RESEARCH   **Open Access**

Energy-saving centric uplink scheduling scheme for broadband wireless access networks

Yen-Wen Chen[1], Yen-Yin Chu[1*] and I-Hsuan Peng[2]

Abstract

This study proposes an energy-saving centric uplink scheduling (ESC-US) scheme to support efficient energy usage and satisfy the quality of service (QoS) requirements of Worldwide Interoperability for Microwave Access (WiMax) networks. The uplink resource allocation is different from that of the downlink resource allocation scheme because the uplink traffic is queued at the mobile station (MS) and the base station (BS) has no information regarding it without using a polling procedure. The considered resource scheduling schemes maximize the sleep efficiency and consider the QoS requirements of individual MSs. The proposed scheduling scheme in this study considers the delay budget of MSs with real-time connections and the required minimum reserved traffic rate (MRTR) of MSs with non-real time connections when maximizing sleep efficiency. Both scheduling schemes for the traffic of real-time polling services (rtPS) and non-real-time polling services (nrtPS) apply the 'just enough QoS' and 'sleep before transmission' (SbT) concepts to achieve this energy-saving centric objective. Exhaustive simulations were conducted to examine the performance of the proposed schemes. The simulation results show that both schemes guarantee the desired QoS and achieve superior energy-savings efficiencies compared to the conventional scheme.

Keywords: Bandwidth requests; Energy-saving; Minimum reserved traffic rate; Scheduling; WiMax

1 Introduction

The rapid progress of broadband wireless access technologies, such as Worldwide Interoperability for Microwave Access (WiMax) [1] and long-term evolution (LTE) [2], and power mobile devices stimulates the flourishing deployment of mobile internet services. However, the slow improvement in battery technologies [3] has led to an exponentially increasing gap between available and required battery capacities [4]. Therefore, conserving energy is a crucial factor for mobile devices in practical applications. Because wireless internet is a shared medium, device energy efficiency is affected not only by the layers that compose the point-to-point communication link but also the interaction between the links in the entire network [5]. Therefore, the efficient conservation of energy to achieve longer mobile station (MS) operation times is vital to the success of mobile internet services. Furthermore, most wireless access networks provide the quality of service (QoS)

mechanism to ensure that users receive the desired service qualities of various services, such as data, voice, video, and multimedia, through appropriate radio resource allocation and scheduling at the base station (BS). WiMax defines three types of power-saving classes (PSC) for various types of traffic for the purposes of energy saving. Type I of PSC is recommended for the connection of best effort (BE) and non-real-time variable rate (NRT-VR) services. Type II of PSC is recommended for the connection of unsolicited grant service (UGS) and real-time variable rate (RT-VR) services. Type III of PSC is recommended for multicast connections and management operations. PSC I and PSC II operate alternatively between an awake and sleep mode. The sleep mode includes two operational windows: sleep window and listen window. PSC III does not contain a listen window, and its sleep cycle consists of only one predefined sleeping interval. The MS can issue a sleep request (MOB_SLP-REQ) message to its serving BS for the approval of entering sleep mode. The BS transmits the sleep response (MOB_SLP-RSP) message to indicate whether the sleep request is accepted or rejected. Sleep parameters are

* Correspondence: yenyinchu@gmail.com
[1]Department of Communication Engineering, National Central University, Jhongli City, Taoyuan County 01, Taiwan
Full list of author information is available at the end of the article

provided if the request is accepted. The sleep parameters include the starting sleep time, initial sleep window size, final sleep window size, and the listen window size, which are adopted by an MS in its sleep mode. PSC I doubles its sleep window size if the BS does not inform it of the presence of data packets at its listen interval, but it is constrained by the final sleep window size. The sleep window of PSC II is consistently the same size because of the real-time requirement. The sleep management procedure defined in WiMax is generally an operational procedure that does not consider the energy-saving efficiency in real applications, especially for uplink traffic. The network side generally does not know whether mobile devices require the uplink radio resource; thus, the BS schedules the uplink radio resource based on the bandwidth request issued by mobile devices. In WiMax, an MS awakes and issues a bandwidth request (BR) message when it has traffic to send. It remains awake and waits for the uplink radio resource allocated by the BS. This situation may cause the MS to halt the sleeping mode and switch to the awake mode even when no sufficient radio resources exist for its uplink transmission. The BS is responsible for radio resource allocation and the MS may wake without transmitting any information if the BS does not allocate resources for it because of insufficient radio resources. The MS wastes its limited energy in this scenario. From an energy-saving viewpoint, an inefficient uplink scheduling scheme may cause the MS to halt the sleeping mode and switch to the awake mode even when insufficient radio resources exist for its uplink transmission. Conversely, an efficient sleep mechanism provides a longer sleep interval to reduce energy consumption and maintains QoS requirements. This study proposes the energy-saving centric uplink scheduling (ESC-US) scheme, which properly manages sleep periods to conserve the energy of MSs under the QoS constraint. In this approach, the BS schedules the traffic for a better sleeping arrangement when receiving the BR from the MS to achieve better sleep efficiency and reduce switching between the sleep and awake modes. Both real-time and non-real-time connections were considered in the proposed scheduling schemes to ensure the desired QoS of each user. The proposed ESC-US scheme applies the 'sleep before transmission' (SbT) concept using hybrid sleep arrangement policies for connections with different services to achieve better radio resource usage in a heuristic manner.

