Computer Communications 31 (2008) 2350–2359

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Energy-efficient packet scheduling algorithms for real-time communications in a mobile WiMAX system

Shiao-Li Tsao *, You-Lin Chen

Department of Computer Science, National Chiao Tung University, Room EC426, No. 1001 Ta Hsueh Road, Hsinchu 300, Taiwan, ROC

ARTICLE INFO

Article history:
Received 17 March 2007
Received in revised form 23 February 2008
Accepted 23 February 2008
Available online 7 March 2008

Keywords:
Mobile WiMAX
Energy efficiency
Power saving
Packet scheduling

ABSTRACT

Broadband wireless access systems usually provide flexible sleep-mode operations for mobile stations to conserve their energy during idle or active mode. For example, Mobile WiMAX, i.e. the IEEE 802.16e, offers several power-saving classes that can be associated with different types of network connections to minimize power consumption of mobile stations. Unfortunately, previous studies did not fully utilize the sleep-mode features to save the energy of a mobile station with multiple real-time connections, and power consumption of a mobile station is not yet optimized. In this work, two energy-efficient packet scheduling algorithms for real-time communications in a Mobile WiMAX system are proposed. The schemes not only guarantee the quality of services (QoSs) of real-time connections but also minimize power consumption of mobile stations. Simulation results demonstrate that the proposed schemes outperform the traditional approach.

1. Introduction

One of the essential features for a broadband wireless access (BWA) system which is designed for portable and battery-operated devices is the power-saving function. For example, the IEEE 802.11, i.e. WLAN, defines a power-saving mode which can be utilized to improve energy efficiencies for web accesses, voice over IP (VoIP), and other applications over WLAN [7,8]. The IEEE 802.16e, so called Mobile WiMAX, that has been newly developed also provides flexible power-saving classes to facilitate mobile stations to conserve their energy during active and sleep mode [1,2,4]. According to the specification, an IEEE 802.16e mobile station can switch to sleep mode for a sleep period, and wakes up to send or receive packets in a listen period. During sleep periods, a base station (BS) must buffer incoming packets sent to the mobile station, and then after the mobile station switches to listen periods, the base station sends the queued packets to the mobile station. To accommodate different characteristics of applications and services, the IEEE 802.16e specifies three power-saving classes and each power-saving class implies a particular sleep and listen behavior for a mobile station. A mobile station can thus associate a power-saving class with a connection and negotiates the parameters of the power-saving class such as the time to sleep and listen, and the length of each sleep and listen period with the base station for the connection. Obviously, the parameters of a power-saving class associated with a network connection should be carefully decided in order to maximize the energy efficiency of a mobile station without violating the QoS requirements of that connection.

Admission control schemes and scheduling algorithms are designed to offer QoS services in wireless networks, and a number of studies have investigated these issues in IEEE 802.16/802.16e networks [3,5,9,10]. For example, Wongthavarawat and Ganz [12] proposed the architecture of the uplink scheduler. They assign different service types with strict priorities, and schedule each service type by using a particular scheme. Unfortunately, these scheduling algorithms do not consider the power consumption of a mobile station. Several studies [6,13,14,16] investigated the power consumption issues of IEEE 802.16e and suggested algorithms to determine the sleep interval in improving its energy efficiency. However, above studies mainly consider non-real-time connections in IEEE 802.16e networks. Shi et al. [15] proposed a burst scheduling mechanism which guarantees the minimum data rates of mobile stations and schedules packets in a busty basis so that it can maximize the sleep time of mobile stations. Unfortunately, the burst scheduling approach may not be suitable to these delay sensitive applications and services. Although a number of energy saving mechanisms have been proposed for the IEEE 802.16e, to minimize power consumption of IEEE 802.16e mobile stations with multiple real-time connections has not yet been investigated. In this paper, we study this problem and propose two packet scheduling schemes to maximize sleep periods of a mobile station without violating the QoS requirements of the real-time connections.

* Corresponding author. Tel.: +886 3 5712121x54717; fax: +886 3 5721490.
E-mail address: sltsao@cs.nctu.edu.tw (S.-L. Tsao).

0140-3664/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2008.02.030

The rest of the paper is organized as follows. The IEEE 802.16e sleep-mode operations and the power consumption problems for a mobile station with multiple real-time connections are described in Section 2. The proposed scheduling schemes, called periodic on–off scheme (PS) and aperiodic on–off scheme (AS), are presented in Section 3. The simulation environment and simulation results are discussed in Section 4, and finally, Section 5 concludes this study.

2. The IEEE 802.16e sleep-mode operations and power consumption problem

In a Mobile WiMAX system, a mobile station can switch to sleep mode if there is no packet to send or receive in order to save power. The IEEE 802.16e defines three power-saving classes to accommodate network connections with different characteristics. According to the specification, each connection on a mobile station can be associated with a power-saving class, and connections with a common demand property can be grouped into one power-saving class. The parameters of a power-saving class, i.e. the time to sleep and listen, the length of a sleep period and a listen period can be negotiated by a base station and a mobile station.

The type-one power-saving class specifies that a mobile station sleeps for a period, wakes up to listen for incoming packets, and repeats sleep and listen operations. If there is no packet to send or receive during a listen period, a mobile station doubles the period for the next sleep. This power-saving class is suitable for the connections of web browsing or data access services. The type-two power-saving class requires a mobile station to repeat the sleep and listen on a round-robin basis, and the sleep and listen period are fixed. This sleep mode is appropriate for real-time connections such as VoIP and video streaming services that have packets to send or receive periodically. Based on the type-two sleep mode, a mobile station only needs to wake up to send or receive packets in those listen periods without violating the QoSs of the real-time connections. The type-three power-saving class defines the length of a sleep period, and a mobile station sleeps for that period and then returns to the normal operation. Fig. 1 illustrates examples for the three power-saving classes.

If a mobile station establishes multiple connections with different demand properties, the periods that a mobile station can sleep are determined by the sleep-mode behaviors associated with all connections. Fig. 2 shows an example that a mobile station has three connections. The connections have different demand properties, and associate with their preferred power-saving classes and parameters. It can be seen that the actual periods that a mobile station can sleep are the slots that three connections are all in a sleep period. Obviously, without a proper schedule of the sleep-mode operations for multiple real-time connections on a mobile station, the power consumption of a mobile station might not be reduced even the sleep mode is applied.

