describe its functionality. We present a protocol designed to implement mPS service. In Section IV, we present the simulation results with respect to typical scenarios, followed by the conclusions in Section V.

II. Conventional Scheduling Services in WiMAX

A. Scheduling services in the IEEE 802.16e standard

1) UGS: UGS enables the BS to allocate fixed-size bandwidth periodically to the MS.

2) rtPS: rtPS is designed for RT applications. It enables the BS to poll the MS at fixed intervals for bandwidth requests.

3) ertPS: ertPS is designed for voice application. During a voice talkspurt, the BS behaves as UGS to grant unsolicited bandwidth to the MS periodically. During voice silence, ertPS decreases the grant size to save bandwidth.

4) nrtPS and BE: nrtPS and BE are two scheduling services for delay-insensitive applications by contention-based bandwidth requests.

In summary, these scheduling services are basically designed for a single-hop WiMAX system without considering the requirements of a multi-hop system. In addition, these scheduling services are not appropriate for bursty RT applications, as these approaches either can not quickly obtain a grant from the BS (nrtPS, BE), or are not spectrally efficient (UGS, rtPS, ertPS).

B. Enhanced Scheduling Services in the IEEE 802.16j Baseline

1) Distributed BWAlloc Algorithms: In MMR scenarios, several distributed scheduling methods have already been proposed in the IEEE 802.16j task group [7], [8]. While these contributions address the *BWAlloc* issue in an MMR network, they are all distributed algorithms, which enable an RS-based *BWAlloc* capability. These distributed algorithms complicate the design of the RS and the associated signaling.

2) Alcatel-Lucent BWAlloc Algorithm: A centralized *BWAlloc* algorithm was proposed in [9] for the downlink (DL) of multi-hop WiMAX system. However, it does not include the associated signaling operation in the UL of a 802.16j network. Besides, this algorithm does not take traffic patterns into account, and thus is not suited for bursty applications.

III. Proposed Polling Algorithm- Multi-hop PS

In this paper, we are interested in supporting RT applications in multi-hop WiMAX networks. We will introduce a generic RT traffic model and proceed with a multi-hop polling service (mPS) as our proposed scheduling service.

A. An RT Bursty ON/OFF Traffic Model

RT applications, which recently have become increasingly popular, usually exhibit bursty traffic characteristics. We illustrate a typical bursty application in Fig. 1.

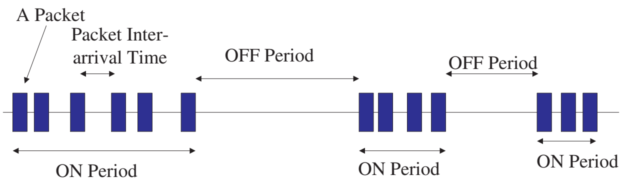

Fig. 1. The Generic Bursty RT Data Applications.

As shown in Fig. 1, the traffic pattern of applications, such as VoIP and gaming, are characterized by bursty ON/OFF flows with varying rates and packet sizes. During each ON period, the variable-sized packets arrive in bursts with variable packet interarrival time. During each OFF period, no packet is generated and the MS is idle. Moreover, these applications also have latency constraints on packet delivery. Therefore, a novel multi-hop polling service is needed to provide timely data delivery and minimize the involved signaling overhead associated with MMR *BWAlloc* in IEEE 802.16j system.

B. Multi-hop Scenarios in 802.16j

In IEEE 802.16j, the MMR system model is illustrated in Fig. 2. A BS is located in the center of a cell and several RSs surround the BS. Each RS serves a number of MSs. In this paper, we consider a non-transparent centralized scheduling mode with only one intermediate RS between the MS and the BS.

C. The Multi-hop Polling Service (mPS) Mechanism

We assume a basic IEEE 802.16j WiMAX system model with one BS, one RS and M MSs. Suppose MSs are far from the BS so that they indirectly communicate with the BS via the RS. The RS serves as relay node to forward both signaling and data traffic between the BS and MSs. This basic relay model, although simplified, generalizes the non-transparent centralized 802.16j network and can be extended to a general multi-hop and multi-RS topology.