The remainder of this study is arranged as follows: Section 2 presents reviews of related studies for this topic, Section 3 presents the proposed system model, Section 4 shows this study's approach and details the proposed ESC-US scheme, Section 5 provides the simulation results of the proposed schemes with discussions, and Section 6 offers a summary of the conclusions of this study.

2 Related works

Several recent studies have investigated the issue of energy consumption with various sleep mode operations. Xiao [6] proposed an analytical model for the PSC I to evaluate the effects of sleep parameters and traffic load by solely considering the downlink transmission. Moreover, both the downlink and uplink transmissions were considered and proposed analytical model to evaluate the energy management in the IEEE 802.16e [7]. In [8], the authors examined the performance in terms of the dropping probability and mean waiting time of packets in the queue of the BS by using an M/GI/1/N queuing model. In [9], the authors presented a semi-Markov chain queuing model to derive the performance and discuss the selection of proper sleep parameters. The performance of PSC II for VoIP traffic was examined in [10] for the given network delay model by using a simulation method that additionally provided a guideline to determine the sleep interval and discover the most energy-efficient sleep interval length which satisfied the given delay constraints from the results. The authors of [11] proposed a maximum unavailability interval (MUI) scheme for the given sleep parameters of PSC II to determine the start frame number for each of the connections with various classes by applying the Chinese reminder theorem. Conversely, in [12], the authors evaluated and compared the performance of PSC I and II. It proposed a sleep mode switching scheme based on analytical results to achieve optimal energy performance according to various traffic conditions.

Conversely, a number of studies are currently considering not only energy savings but also the QoS issues as well [13-17]. In [13], the authors proposed an energy-saving centric scheme for downlink scheduling. The authors presented a longest virtual burst first (LVBF) scheduling algorithm to improve the energy efficiency of MSs and meet the requirement of minimal data rates for the connection with PSC I [14]. A minimum wakeup time (MWT) scheduling scheme for PSC II was proposed in [15] to determine the maximal sleeping time for a single MS with constant bit rate (CBR) traffic and analyzed the upper bound for multiple MS power-saving scheduling with a MWT. Tsao and Chen [16] proposed two power-saving scheduling algorithms. The first algorithm is a periodic on-off scheme based on the idea of allowing an MS to sleep for a fixed interval and, subsequently, to listen for another fixed interval in a round-robin manner to determine feasible solutions for minimal energy consumption

under the delay constraint. The aperiodic on-off scheme was proposed to solve the energy-wasting problem of no traffic transmission during a fixed listen period, which adjusts the length of sleep and listen periods according to the traffic status. The energy-efficient architecture for two-level scheduling was proposed in [17]. The proposed architecture *sets* a priority order for the first-level scheduling by assuming that a mobile and fixed station coexist in the network and that mobile stations have a higher priority than fixed stations from an energy-saving viewpoint. Their study proposed a two-timer mechanism (i.e., empty timer and overloading timer) to dynamically adjust the state transition timing to achieve optimal energy efficiency. Most previous studies considered either sleeping period management for one type of connection or sleep synchronization of multiple connections, but resource scheduling for uplink traffic from an energy saving viewpoint was seldom considered. However, a BS must serve various types of connections for multiple MSs in a practical environment. Hence, the purpose of this study is to schedule sleeping and resources in a tradeoff manner so that an MS can conserve as much energy as possible and sufficiently provision its desired QoS.

3 The proposed ESC-US system model

WiMax adopts orthogonal frequency division multiple access (OFDMA) technique to achieve broadband wireless access. In addition to allocating the radio resource for downlink traffic, WiMax BS manages and arranges the radio resource for the uplink traffic of each MS. In order to obtain the required bandwidth of uplink, each MS needs to issue a BR to ask BS for the radio resource for uplink transmission. On receiving the bandwidth request from the MS, traditionally, the BS allocates the radio resource to the MS as soon as possible by considering the QoS requirements, such as the delay tolerance or minimum reserved traffic rate (MRTR). However, this approach may lead the BS to continue scheduling the radio resources when the uplink traffic is still being generated and not helpful for saving energy. A better method for efficient energy savings is to allow the MS to accumulate its uplink traffic for a period without activating its transmitter and receiver (i.e., in sleep mode) and immediately provide enough radio resource for the accumulated traffic when the MS is initiating transmission.