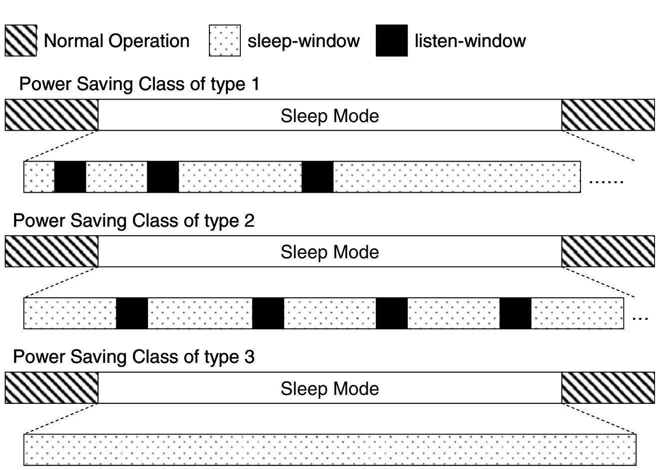

Fig. 1. Power-saving classes defined in the IEEE 802.16e.

In this work, only packet scheduling issues for mobile stations with multiple real-time connections are considered. Non-real-time packets that can tolerate delays could be scheduled in any listen period with available radio resources for a mobile station.

3. Energy-efficient packet scheduling algorithms

3.1. Periodic on–off scheme (PS)

The idea behind the first proposed approach, called periodic on–off scheme (PS), is to allow a mobile station to sleep for a fixed period and then to listen for another fixed period on a round-robin basis. The concept of on–off scheduling algorithms has been applied to wireless devices in reducing the power consumption [17,18]. In this study, we extend the concept and apply it to Mobile WiMAX system. Our scheme maximizes the length of a sleep period in the type-two power-saving class defined in the IEEE 802.16e without violating QoSs of all connections. During listen periods, a mobile station transmits and receives packets, and on other hand, the mobile station sets the interface idle to conserve the energy during sleep periods. Fig. 3 gives an example of a packet schedule for two real-time connections by applying the PS approach.

The PS is performed in two steps. The first step is to compute the length of a sleep period and a listen period for a mobile station. The second step is to let a mobile station enter a periodical sleep mode and schedule the packets according to the parameters obtained in the first step. A mobile station stays idle during sleep periods, and only wakes up to transmit data in listen periods. Packets sent to the mobile station during sleep periods are buffered at the base station and are delivered to the mobile station till listen periods. In other words, the mobile station only needs to receive and transmit data in listen periods and stay idle to conserve energy during sleep periods. The next paragraphs describe the detail of the first-step procedures of the PS. Also, notations used in this paper are summarized in Table 1.

To minimize power consumption of a mobile station with multiple real-time connections, the PS determines the length of a sleep period and a listen period under the radio resource and QoS constraints. Considering a mobile station with N real-time connections, the QoS parameters of connection i can be denoted as $Q_i\{S_i, TI_i, D_i\}$, where D_i is the delay constraint in milliseconds of any two consecutive packets for connection i, S_i is the average packet size in bytes for connection i, and TI_i is the average inter-packet arrival time in milliseconds for connection i. In this paper, these connections could be either downlink from a base station to a mobile station or uplink from a mobile station to a base station. To schedule downlink packets, the proposed algorithms should be implemented on base stations. On the other hand, the proposed mechanisms have to be realized on both base stations and mobile stations if the proposed methods are applied to the uplink packet scheduler. The base station can know the resource requirements of all mobile stations by negotiations in advanced or bandwidth requests from the mobile stations. Thus, the base station scheduler can determine the uplink packet schedule according to the proposed algorithms, and provides transmission opportunities to mobile stations. Then, mobile stations transmit uplink packets through the given OFDM frames. Without loss of generality, this study considers the above-mentioned QoS parameters to present the basic idea behind the proposed scheduling schemes. Other parameters such as delay jitters can be also

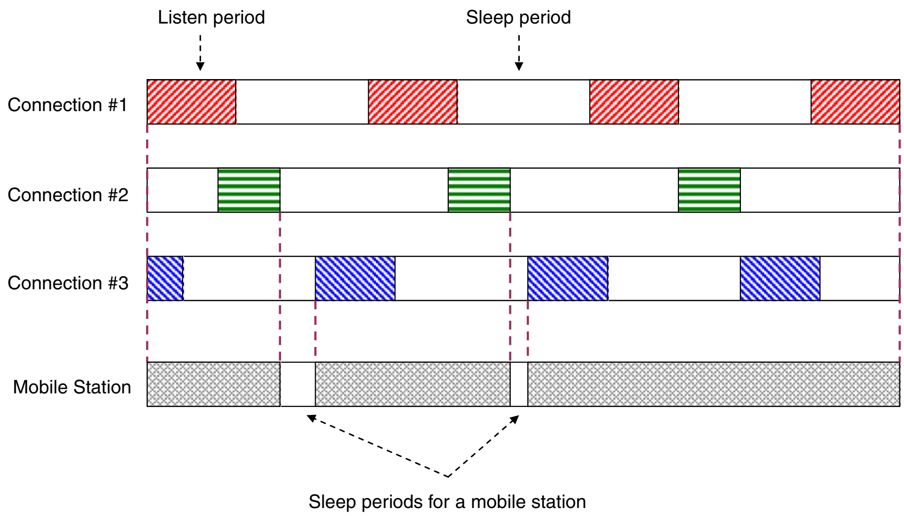

Fig. 2. Sleep periods for a mobile station with three connections.

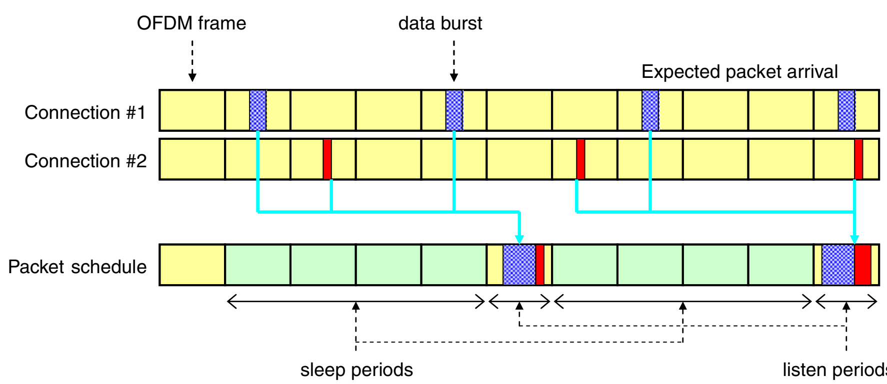

Fig. 3. Periodic on–off scheme (PS).