Fig. 3 illustrates the mPS mechanism. The BS polls the RS at intervals of T_0, while the RS polls MSs at intervals of

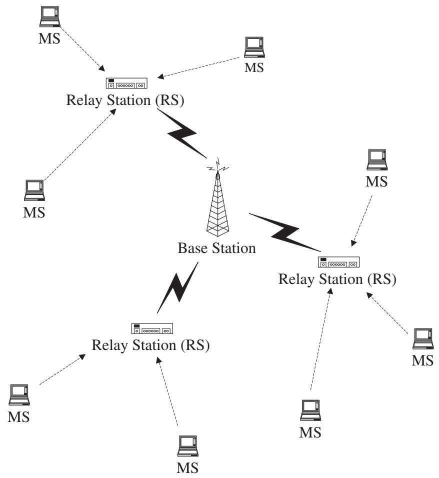

Fig. 2. The Centralized Multi-hop WiMAX Network.

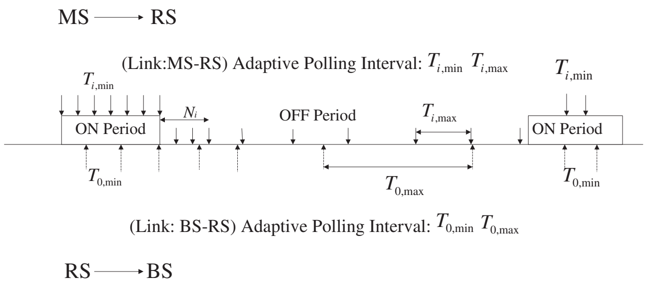

Fig. 3. Multi-hop Polling Service.

T_i, where i denotes the ith ($1 \leq i \leq M$) MS. The polling intervals of AL and RL, T_i, ($0 \leq i \leq M$) are variable and jointly given by (1) below. See Table I for notation, which will be further described later. Instead of relying on an MS state report as prior approaches do, the polling intervals of mPS in AL and RL are both adaptively updated at the RS and the BS by observing the UL data transmission pattern.

TABLE I
MULTI-HOP PS DEFINITION

Parameters	Meaning
M	Number of MSs within the cell
T_i	Polling interval of the ith MS ($1 \leq i \leq M$) or RS (i=0)
$T_{i,min}$	Minimum polling interval used by the ith MS ($1 \leq i \leq M$)
$T_{i,max}$	Maximum polling interval used by the ith MS ($1 \leq i \leq M$)
$T_{0,min}$	Minimum polling interval used by the RS ($i = 0$)
$T_{0,max}$	Maximum polling interval used by the RS ($i = 0$)
N_i	Number of initial polls with $T_{i,min}$ during idle period of the ith MS
N_0	Number of initial polls with $T_{0,min}$ during idle period of RS

$$T_i = \begin{cases} T_{i,min} & n = 1, 2, 3, ..., N_i \\ min\{2^{n-N_i} \times T_{i,min}, T_{i,max}\} & n > N_i, \end{cases} \tag{1}$$

where $0 \leq i \leq M$ and n denotes the number of polls after an idle period starts.

Specifically, let us discuss the mPS in more details. At the beginning, the ith MS is polled by RS with a fixed interval $T_{i,min}$ over the AL. When no bandwidth is requested by the MS for N_i consecutive polls, the BS starts to poll the ith MS with an exponentially increasing interval until $T_{i,max}$ is reached. After that point, the polling interval is fixed at $T_{i,max}$. When a polling message leads to a *BWReq* by the ith MS for data transmission, T_i is reset by RS to $T_{i,min}$. Usually each MS may have different $T_{i,min}$ and $T_{i,max}$ depending on their respective QoS constraints, such as delay and delay jitter. By adaptively updating polling intervals of MSs, $BWReq$ is substantially reduced so as to save considerable signaling bandwidth, albeit at the cost of increased delay. The proposed adaptive polling interval provides good performance under the ON-OFF traffic pattern, shown in Fig. 1. Moreover, even under a general bursty traffic pattern without clear ON-OFF periods, adaptive polling intervals still offer reasonably good performance, as shown later in simulations. In a similar manner, the RS is also successively polled by BS with an initial interval $T_{0,min}$ updated in accordance with (1).

In addition to adaptive polling intervals in AL and RL, *BWReq* aggregation also contributes to a reduction in signaling. When the RS is polled by the BS, RS collects all *BWReq* from its subordinate MSs and generates a new *Aggregated BWReq* requesting the total bandwidth of all MSs from the BS. This *Aggregated BWReq* is forwarded to the BS. Once it receives a grants from the BS, the RS will further assign specific grant to individual MSs. Obviously, the use of *Aggregated BWReq* at the RS considerably decreases signaling overhead.