The system model is provided in Figure 1 and the proposed scheme focuses on two major algorithms to address the scheduling, which are the bandwidth allocation algorithm and the sleep cycle decision algorithm. The uplink traffic is generated and queued at the MS and waits for scheduling. When the MS has traffic waiting for transmission, it issues a BR message to the BS to request radio resources on the reception of a polling message from the BS. The bandwidth allocation algorithm properly allocates the radio resources
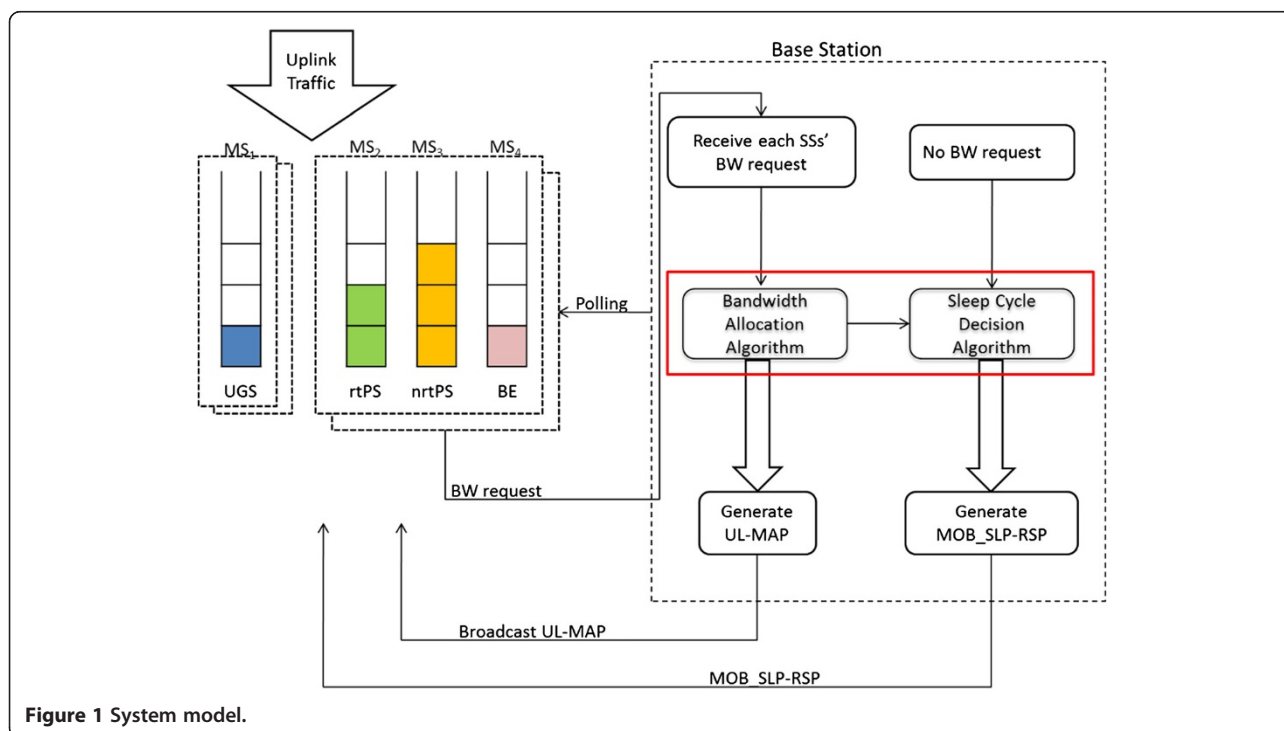

Figure 1 System model.

to the MS and decides the sleep period based on the allocation condition. The MS does not awake during the sleep period assigned by the BS despite arriving traffic during this period.

The descriptions and analysis of the proposed scheme is detailed in the following section, and the definitions of important parameters used during the analysis are listed in Table 1.

4 The proposed ESC-US scheme

An MS may generally issue a sleep request in a WiMax network to a BS when it has no data to send. The BS determines whether to accept this request according to the buffer status of the MS and the QoS. Thereafter, the sleep parameters are determined together with the scheduling of the downlink traffic. In the LTE network, the user equipment (i.e., the mobile device) automatically enters the discontinuous reception and transmission (DRX) mode (i.e., sleep mode) on the expiration of the inactivity timer that is specified by eNodeB (i.e., the BS). Therefore, the BS has explicit indications and information for arranging the sleep procedure of the MS during downlink scheduling. In [18], the energy-saving centric concept was also adopted for downlink scheduling. However, the operational scenario of the sleep arrangement with the uplink scheduling is quite different from the downlink scheduling. First, the BS has no information regarding the status of the uplink traffic, which is buffered by the MS. Second, the MS will not issue the sleep request when the uplink traffic is waiting for transmission.

To achieve this concept, the proposed ESC-US scheme schedules the radio resource for the MS in a reserving manner. Moreover, the SbT concept is applied to maximize the sleep efficiency when receiving the bandwidth request from the MS. Figure 2 shows that the BS reserves the radio resources of later frames without violating the QoS requirement when receiving the BR, and it requests the MS to enter a sleep period before its transmission at the later reserved frames in the SbT approach. The MS sends the bandwidth request, which accumulates the previously required bandwidth BR1 and the newly generated bandwidth requirement BR2 during the sleep period to the BS when it awakes. Thereafter, the BS allocates the denser radio resource to the MS to convey the uplink traffic. It shows that the proposed concept achieves better energy usage. The PSC III originally specifies the single sleeping window for the sleep management of the multicast connection. The sleep request is triggered by the BS or MS to inform the MS to sleep for only one window. Thereafter, the proposed scheme applies this characteristic for the sleep arrangement when performing uplink scheduling.

Based on the reservation concept, the BS begins allocating the bandwidth of the farthest frame backward to the nearest frame until the required bandwidth is satisfied. Moreover, the bandwidth of the farthest frame does not violate the QoS constraint of the MS. If the BS can allocate the bandwidth for the MS in the farther frames, then the BS will inform the MS to sleep until the frame with the reserved bandwidth arrives. In this scenario, the MS receives the sleep response (MOB_SLP-RSP) from the BS with the sleep duration when it issues the bandwidth request. Thereafter, the MS will immediately obtain the reserved radio resource according to the allocation specified in the uplink map (UL_MAP) message to transmit its traffic when it awakes. The purpose of the proposed algorithms is to maximize the sleep period and maintain the desired QoS. Figure 3 shows