Table 1
Notations and their descriptions

Notation	Description
Q_i	The QoS requirement of connection *i*. It is denoted as $Q_i\{S_i, TI_i, D_i\}$
S_i	The average packet size in bytes for connection *i*
TI_i	The average inter-packet arrival time in milliseconds for connection *i*
D_i	The delay constraint in milliseconds for connection *i*
T_{frame}	The duration of an OFDM frame in milliseconds
B_{frame}	The maximal capacity of data in bytes that a base station can transmit in an OFDM frame
N_S	The number of OFDM frames in a sleep period
N_A	The number of OFDM frames in a listen period
P_S	The power consumption of a mobile station in Watt during sleep mode
P_A	The power consumption of a mobile station in Watt during listen mode
P_{avg}	The average power consumption of a mobile station in Watt
$B_{j,i}$	The amount of data in bytes which are requested by connection *i* in the *j*th OFDM frame
B_k	The amount of data in bytes which has been allocated by the base station in the *k*th OFDM frame
$\text{Block\_Set}_m(B_{j,i})$	A set of feasible scheduling frames of mobile station *m* for $B_{j,i}$

specified as the QoS of a connection and taken into account in the presented approaches.

To satisfy the QoS requirements of the connections on a mobile station, both bandwidth and delay constraints specified by all connections need to be considered. For the bandwidth constraint, since a mobile station cannot transmit and receive packets during a sleep period, the total amount of packets that a mobile station can transmit and receive during a listen period must be large enough to provide the needs for all connections during the listen and sleep period. For the delay constraint, the length of a sleep period must not exceed delay requirements of all connections. Assume the duration of an OFDM frame is T_{frame} milliseconds, and a base station can supply the maximal resources, say B_{frame} bytes, in an OFDM frame to the mobile station. The relationship between the number of OFDM frames in a sleep period, say N_S, and the number of OFDM frames in a listen period, say N_A, for the mobile station can be derived. First, N_S and N_A must satisfy the bandwidth constraint. That is:

$$N_S \geqslant 0,\ N_A \geqslant 0,\ N_S + N_A \geqslant 1,$$
$$\sum_{i=1}^{N}\left(S_i \times \left\lceil \frac{(N_S + N_A) \times T_{frame}}{TI_i} \right\rceil \right) \leqslant N_A \times B_{frame}. \quad (1)$$

Eq. (1) presents the maximal feasible amount of data that a mobile station needs to transmit and receive during a sleep cycle, i.e. $N_S + N_A$ OFDM frames, must be less than the total amount of data that the base station can supply a mobile station during $N_A \times B_{frame}$ frames for all N connections. Second, N_S and N_A must also satisfy the delay constraint. That is: $(N_S + N_A) \times T_{frame} \leqslant D_i,\ \forall i,$ which means the maximal delay between any two consecutive packets for any connection must be smaller than its delay requirement. Assume that the power consumption of a mobile station in sleep mode is P_S, the power consumption of a mobile station during listen mode is P_A, and the average power consumption of the mobile station is: $P_{avg} = \frac{P_S \times N_S + P_A \times N_A}{N_S + N_A}$. The equation can be rewritten as:

$$P_{avg} = \frac{P_S \times (N_S + N_A) + (P_A - P_S) \times N_A}{N_S + N_A} = P_S + \frac{(P_A - P_S)}{\frac{N_S}{N_A} + 1}.$$

Since P_S is less than P_A, it can be seen from the above equation that the maximal $\frac{N_S}{N_A}$ achieves the minimal power consumption of a mobile station. By applying the integer programming technique to the above equations, the optimal N_S and N_A under the constraints can be derived. Fig. 4 shows an example. The non-negative integer points in the region of the feasible solutions in Fig. 4 are possible pairs of N_S and N_A which guarantee the QoS requirements. Then, the solution of N_S and N_A which has maximal $\frac{N_S}{N_A}$ is the most energy-efficient solution. After N_S and N_A are determined, a mobile station can perform the sleep-mode operations based on the parameters.

3.2. Aperiodic on–off scheme (AS)

Since the PS requires a mobile station to always sleep for a fixed period and listen for another fixed period in a round-robin basis, a mobile station might have to stay awake in some frames in the listen period even there is no packet to send or to receive. Thus, an aperiodic on–off scheduling scheme (AS) is further proposed to determine if a mobile station should go to sleep or not in a frame basis. In other words, the AS tries to schedule the packet transmission in the minimal number of OFDM frames without violating the QoSs of all connections. The length of sleep and listen periods are variable.

While a new connection on a mobile station is initiated or any existing connection is released, the AS on a base station is activated to schedule or re-schedule resources in the following frames for the mobile station. First, the AS sorts all connections on a mobile station based on their delay requirements, and schedules these connections with tight delay requirements first. The reason to schedule connections with tight delay requirements first is that packets of these connections need to be sent or received within a small time window. The scheduler has to consider these packets first in order not to violate their QoSs. Conversely, for packets that could tolerate more delays, the scheduler can find more feasible OFDM frames to schedule the packets without violating the delay requirements. After the scheduler decides the scheduling priorities of connections, the packets from the first priority connection, e.g.

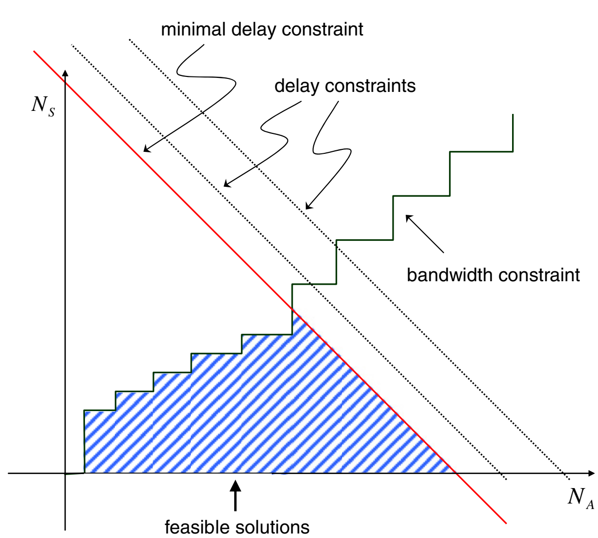

Fig. 4. Feasible solutions for the PS scheduling algorithm under constraints.

connection i, are scheduled. $B_{j,i}$ is defined as the amount of data in bytes that are requested by connection i in the jth OFDM frame. The AS tries to group this request with requests from other connections of the same mobile station together. Assume that the kth OFDM frame, where $k \geqslant j$, has been already scheduled B_k bytes data for other requests of the mobile station. The AS schedules $B_{j,i}$ into the kth frame if both the bandwidth and delay constraints can be satisfied. That is, $B_{j,i} \leqslant B_{frame} - B_k$, and $(k - j + 1) \times T_{frame} \leqslant D_i$. The kth frame can be any frame after the jth frame, but both the bandwidth and delay requirements must be satisfied. In the next paragraph, the determination of the frame k is discussed.