An interesting topic is how to properly configure polling parameters to achieve the desired delay performance. $T_{0,max}$+$T_{i,max}$, denoting the longest possible access delay for a *BWReq*, should meet the tolerable delay budget of the RT traffic, allowing enough time for scheduling delays. Since the BS has the knowledge of scheduling parameters, such as the state of the UL grant request queue, it can estimate the scheduling delay and properly manage $T_{0,max}$ and $T_{i,max}$. Since packet inter-arrival times vary, the BS can utilize a statistical characterization of the traffic pattern to determine a good choice of N_i. For typical applications, a small N_i should suffice. The basic ideas of mPS are summarized by

- MSs send their *BWReq* to their superordinate RS at their respective polling intervals.
- The RS collects all *BWReq*s from MSs and generates an *Aggregated BWReq* to the BS.
- The BS, instead of allocating bandwidth to MSs directly, grants bandwidth to the intermediate RS, which then allocates bandwidth to individual MSs.
- As traffic from the MS is bursty, the polling intervals in both AL and RL are adapted to the transitions of ON/OFF cycles. During ON periods, polling intervals are fixed and short, while during OFF periods polling intervals are lengthened exponentially.
- The main advantage of mPS employing *BWReq* aggregation is to significantly decrease the signaling overhead

without compromising delay performance.

- mPS can incorporate the use of piggybacking and bandwidth stealing defined in the standards, with the aim of further improving the bandwidth efficiency.

Consequently, the polling mechanism is carried out on both AL and RL. through appropriately selecting polling parameters, the delay and signaling overhead can be jointly optimized.

D. The Protocol design of mPS

The proposed bandwidth rquest algorithm to be implemented in mPS is illustrated in Fig. 4.

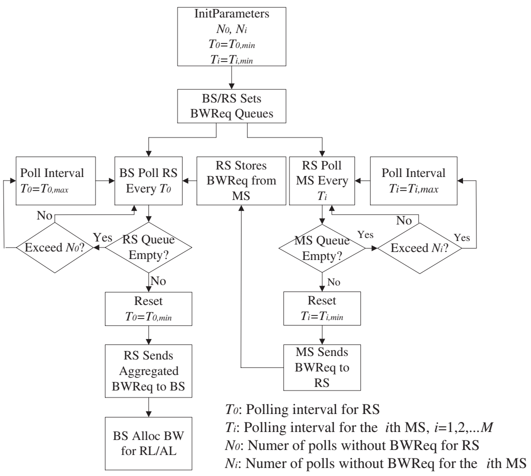

Fig. 4. The Proposed Bandwidth Request Algorithm (mPS) in WiMAX.

In Fig. 4, firstly, the BS and RS determine system parameters, $T_{i,min}$ and $N_i (0 \leq i \leq M)$, based on the application's QoS requirements. The ith MS in the AL is polled by RS at intervals of T_i. The MS with data to send responds with a *BWReq* which is stored in the *BWReq* queue at the RS. At the same time, the RS is polled by the BS at intervals of T_0. Then RS aggregates *BWReq* from MSs into a single *Aggregated BWReq* and transmits it to BS.

The BS, upon receiving *Aggregated BWReq* from RS, schedules its resources and sends back a grant to the RS which assigns bandwidth on both the AL and RL. The BS allocates successive frames to AL and RL to minimize the data delivery delay. During the process of mPS, T_i, $(0 \leq i \leq M)$ is updated by equation (1). In case of ongoing data traffic over the RL, the RS can piggyback data to send an *Aggregated BWReq* to the BS directly without waiting the next polling message. This piggyback function further improves bandwidth efficiency.

IV. Performance Evaluation

We conducted Monte Carlo simulations to demonstrate the efficiency of mPS. The simulation environment consists of one BS, one RS and multiple MSs, comparing rtPS with mPS for the same application. We study UL traffic and focus on the impact of mPS and *BWReq* aggregation. The configurations of rtPS and mPS are defined in Table II. We suppose $T_{i,min}$, $T_{i,max}$, and N_i to be the same for AL and RL.

We measure the average delay performance in terms of air-link transmission delay (95% confidence interval), i.e., the time from a packet arrival till reception at the BS. The signaling overhead is measured in terms of the average data rate of *BWReq* from the MS, where each *BWReq* message consumes 6 bytes and each *Aggregated BWReq* is assumed to be 12 bytes in our experiments.