Table 1 The definitions of important parameters of the proposed scheme

| Parameters	Descriptions
α	The predicted number of frames that the traffic generated prior to the BR issue
$BR_{m,n}$	The mth bandwidth request issued at the nth frame for a specific MS i
K_m^{rt}	Number of frames that BS will request the MS to sleep for the $BR_{m,n}$ issued by rtPS MS
B_m	The bandwidth that needs to be allocated in the newly generated frames
F_i	The average minimum traffic rate of each frame
M_i	The MRTR of the MS i, M_i indicates that the BS can calculate the average bandwidth that must be allocated in each frame F_i
N_i	BS can estimate the farthest frame of this allocation based on the desired bandwidth specified in $BR_{m,n}$
t_{frame}	The frame duration
β	BS can arrange the delay of the nrtPS traffic because of the sleep period
K_m^{nrt}	Number of frames that BS will request the MS to sleep for the $BR_{m,n}$ issued by nrtPS MS
P_x^{nrt}	Number of bits of preallocated nrtPS bandwidth from the xth frame
BW_{frame}	Number of bits that BS can allocate for MS in each frame
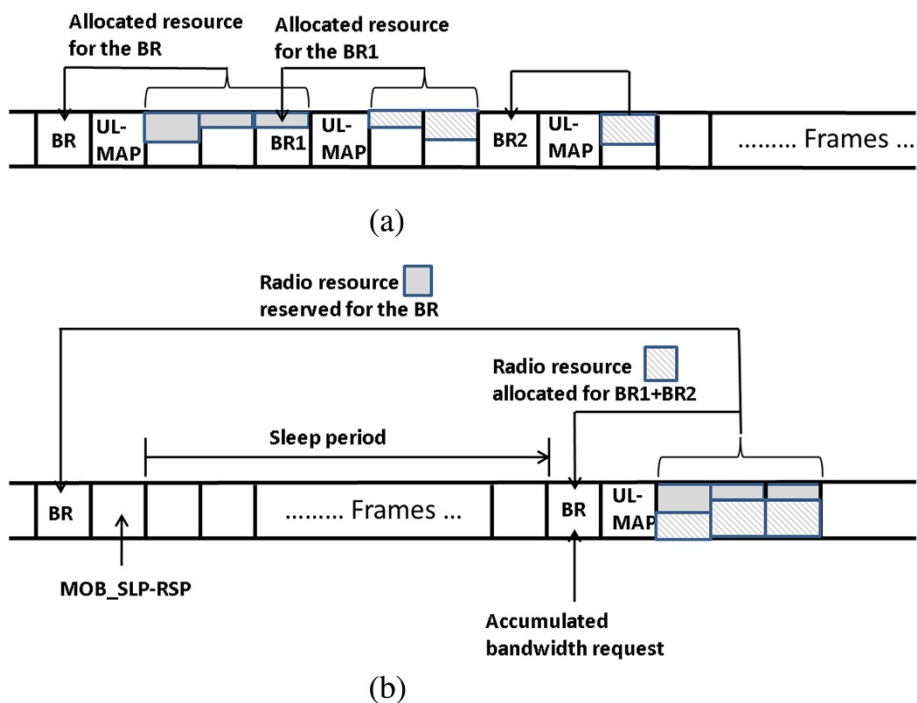
Figure 2 Traditional uplink traffic scheduling (a) and proposed reservation based uplink traffic scheduling (b).

the operational scenario between the MS and BS. This reservation-based approach allows the MS more sleeping time compared to the traditional scheme and can still maintain the QoS requirement. Moreover, the BS notifies the MS to enter a sleep period if it does not receive a bandwidth request from the MS.

When the MS awakens from a sleep period, it must issue a BR and wait for the UL_MAP for the resource allocation information if it has uplink traffic waiting for transmission. The bandwidth request issued by the MS can be categorized into the incremental based and aggregated based in the specification. The incremental-
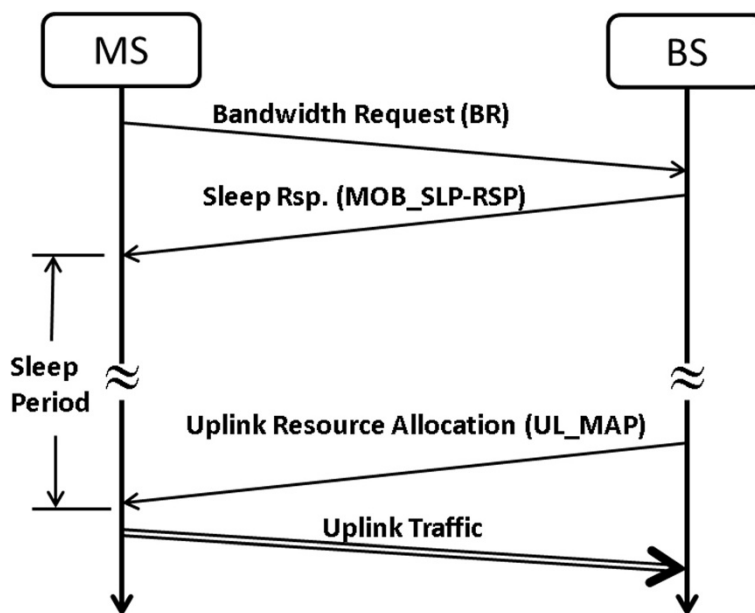

Figure 3 Operational scenario between the MS and BS.