The AS has to schedule $B_{j,i}$ between the jth frame and the $\left(j + \left\lfloor \frac{D_i}{T_{frame}} \right\rfloor - 1\right)$th frame to satisfy the delay constraint of connect i. Block_Set$_{\mathrm{m}}(B_{j,i})$ is defined as a set of feasible scheduling frames for $B_{j,i}$, i.e. the frames between the jth frame and the $j + \left\lfloor \frac{D_i}{T_{frame}} \right\rfloor - 1$th frame which have available resources for mobile station m. To allocate radio resources in one or more frames from Block_Set$_{\mathrm{m}}(B_{j,i})$ to $B_{j,i}$, the AS follows the below steps. (1) The frames in Block_Set$_{\mathrm{m}}(B_{j,i})$ that have been already scheduled packets for the mobile station, called in-used frames for mobile station m, receive the highest priorities to be scheduled to $B_{j,i}$. That is because the AS aims to reduce the number of listen frames and increase the number of sleep frames. For a frame which is already scheduled packets, a mobile station cannot sleep. Therefore, in-used frames are assigned first if the resources of the in-used frames are still available to accommodate $B_{j,i}$. (2) If there are two or more in-used frames in the set, the AS schedules the packet to the first feasible in-used frame. That is, the AS picks up the qth frame rather than the pth frame if the frame q and p are both in used and $q \leqslant p$. The reason that the AS should schedule the packet as early as possible is to reduce the delays. Moreover, the pth frame has opportunities to be scheduled for sending or receiving packets whose deadlines are after the qth frame, but the qth frame cannot be scheduled for serving these packets. (3) If the AS cannot find in-used frames from the set or the in-used frames are all fully occupied, un-used frames are scheduled. The AS picks up the last un-used frames from the set to serve $B_{j,i}$. The last un-used frame is selected is because once a frame is scheduled to transmit or receive packets, the frame becomes an in-used frame and the mobile station is unable to sleep. If a latter frame can be selected, this frame potentially gains more opportunities to serve other packets in the following OFDM frames. Therefore, the last un-used frame is assigned. After the above steps, the AS schedules $B_{j,i}$ to the selected frame.

Fig. 5 illustrates an example of a packet schedule by applying the proposed AS. In Fig. 5, the AS selects a suitable frame to schedule the jth packet. There are five feasible frames for the packet under the delay constraint. The first frame and the fourth frame have higher priorities than others because they are in-used frame, and the earliest in-used frame gets the highest priority. For these un-used frames, they have lower priorities than the in-used frames, and the latter un-used frames receives higher priority than former un-used frames. Based on this priority assignment policy, the priorities for the five frames are 1, 5, 4, 2, and 3. Finally, the fourth frame is chosen since the first frame is fully occupied. The AS tries to schedule the packets in the minimal number of frames and can also guarantee the QoSs of all connections. Different from the PS which defines the fixed length on and off periods, the AS needs to exchange the on–off schedule information for each frame between a mobile station and a base station. The exchange can be done once after the connection is established or can be performed between a mobile station and a base station periodically. The proposed AS approach can be

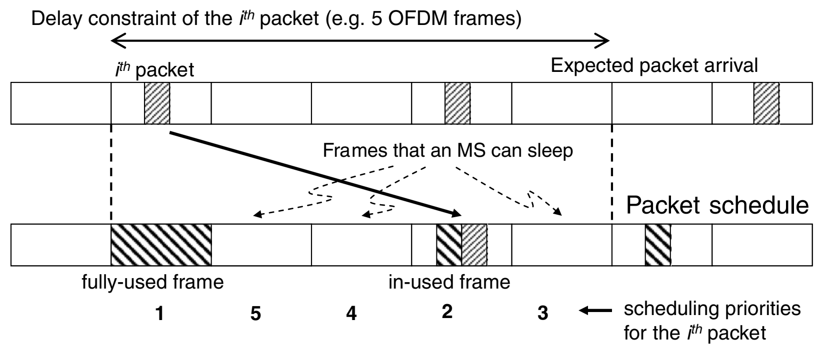

Fig. 5. Aperiodic on–off scheme (AS).

implemented by utilizing the type-three power-saving class defined in the IEEE 802.16e.

3.3. Periodic on–off scheme and aperiodic on–off scheme for multiple mobile stations

The proposed scheduling schemes presented in the previous paragraphs consider multiple connections on a single mobile station, but usually a base station serves multiple mobile stations. Therefore, to apply the PS and AS to a base station for scheduling multiple mobile stations should be further discussed. The proposed schemes both assume that a base station supplies the amount of resources, say B_{frame}, in an OFDM frame to a mobile station. Given the amount of resources allocated to a mobile station, the base station determines the parameters of the sleep operations for the mobile station. According to Eq. (1), the maximal amount of resources in an OFDM frame that a base station can supply a mobile station significantly influence the length of a sleep and listen period of the mobile station for the PS approach. Also, for the AS approach, the listen and sleep frames are derived from $\text{Block\_Set}_\text{m}(B_{i,j})$ which is obtained based on the available resources that a base station can offer the mobile station. Hence, various resource allocation policies that assign different amount of resources to mobile stations result in different power consumption for each mobile station. The optimal solution is to properly allocate resources of an OFDM frame to all mobile stations so that the total sleep periods of all mobile stations are maximal. Unfortunately, the optimal solutions for both PS and AS are difficult to implement since that the available resources of an OFDM frame change while a new connection is established or an existing connection is released. Then, the resource allocations in an OFDM frame to all mobile stations and the sleep-mode operations of all mobile stations need to be rearranged while the current available resources are updated. The rearrangement of the resources and updates of sleep-mode operations for all mobile stations introduce management overheads. Therefore, a greedy strategy that is easy to implement is proposed to cooperate with the PS and AS for a base station to determine the resource allocations for mobile stations.

A base station performs the admission control when a mobile station requests a new connection. Once a base station receives the connection request, it allocates all of its available resources to this new connection request and determines the lengths of the sleep and listen periods for the mobile station according to the PS or AS approach. If a mobile station releases a connection or establishes a new connection, the base station re-schedules the sleep-mode operations for that mobile station based on the current available resources. On the other hand, if a mobile station does not establish or release connections, a base station does not re-schedule that mobile station even when the base station obtains new resources released by other mobile stations. The simulations in the next section demonstrate that the proposed PS and AS approach with the greedy resource allocation policy also efficiently improve the sleep periods for multiple mobile stations.