TABLE II
rtPS and mPS Parameters

Parameter Names	Values
rtPS Polling Interval	20ms
mPS parameter N_i, $0 \leq i \leq M$	1
Scenario 1 (Min,Max) T_i Interval	(20ms,40ms)
Scenario 2 (Min/Max) T_i Interval	(20ms,80ms)
Scenario 3 (Min/Max) T_i Interval	(20ms,160ms)
Scenario 4 (Min/Max) T_i Interval	(20ms,320ms)
Scenario 5 (Min/Max) T_i Interval	(20ms,640ms)
Scenario 6 (Min/Max) T_i Interval	(20ms,1280ms)

A. A generic bursty ON/OFF application

We first consider a generic bursty ON/OFF UDP traffic models to illustrate the performance tradeoff in mPS, with the simulation parameters listed in Table III.

TABLE III
UDP System Parameters

Parameter Names	Values
Flow Length Distribution (seconds)	Exponential (200)
(On, Off) State Distribution (seconds)	Exponential (10, 3)
Packet Inter-arrival Time Distribution (seconds)	Exponential (0.1)
Max Packet Size (bytes)	1500

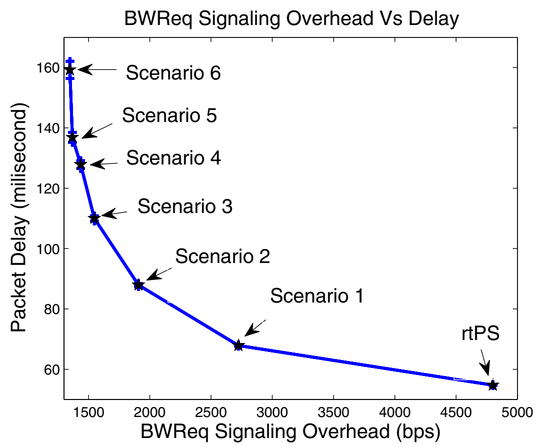

Fig. 5. Generic UDP: Delay vs. BWReq signaling overhead.

Fig. 5 displays the delay performance and signaling overheads under several scenarios. It is clear that rtPS has the least delay but its signaling overhead is the highest. The delay of mPS slowly increases from 80ms to 135ms when the maximum polling intervalis lengthened. Nevertheless, the signaling overhead of mPS is appreciably cut down by about 44% ˜ 75%. The increase in delay is only marginal and is still acceptable to most applications.

B. An online game application

We consider a popular server/client-based online game named Age of Kings [10]. Game traffic is characterized by fairly bursty packets, but without clear ON-OFF periods. The packet arrivals exhibit a large jitter, which may be exploited by mPS to offer overhead reduction. The parameters of the game is given in Table IV.

TABLE IV
ONLINE GAME (AGE OF KINGS) PARAMETERS

Parameter Names	Values
Traffic Direction	Client to Server
Packet Inter-arrival Time Distribution (seconds)	Normal (0.136, 0.296)
Mean Bit Rate (kbps)	3.06

Fig. 6 illustrates the delay and *BWReq* signaling overhead for online gaming. The range of packet delay is 80ms-130ms for mPS, compared to the 80ms delay for rtPS. Although this game does not have clear ON-OFF period, mPS still substantially reduces the signaling overhead by 46-75% with tolerable delay. Such overhead reduction is critical for spectral efficiency since the online game uplink average traffic rate is only 3.06 kbps, (Table IV), while rtPS signaling overhead consumes bandwidth of the same order.

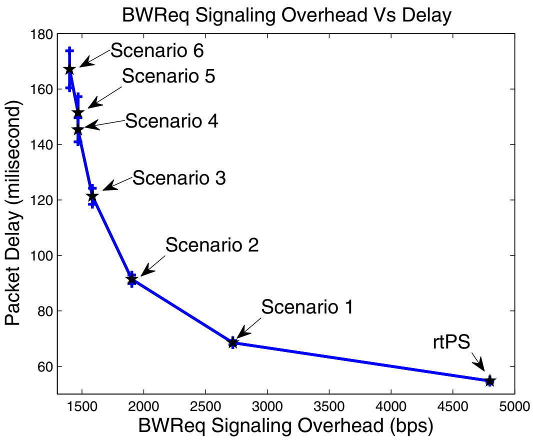

Fig. 6. Online gaming: packet delay vs. BWReq signaling overhead.