based BR approach assumes that the BS remembers the unallocated bandwidth of the previously issued BR, and the current BR is the required bandwidth of the newly generated traffic; thus, the BS adds the quantity of requested bandwidth to its current perception regarding the bandwidth needs of the connection for the incremental-based approach. When the BS receives an aggregated bandwidth request, it replaces its perception of the bandwidth needs of the connection with the full quantity of requested bandwidth. In the incremental-based approach, the MS issues a BR at each frame whenever it has traffic to be sent. Although this approach provides real-time information regarding the BR to the BS, remaining awake at each frame is inefficient from an energy-saving viewpoint. Therefore, the aggregated-based approach is adopted in this study and the MS issues the aggregated-BR during a period. When the MS wakes from the sleep period, it issues a BR to the BS if any uplink traffic is queued for transmission; otherwise, it begins a new sleep period. An MS may request more bandwidth than specified in the previous BR because traffic may be generated during the sleep period; thus, the bandwidth requested by the MS may include the bandwidth specified in the previous BR. The bandwidth requested in the previous BR message should be prescheduled by the BS at that moment, and the MS originally transmits the scheduled traffic when it awakes. Because the MS remains awake to transmit data in the prescheduled (reserved) frames, the allocation of the latest generated traffic by the BS in those prescheduled frames is better from an energy-savings and resource-usage viewpoint if those frames have available bandwidth. Therefore, the BS initially allocates the available bandwidth of the reserved frames to the required

bandwidth of the MS on receiving the BR. If the residual bandwidth of the reserved frames cannot satisfy the required bandwidth, the BS seeks to schedule the unallocated bandwidth as far as possible so that the MS has a longer sleep period. The allocation scheme of the proposed ESC-US scheme is shown in Figure 4, which shows that the proposed scheme schedules the required bandwidth to use the residual bandwidth of the frames that the MS had initially been allocated for better resource usage. If those frames are insufficient to satisfy the required bandwidth, the BS begins allocating the bandwidth backward from the farthest frame, as shown in Figure 4. This also indicates that the MS wakes two frames prior to the frame that was allocated with the resource because the MS must send a BR message and await the resource allocation information through the UL_MAP message before transmitting the uplink traffic. For example, assume that the maximum tolerable delay of the MS i is Di frames. The MS issues a bandwidth request and the BS determines that the farthest tolerable frame for the bandwidth allocation is the $(n + D_i)$th frame, as shown in Figure 4, which means that the BS will allocate bandwidth for the MS no later than the next D_i frames. Thereafter, the BS can allocate the resource backward for the MS from the $(n + D_i)$th frame to the nearest frame until the required bandwidth is satisfied. Assume that the BS can allocate (reserve) the required bandwidth of the MS at the farthest three frames (i.e., the $(n + D_i)$th, $(n + D_i - 1)$th, and $(n + D_i - 2)$th frames of the current frame). Thereafter, the BS can request the MS to enter sleep mode until the $(n + D_i - 4)$th frame. The MS wakes to send its BR to the BS at the $(n + D_i - 4)$th frame, and the BS provides the radio resource allocation information in the UL-MAP at the

Figure 4 An example showing BS receiving BR at the current frame.
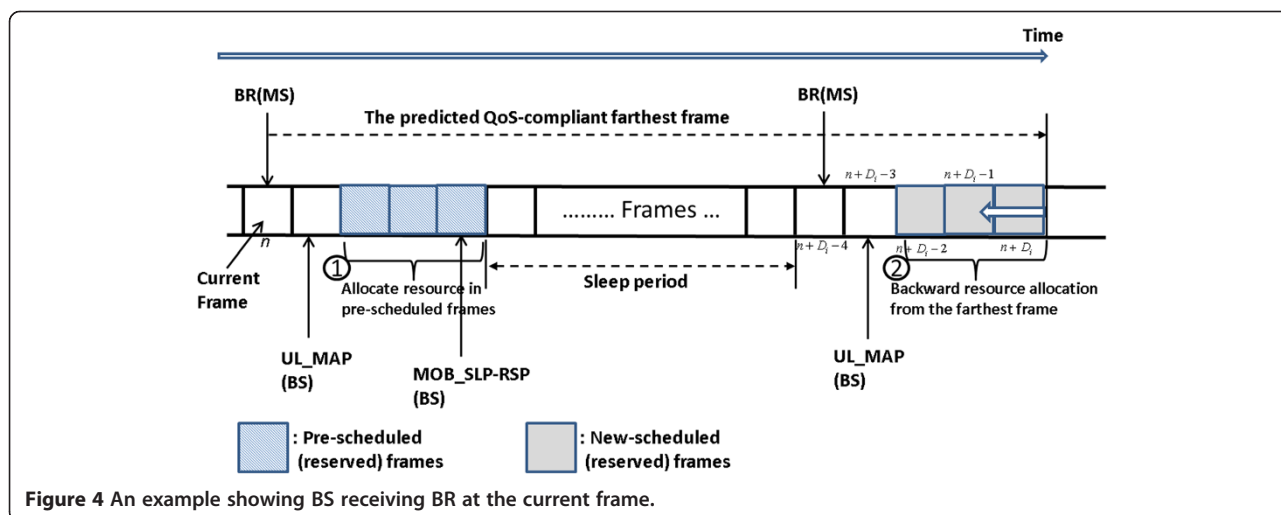
next $(n + D_i - 3)$th frame so that the MS can begin transmitting its uplink traffic at the next $(n + D_i - 2)$th frame of the current frame. For the example shown in Figure 4, when the BS receives the BR at the current frame, it initially allocates the residual bandwidth of the three prescheduled frames and allocates the bandwidth of the farthest three frames for the MS. Thereafter, the BS will provide the resource allocation information at the next frame of the current frame and informs the MS to enter sleep mode at the last prescheduled frame. In the proposed scheme, the resource allocation bases regarding the provisioning of the 'just enough' QoS and length of the sleep period on effective resource allocation. The BS may not allow the MS to sleep if the sleep period is shorter than a specific threshold because the frequent state change may result in a higher consumption of energy. If the MS transmits its data completely and has no queued data to send, it will not issue a BR when polled by the BS. In this scenario, the BS will request MS to enter sleep mode for the period of the tolerable delay.