4. Simulation environment and analysis of simulation results

4.1. Simulation environment

An IEEE 802.16e MAC-layer simulator written in C++ was developed to evaluate the performance improvement by employing the proposed schemes. The simulations in this study use WirelessHUMAN(-OFDM) profile, i.e. ProfP3_10 defined in [1], and 10 MHz channel, 5 ms frame length, and 64-QAM with 3/4 coding rate are assumed. The changes of channel conditions and the adaptive modulation and coding (AMC) are not considered in the simulation. In the simulations, a mobile station could be a standard mobile station which establishes multiple real-time connections such as a multi-party conference call, or a mobile relay router which bridges multiple real-time connections for other mobile stations. These connections on a mobile station have different QoS requirements and real-time characteristics. Four types of real-time connections with different QoS requirements are defined and their descriptions are summarized in Table 2. A_L and A_H denote a low-bite audio connection and a high bit-rate audio connection individually. V_L denotes a low bit-rate video connection, and V_H denotes a high bit-rate video connection. The delay constraints for audio and video connections are set to 50 ms and 100 ms, respectively. The audio connections are assumed constant-bit-rate (CBR), and they are classified as unsolicited grant service (UGS) connections. On the other hand, the video connections are variable-bit-rate (VBR). The sizes of video packets are generated by a Gamma distribution suggested by [11]. The video connections are classified as real-time polling service (rtPS) connections. Also, IP/UDP/RTP and MAC headers and the signaling overheads to exchange the control messages of sleep-mode operations for the proposed PS and AS are all taken into considerations in the simulation program.

4.2. Analysis of simulation results

First, the power consumption of a mobile station by employing three different scheduling schemes, i.e. the proposed PS and AS, and the traditional approach, are investigated. The traditional

Table 2
Four real-time connections with different characteristics

Connection	Description
A_L	Low bit-rate audio connection using G.723.1 codec at 6.4 Kbps and 30 ms frame length
A_H	High bit-rate audio connection using G.711 codec at 64 Kbps and 20 ms frame length
V_L	Low bit-rate video connection at 100 Kbps and 30 frames per second
V_H	High bit-rate video connection at 300 Kbps and 30 frames per second

approach implies that each connection associates with its preferred power-saving class and parameters which minimize the packet delay and power consumption for that single connection. In the first simulation, the percentage of sleep periods of a mobile station is evaluated while a base station serves different number of mobile stations simultaneously. In this simulation, there are three different configurations. In the first configuration, each mobile station establishes two high bit-rate audio and two high bit-rate video connections, denoted as $(A_H + V_H) \times 2$. In the second configuration, each mobile station establishes two low bit-rate audio and two low bit-rate video connections, denoted as $(A_L + V_L) \times 2$. In the third configuration, each mobile station establishes a pair of low bit-rate audio and video connections, and a pair of one high bit-rate audio and video connections, denoted as $A_L + V_L + A_H + V_H$. Arrival time of an audio or video packet to the destination that exceeds its delay constraint is counted as a dropped packet. In the simulation, the three packet scheduling schemes all guarantee the packet drop rate to be less than 1%.

Fig. 6 shows the percentage of sleep periods for a mobile station under different base station loadings, i.e. under that a base station serves different number of mobile stations simultaneously. The higher percentage of sleep periods a mobile station stays, the less energy the mobile station consumes. According to the simulation results, the maximal numbers of mobile stations served by a base station for the above three configurations, i.e. each mobile station with $(A_H + V_H) \times 2$, $(A_L + V_L) \times 2$, and $A_L + V_L + A_H + V_H$ connections, are around 30, 42 and 35, respectively, without violating the packet drop rate and packet delay constraints. The maximal numbers of mobile stations served by a base station are dominated by the radio resources consumed by mobile stations, and the consumed radio resources for the above three configurations are different.

Regarding of the percentage of sleep periods of a mobile station, the simulation results reveal that the AS achieves the best performance among three approaches. By employing the AS, a mobile station always can have more than 75% sleep periods, about 50–80% improvement over the traditional approach. While the system loading is not heavy, a mobile station can have 90% sleep periods, i.e. 80–120% improvement over the traditional approach. The reason is that the AS minimizes the number of listen frames and also guarantees the QoS requirements of the connections. On the other hand, the percentage of sleep periods of a mobile station can be increased about 80–120% by employing the proposed PS while a base station serves few mobile stations. While the number of serving mobile stations for a base station increases, the improvements of the sleep periods for the PS approach decreases. This is because while a base station suffers from moderate or heavy loading, it is difficult for a base station to find sufficient radio resources in few consecutive frames to accommodate a mobile station to perform regular on–off transmissions. Although the improvement of the percentage of sleep periods for a mobile station by applying the PS decreases while a base station serves more mobile stations, a mobile station always gain more sleep time than that by applying the traditional approach.

Fig. 7 shows the average packet delay for a mobile station under different base station loadings. The figure reveals that the traditional approach always achieves the lowest packet delay since it processes packets immediately while packets arrive. The average packet delays for a mobile station by employing the proposed PS are higher than that by employing the traditional approach, but are less than that by employing the AS. This is because the AS may buffer the packet to the maximal delay constraint in order to gain more sleep time for a mobile station. Although the delays by employing the proposed PS and AS increase, the QoS requirements of the connections are still satisfied.

By comparing Figs. 6 and 7, it can be seen that the AS achieves better energy efficiency than the other approaches

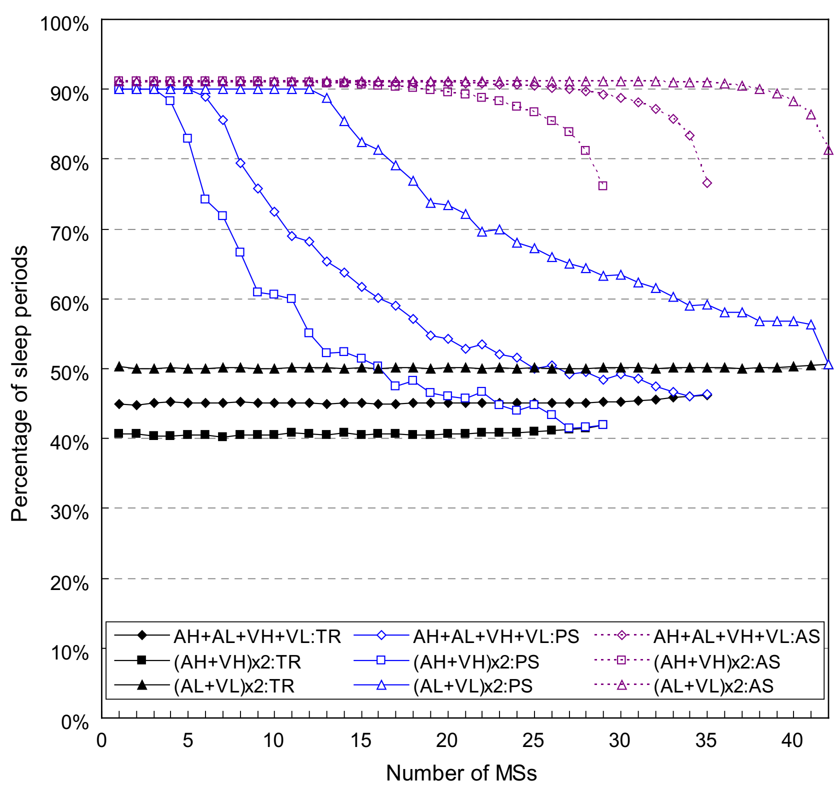

Fig. 6. Percentage of sleep periods for a mobile station with four real-time connections by employing different scheduling schemes (TR: traditional scheme, PS: periodic on–off scheme, AS: aperiodic on–off scheme).