C. An online game application with BWReq Aggregation

When bandwidth request aggregation is carried out with mPS, the consumed signaling overhead is further reduced over the relay link. Accordingly to IEEE 802.16e standards, the size of each *BWReq* is 6 bytes long, of which 11 bits are used to specify the length of requested bandwidth in units of bytes. We assume that each *Aggregated BWReq* is twice the size of regular *BWReq*. Here we suppose 8 homogeneous MSs with the same online game traffic are served in the cell.

From Fig. 7, delay and signaling overhead are both further improved with *BWReq* aggregation, e.g., scenarios 2-4. This is because by using *BWReq* aggregation, less bandwidth is required over the RL to deliver *Aggregated BWReq*. Moreover, *BWReq* arrivals from MSs are multiplexed at the RS so that the interval between successive *Aggregated BWReq* is shortened, which leads to a reduced access delay caused by polling.

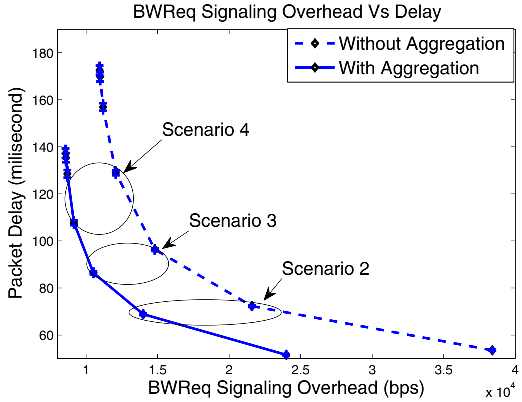

Fig. 7. Online Game with aggregation: Delay vs. BWReq signaling overhead.

V. CONCLUSION

In this paper, we have developed a novel multi-hop polling service for the IEEE 802.16j MMR network. Our proposed mPS adapts the length of polling intervals to idle periods. The mPS consumes considerably less bandwidth while still satisfying the delay constraints of most RT applications. From numerical results, our mPS mechanism enhances bandwidth efficiency significantly, especially when combined with *BWReq* aggregation. Furthermore, the proposed algorithm is easy to implement without any modification to MSs.

REFERENCES

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

[2] IEEE 802.16e, "IEEE standard for local and metropolitan area networks-part 16: Air interface for fixed broadband wireless access systems, amendment 2 and corrigendum 1", *IEEE Standard*, Feb. 2006.

[3] IEEE 802.16j, "Part 16: Air interface for fixed and mobile broadband wireless access systems: Multihop relay specification", *IEEE Baseline Document for Draft Standard for Local and Metropolitan Area Networks*, Apr. 2007.

[4] M. Venkatachalam, "Optimizing the uplink bandwidth request latency", *IEEE 802.16 Broadband Wireless Access Working Group*, Nov. 2005.

[5] H. Lee, D. Cho, T. Kwon, K. Lee, S. Choi, J. Kim, S. Cho, S. Yun, W. Park, Y. Kim, Y. Chang, G. Lim, H. Chang, J. Kim, and T. Kim, "Extended rtPS for VoIP services", *IEEE 802.16 Broadband Wireless Access Working Group*, Aug. 2004.

[6] C. Nie, M. Venkatachalam, and X. Yang, "Adaptive polling service for next-generation IEEE 802.16 wimax networks", in *Proc., IEEE Globecom*, Nov. 2007.

[7] S. Feng, Y. Lu, T. Li, L. Zhang, M. Okuda, Y. Zhou, K. Loa, Y. Tsai, and Y. Lee, "Optimized distributed bandwidth request and allocation in 802.16j system", *IEEE C802.16j-07/187r3, IEEE 802.16 Broadband Wireless Access Working Group*, 2007.

[8] H. Qu, S. Cai, M. Chion, Y. Liu, and Y. Chen, "Signaling scheme for bandwidth allocation in MMR network with distributed scheduling", *IEEE C802.16j-07/207r2, IEEE 802.16 Broadband Wireless Access Working Group*, 2007.

[9] E. Liu, D. Wang, J.Liu, G. Shen, and S. Jin, "Performance evaluation of bandwidth allocation in 802.16j mobile multi-hop relay networks", in *Proc., IEEE Veh. Technology Conf.*, 2007, pp. 939 – 943.

[10] J. Lakkakorpi, A. Heiner, and J. Ruutu, "Measurement and characterization of internet gaming traffic", in *Research Seminar on Networking*, Helsinki University of Technology, Networking Laboratory, Espoo, Finland, Feb. 2002.