The proposed scheme shall decide the farthest frame that can be allocated for the issued BR without violating the desired QoS. The real-time polling service (rtPS) and non-real-time polling service (nrtPS) have different QoS requirements. The delay tolerance is critical for rtPS connections, but the MRTR is a major parameter for nrtPS connections when fair resource usage is concerned. Therefore, the acceptable farthest frames of the rtPS connection and nrtPS connection are decided in a different manner.

The proposed ESC-US scheme adopts the reservation-based approach to allocate the radio resources for the MS to maximize the sleep period or minimize the idle period. The resource allocation scheme tends to support the just enough QoS requirement, and the SbT concept manages the sleep operation more efficiently to conserve more energy. For the rtPS connection, the proposed scheme uses the delay tolerance as the performance index for traffic scheduling. On receiving the BR from MS, if the MS has not been allocated with radio resources for the required bandwidth or the residual bandwidth of the pre-scheduled frames for the MS, they will be insufficient for the required bandwidth; thus, the BS will decide the farthest frame, which satisfies the maximum delay tolerance to backwardly allocate the radio resource. The MS may request more bandwidth than that specified in the previous BR because traffic may be generated during the sleep period. In this scenario, bandwidth needs to be allocated in addition to the newly generated traffic to satisfy the delay requirement. The BS can determine its delay tolerance as $D_i - \alpha$ in the same manner as the traffic generated during the sleep period. Assume

that the BS has arranged f frames to convey the required bandwidth issued by $\text{BR}_{m-1,x}$, and the MS will transmit its uplink data from the $(n + 2)$th frame to the $(n + 1 + f)$th frame when it awakes at the nth frame. Therefore, the difference of the bandwidths specified in $\text{BR}_{m-1,x}$ and $\text{BR}_{m,n}$ is the traffic generated during the sleep period after $\text{BR}_{m-1,x}$ was issued. If $\text{BR}_{m,n}$ equals $\text{BR}_{m-1,x}$, it means that no traffic was generated during the sleep period. Because the bandwidth specified in $\text{BR}_{m-1,x}$ is allocated, the BS is only required to schedule the newly generated traffic. The BS initially begins allocating the available radio resource to the required bandwidth of the newly generated traffic from the $(n + 2)$th frame to the $(n + 1 + f)$th frame and allocates the resource from the farthest frames if it is insufficient. Therefore, the BS will request the MS to sleep for the period of K_m^{rt} frames for the $\text{BR}_{m,n}$ issued by MS, which can be expressed as

$$K_m^{\text{rt}} = \text{BWD}(n + D_i - \alpha, B_m) - 2 - (n + 1 + f) \qquad (1)$$

where

$$B_m = \max\{0, [|\text{BR}_{m,n}| - |\text{BR}_{m-1,x}| - \text{Av}_i((n + 2), (n + 1 + f))]\} \qquad (2)$$

where $|\text{BR}_{m,n}|$ and $|\text{BR}_{m-1,x}|$ denote the required bandwidth specified in the $\text{BR}_{m,n}$ and $\text{BR}_{m-1,x}$ messages, respectively. The function of $\text{BWD}(x,y)$ allocates the bandwidth from the farthest frame x backward to the nearest frame for the required bandwidth y and returns the number of the nearest frame after allocation. It can be calculated by decreasing the bandwidth that can be allocated in prescheduled frames from the newly generated bandwidth, which is the difference between the bandwidths specified in the $\text{BR}_{m-1,x}$ and $\text{BR}_{m,n}$ messages. The function $\text{Av}_i(a, b)$ in (3) calculates the available bandwidth from the ath frame to the bth frame of MS i. The newly generated required bandwidth may be fully allocated in the prescheduled frames. Thereafter, B_m is zero. In this scenario, $\text{BWD}(n + D_i - \alpha, 0)$ equals $n + D_i - \alpha$, and it makes the sleep period k_m^{rt} equal to $(n + D_i - \alpha) - 2 - (n + 1 + f) = D_i - \alpha - 3 - f$ in (2), as shown in Figure 4.

For the nrtPS connection, the proposed scheme uses MRTR as the performance index to maximize the sleep efficiency. The MRTR specifies the minimum bandwidth that the BS provides for nrtPS traffic. Although the nrtPS does not specify the maximum delay tolerance, the required MRTR implicitly indicates the acceptable delay or the farthest frame for bandwidth allocation. In this scenario, the BS is on average assumed to provide the MRTR of the MS in each frame when it issues a BR to the BS. However, rather than

allocating even bandwidth to the MS in a frame-by-frame manner, the BS requests the MS to enter sleep mode and allocates the burst bandwidth to the MS when it awakes. The average minimum traffic rate of each frame F_i can be derived from the MRTR and the length of the frame duration is as follows:

$$F_i = M_i \times t_{\text{frame}} \tag{3}$$

and

$$N_i = \left|\text{BR}_{m,n}\right|/F_i \tag{4}$$

For example, if M_i is 1 Mbps and the frame duration is 10 ms, then F_i is 10 kbits. The value of B_m for nrtPS traffic is the same as (3) and the sleep period of the nrtPS connection K_m^{nrt} is similar to that of rtPS traffic as shown in Equation 6.

$$K_m^{\text{nrt}} = \text{BWD}(n + N_i - \beta, B_m) - 2 - (n + 1 + f) \tag{5}$$

For the nrtPS traffic, there is no delay constraint. Thus, the BS is not required to predict when the uplink was generated. The larger value of β is the shorter sleep period and the transmission delay the MS acquires. A negative β value indicates that the MS can acquire a much longer sleep period, but the MRTR requirement may not be satisfied if the MS has data to transmit.