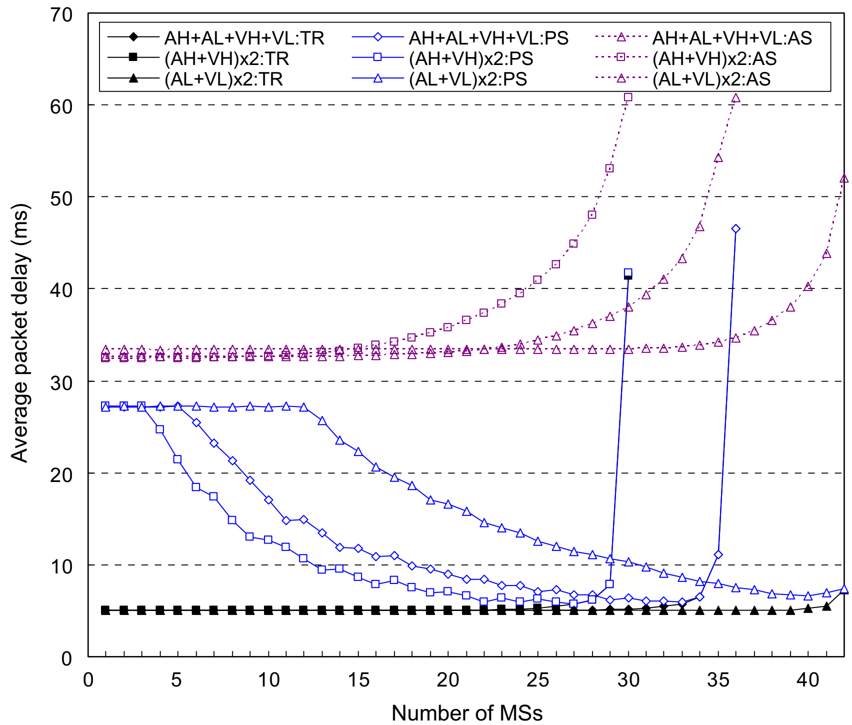

Fig. 7. Average packet delay for a mobile station with four real-time connections by employing different scheduling schemes (TR: traditional scheme, PS: periodic on–off scheme, AS: aperiodic on–off scheme).

under the same packet delay constraints. However, the average packet delays of the AS increase while the loading of base stations grows. It reveals that the AS cannot benefit from buffering packets while the system has insufficient radio resources. On the other hand, the average packet delays and the percentage of sleep periods for the PS approach both decrease while the loading of the base stations increases. With limited radio resources, the PS must increase the length of listen periods to satisfy the QoS requirements of the real-time connections. Therefore, both average packet delays and energy efficiencies decrease. For a heavy loading system, PS performs similar to the traditional scheme since the radio resources of OFDM frames are insufficient to offer more bandwidth to mobile stations and accommodate more sleeping OFDM frames.

In the next simulation, the percentage of sleep periods of a mobile station while a mobile station has more real-time connections is evaluated. Obviously, if a mobile station establishes more real-time connections, the number of mobile stations that a base station can serve reduces. Also, in this case, a mobile station has more real-time packets to transmit or receive, and a mobile station must stay awake in more frames to process the packet transmissions. Figs. 8 and 9 show the similar results as those in Figs. 6 and 7, but each mobile station establishes four audio connections and four video connections in this simulation. The real-time connections for a mobile station are four high bit-rate audio and four high bit-rate video connections denoted as $(A_H + V_H) \times 4$, four low bit-rate audio and four low bit-rate video connections denoted as $(A_L + V_L) \times 4$, and eight connections that include four different types denoted as $(A_L + V_L + A_H + V_H) \times 2$. As can be seen from the figure, the maximal number of serving mobile stations for a base station decreases. Also, the percentage of sleep periods of a mobile station decreases to 15–20% for the traditional approach. Fig. 8 shows that a mobile station still can gain 75–90% sleep time by applying the proposed AS. While a base station serves 10 mobile stations and implements the PS, each mobile station can have 50% more sleep time than that by applying the traditional approach. Fig. 9 shows that the traditional approach always achieves the lowest packet delay. The average packet delay of a mobile station by applying the AS is more than that by applying the other two schemes.

Finally, the percentages of sleep periods of a mobile station under different delay constraints that real-time connection can tolerate are investigated. Fig. 10 shows the percentage of sleep periods of a mobile station by applying the AS with different delay constraints. In this simulation, a base station is assumed to have a moderate system loading. The number of serving mobile stations is 20 for the situation that each mobile station establishes two audio and two video connections, and the number of serving mobile stations is set to 10 for the situation that each mobile station establishes four audio and four video connections. The delay constrains for both audio and video connections vary from zero OFDM frame, that is 0 ms, to 10 OFDM frame, that is 50 ms. Notably, if the delay constraint is set to zero frame, the AS must schedule all packets immediately without introducing any delay. Therefore, the AS with 0 ms delay constraint for real-time connections becomes the traditional approach. Fig. 10 shows that by applying the traditional approach, i.e. delay constraint is zero, the percentage of sleep periods for a mobile station with eight and four real-time connections is about 15–25% and 40–50%, respectively. While the delay constraint for both audio and video connections is set to one OFDM frame, i.e. 5 ms, the AS can improve the percentage of sleep periods of a mobile station by 15–40%. The percentage of sleep periods increases to about 70–90% if the delay constraints for real-time connections are set to two OFDM to ten OFDM frames. If the real-time connections can tolerate more delays, loose delay constraints can be set and more percentage of sleep periods for a mobile station can be obtained. Although the percentage of sleep periods increases while loose delay constraints are applied, the improvement becomes saturated while the delay constraint is larger than eight OFDM frames. In this situation, the bandwidth constraint becomes the