The resource may be insufficient when the BS allocates the required bandwidth of the rtPS connection backward from the farthest frame to the nearest. In this scenario, the overflow traffic is discarded because the BS cannot satisfy the delay requirement of the rtPS traffic. The proposed scheme may alternatively apply the preemptive approach to allow the BS to reschedule the resource that is allocated to the nrtPS

traffic for the allocation of the rtPS traffic. Moreover, the preemption starts backwards from the farthest frame to the nearest, as shown in Figure 5. Following the preemption, the BS allocates the available resources after the farthest frame of the MS that issues the BR, as shown in Figure 5; thus, if the BS preempts P_x^{nrt} bits, the BS must allocate bandwidth for the P_x^{nrt} bits in the farther frames. Although the preemption alternative provides more bandwidth to satisfy the rtPS traffic, it may affect the nrtPS MS toward transmitting data when it awakes because its reserved bandwidth has been rescheduled to later frames. Because switching from one mode to another consumes energy, the threshold control approach can be applied to define the minimum sleep period to prevent the MS from frequently oscillating between the sleep and active modes; therefore, if the calculated sleep period (either K_m^{rt} or K_m^{nrt}) is shorter than the defined *SLEEP_THRESHOLD*, the BS will not allow the MS to enter sleep mode.

Algorithm 1 shows the procedure of the rtPS scheduling, and lines 1 to 29 state the allocation steps for the rtPS connection. In lines 3 to 5, the BS allocates the bandwidth for the requesting MS with the nearest frames to the previously reserved bandwidth for the MS. The BS allocates the radio resources from the farthest frame in lines 4 to 16 if no radio resources remain in the nearest frame that has pre-reserved radio resources for the MS. The BS initially chooses the frame in which a portion of the radio resources was pre-reserved for the MS during allocation. The BS arranges the MS to enter sleep mode if the number of consecutive frames (not reserved for the MS) exceeds the SLEEP_THRESHOLD after bandwidth scheduling. Lines 30 to 33 determine whether the MS can enter sleep mode and sets the sleep period as K_i^{rt} frames. Because the nrtPS traffic has no delay constraint, the rtPS connection can optionally preempt the previously reserved bandwidth for the nrtPS traffic to reduce the
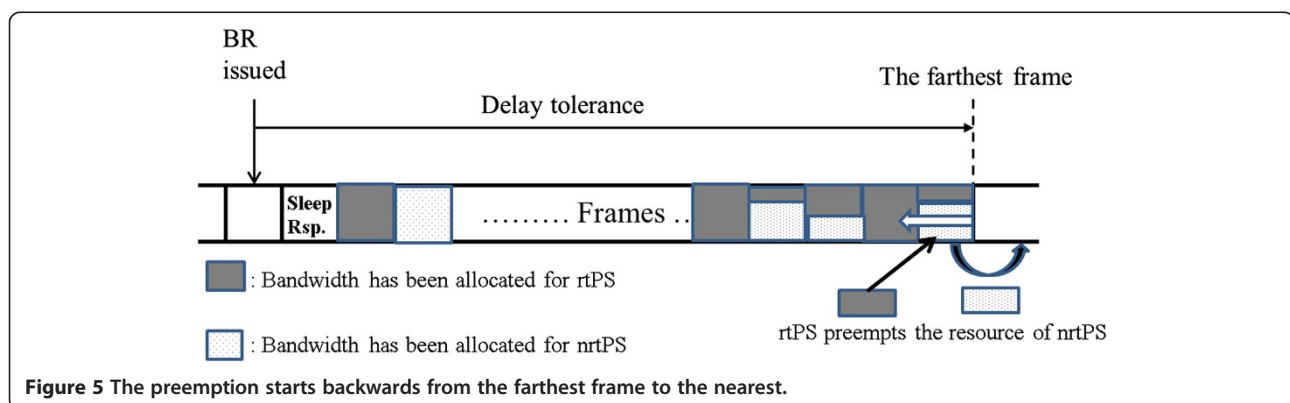

Figure 5 The preemption starts backwards from the farthest frame to the nearest.

packet-dropping date, as stated in lines 19 to 21. The preemption procedure is provided in lines 34 to 43. The preempted bandwidth is rescheduled at the frame in which nrtPS MS has been previously allocated, and if insufficient, the BS will allocate them after the $(n + D_i - \alpha + 1)$th frame.