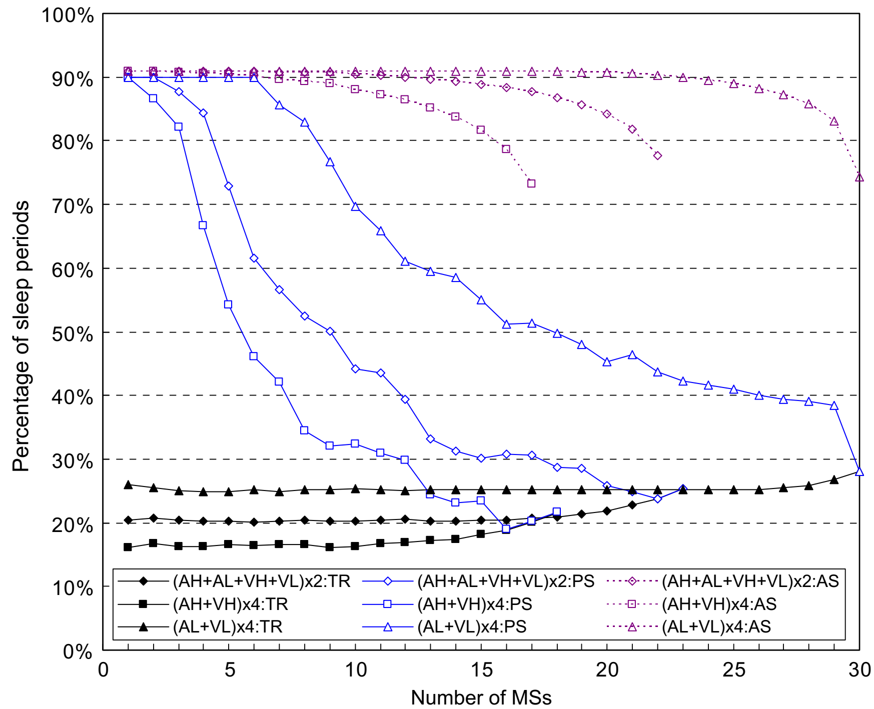

Fig. 8. Percentage of sleep periods for a mobile station with eight real-time connections by employing different scheduling schemes (TR: traditional scheme, PS: periodic on–off scheme, AS: aperiodic on–off scheme).

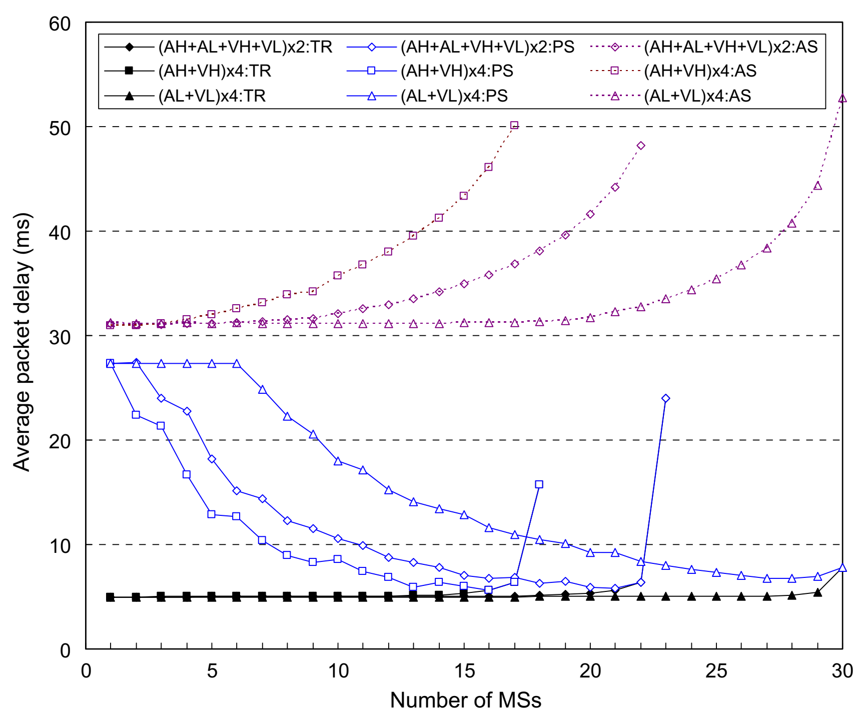

Fig. 9. Average packet delay for a mobile station with eight real-time connections by employing different scheduling schemes (TR: traditional scheme, PS: periodic on–off scheme, AS: aperiodic on–off scheme).

bottleneck and a mobile station needs to stay awake to transfer and receive the data in the minimal number of OFDM frames in order not to violate the bandwidth constraints for these real-time connections.

On the other hand, Fig. 11 shows the average packet delay of the AS with different delay constraints. As can be expected, the average delay linearly increases by the delay constraints. The average delay for a real-time connection and the improvement of the sleep time are tradeoff for a base station that applies to the AS approach. Even the AS is applied to a system with tight delay constraints, the AS still gains much more sleep time than that by employing the traditional approach.

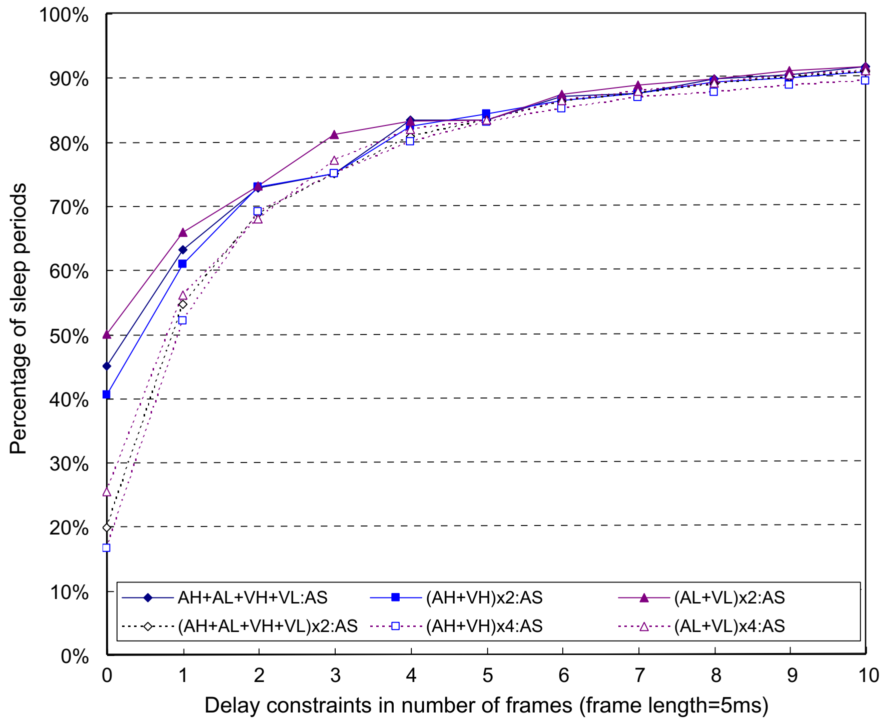

Fig. 10. Percentage of sleep periods for a mobile station while the AS with different delay constraints is applied.

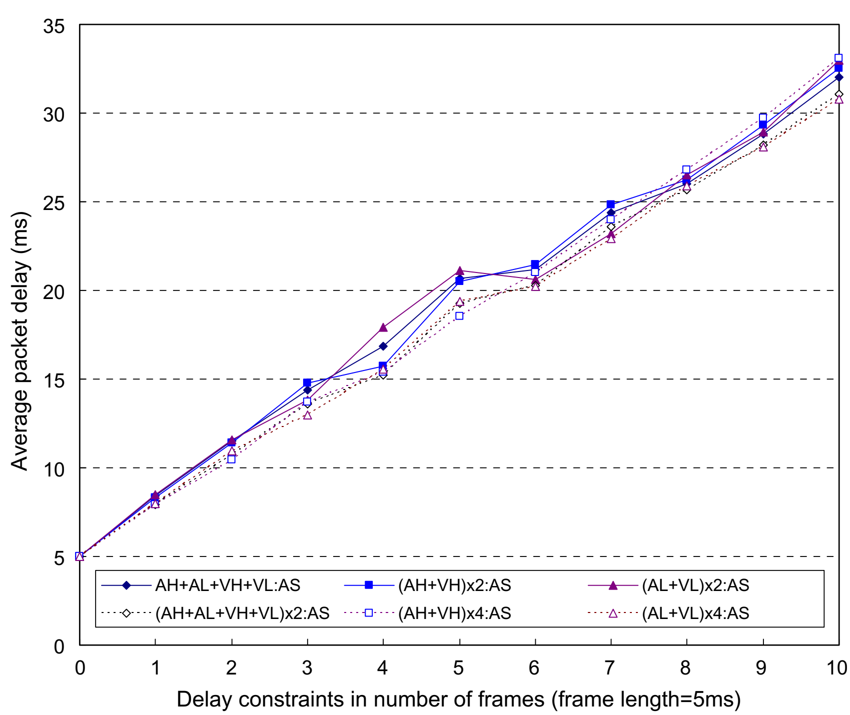

Fig. 11. Average packet delay for a mobile station while the AS with different delay constraints is applied.

5. Conclusions

Two energy-efficient packet scheduling schemes for the IEEE 802.16e were proposed. The proposed periodic on–off scheme (PS) that allows a mobile station to sleep and listen for fixed periods in a round-robin basis is easy to implement. The aperiodic on–off scheduling scheme (AS) that determines if a mobile station should go to sleep or not in a frame basis is further proposed to maximize the length of sleep periods of a mobile station. The implementation of the AS is more complicated than that of the PS since the AS has to maintain the status of each frame for every mobile station and schedules the packet transmission frame by frame. Although, the AS introduces more delay than the PS scheme, the delays by employing the AS are still controlled within the constraints. Simulation results demonstrate that the PS approach introduces few packet delays but increases about 80–120% sleep periods for a mobile station than the traditional approach while a base station serves few mobile stations. On the other hand, a mobile employing the proposed AS gains 15–120% more sleep periods than that using the traditional approach depending on the delay constraints and the number of real-time connections on the mobile station. The proposed schemes minimize the power consumption of mobile stations with multiple real-time connections without violating the QoSs of real-time connections.

Acknowledgement

The authors would like to thank the ZyXEL Communication Corp. for financially supporting this research.

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, IEEE Std 802.16-2004 (revision of IEEE Std 802.16-2001), October 2004.
[2] IEEE Std 802.16e-2005, IEEE standard for local and metropolitan area networks part 16: air interface for fixed and mobile broadband wireless access systems, IEEE Std 802.16e-2005, February 2006.
[3] K. Wongthavarawat, A. Ganz, Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems, in: IEEE Military Communications Conference, 2003.
[4] A. Ghosh, D.R. Wolter, J.G. Andrews, R. Chen, Broadband wireless access with WiMAX/802.16: current performance benchmarks and future potential, IEEE Communications Magazine (2005).
[5] D.-H. Cho, J.-H. Song, M.-S. Kim, K.-J. Han, Performance analysis of the IEEE 802.16 wireless metropolitan area network, IEEE Distributed Frameworks for Multimedia Applications, 2005.
[6] J.-B. Seo, S.-Q. Lee, N.-H. Park, H.-W. Lee, C.-H. Cho, Performance analysis of sleep mode operation in IEEE802.16e, in: IEEE Vehicular Technology Conference, 2004.
[7] Y. Chen, N. Smavatkul, S. Emeott, Power management for VoIP over IEEE 802.11 WLAN, in: IEEE Wireless Communications and Networking Conference, March 2004.
[8] R. Krashinsky, H. Balakrishnan, Minimizing energy for wireless web access with bounded slowdown, in: The 8th International Conference on Mobile Computing and Networking, 2002.
[9] G. Hair, J. Chou, T. Madejski, K. Perycz, D. Putzolu, J. Sydir, IEEE 802.16 medium access control and service provisioning, Intel Technology Journal 8 (3) (2004) 213–228.
[10] A.E. Xhafa, S. Kangude, X. Lu, MAC performance of IEEE 802.16e, in: IEEE Vehicular Technology Conference, vol. 1, September 2005, pp. 685–689.
[11] A. Drigas, S. Kouremenos, Y. Bakopoulos, V. Loumos, A study of H.263 traffic modeling in multipoint videoconference sessions over IP networks, Computer Communications (2005).
[12] K. Wongthavarawat, A. Ganz, Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems, International Journal of Communication Systems 16 (1) (2003) 81–96.
[13] Y. Xiao, Energy saving mechanism in the IEEE 802.16e wireless MAN, IEEE Communication Letters 9 (7) (2005) 595–597.
[14] Y. Zhang, M. Fujise, Energy management in the IEEE 802.16e MAC, IEEE Communication Letters 10 (4) (2006) 311–313.
[15] J. Shi, G. Fang, Y. Sun, J. Zhou, Z. Li, E. Dutkiewicz, Improving mobile station energy efficiency in IEEE 802.16e WMAN by burst scheduling, in: IEEE Global Telecommunications Conference, November 2006.
[16] J.-R. Lee, D.-H. Cho, Performance evaluation of energy-saving mechanism based on probabilistic sleep interval decision algorithm in IEEE 802.16e, IEEE Transactions on Vehicular Technology 56 (4) (2007) 1773–1780.
[17] S. Chandra, A. Vahdat, Application-specific network management for energy-aware streaming of popular multimedia formats, in: Proceedings USENIX Annual Technical Conference, 2002, pp. 329–342.
[18] D. Bertozzi, L. Benini, B. Ricco, Power aware network interface management for streaming multimedia, in: IEEE Wireless Communications and Networking Conference, 2002, pp. 926–930.