Algorithm 1 The procedure of rtPS scheduling

1:    For each MS i who issues BR
2:    **if** MS_i is rtPS connection **then**
3:     **if** $n + D_i - \alpha \geq n + 2$ **then**
4:      **if** MS_i has pre-schedule(reserved) at the $(n+2)$th to $(n+x+1)$th frame AND $BR_{i,n} \leq Av_i((n+2),(n+x+1))$ **then**
5:       Allocate $BR_{i,n}$ at $(n+2)$th to $(n+x+1)$th frame /*allocates the bandwidth for MS_i with the nearest frames in which BS has reserved bandwidth */
6:      **elseif** MS_i has pre-schedule(reserved) at $(n + D_i - \alpha)$th frame **and** $BR_{i,n} \leq Av_i((n+D_i-\alpha),(n+D_i-\alpha))$ **then**
7:       Allocate $BR_{i,n}$ at $(n + D_i - \alpha)$th frame /* allocates the bandwidth for MS_i with the farthest frames */
8:      **else**
9:       **for** $j = n + D_i - \alpha$ to $n + 2$ frame
10:        **if** the jth frame has pre-schedule(reserved) for MS_i **then**
11:         $BR_{i,n} \leftarrow BR_{i,n} - BW_j$ /* allocates the bandwidth for MS_i between the nearest frames and farthest frames in which BS has reserved bandwidth */
12:        **endif**
13:       **endfor**
14:       **if** $BR_{i,n} \geq 0$ **then**
15:        Find BW_K in $n + D_i - \alpha \sim n + 2$ frames
16:        **repeat**
17:         $BR_{i,n} \leftarrow BR_{i,n} - BW_K$ /* allocates the bandwidth for MS_i between the nearest frames and farthest frames*/
18:        **until** $BR_{i,n} = 0$ or BW_K is zero in $n + D_i - \alpha \sim n + 2$ frames
19:        **if** $BR_{i,n} \geq 0$ **then**
20:         **call** *Proc Preemption* /*MS_i preempt bandwidth that allocates for nrtPS*/
21:        **endif**
22:       **endif**
23:      **endif**
24:     **else**
25:      Drop cannot allocate bandwidth
26:     **endif**
27:    **else**
28:     **call** *Proc nrtPS Scheduling*
29:    **call** *Sleep Cycle Decision*
Sleep Cycle Decision:
30:    Calculate K_i^n according to Equation 2
31:    **if** $K_i^n \geq$ SLEEP_THRESHOLD **then**
32:     MS_i can enter sleep mode and number of consecutive frames for sleep is K_i^n
33:    **endif**
Proc Preemption:
34:    **while** $BR_{i,n}$ is not zero
35:     **for** $x = n + D_i - \alpha$ to $n + 2$ frame
36:      P_x^{nrt} is preempted from nrtPS connection's bandwidth
37:      $BR_{i,n} \leftarrow BR_{i,n} - P_x^{brt}$ /* to preempt bandwidth to MS_i for insufficient bandwidth */
38:      Reschedule P_x^{nrt} starting from the $(n + D_i - \alpha + 1)$th frame until P_x^{nrt} is satisfied /*reschedule the bandwidth that is preempted for MS_i*/
39:      if $BR_{i,n} =$ zero **then**
40:       **break**
41:      **endif**
42:     **endfor**
43:    **endwhile**

Algorithm 2 shows the procedure of the nrtPS scheduling. The main difference between nrtPS scheduling and rtPS scheduling is that the nrtPS has no delay constraint and it can be allocated with bandwidth after the $(n + N_i - \beta + 1)$th frame until its bandwidth requirement has been satisfied.

Algorithm 2 The procedure of nrtPS scheduling

1:   **if** MS_i has pre-schedule (reserved) at $(n + 2)$th to $(n + x + 1)$th frame **and** $BR_{i,n} \leq Av_i((n + 2),(n + x + 1))$ **then**

2:       Allocate $BR_{i,n}$ at $(n + 2)$th ∼ $(n + x + 1)$th frames

3:   **elseif** MS_i has pre-schedule (reserved) at $(n + N_i - \beta)$th frame **and** $BR_{i,n} \leq Av_i((n + N_i - \beta),(n + N_i - \beta))$ **then**

4:       Allocate $BR_{i,n}$ at $(n + N_i - \beta)$th frame

5:   **else**

6:       **for** $j = n + N_i - \beta$ to $n + 2$ frame

7:         **if** the jth frame has pre-schedule(reserved) for MS_i **then**

8:             $BR_{i,n} \leftarrow BR_{i,n} - BW_j$

9:         **endif**

10:      **endfor**

11:      **if** $BR_{i,n} \geq 0$ **then**

12:         Find BW_K in $n + N_i - \beta \sim n + 2$ frames

13:         **repeat**

14:             $BR_{i,n} \leftarrow BR_{i,n} - BW_K$

15:         **until** $BR_{i,n} = 0$ or K_i^{nrt} = SLEEP_THRESHOLD /*let MS_i can sleep period is more than sleep_threshold to switch to sleep mode*/

16:         **if** $BR_{i,n} \geq 0$ **then**

17:             Allocate $BR_{i,n}$ from $(n + N_i - \beta + 1)$th until $BR_{i,n}$ is zero /*allocate insufficient bandwidth for MS_i*/

18:         **endif**

19:      **endif**

20:   **endif**

21:   **call** *Sleep Cycle Decision*

Sleep Cycle Decision:

22:   Calculate K^{nrt}_i according to Equation 6

23:   **if** $K^{\mathrm{nrt}}_i \geq$ SLEEP_THRESHOLD **then**

24:     MS_i can enter sleep mode and number of consecutive frames for sleep is K^{nrt}_i

25:   **endif**

---

5 Performance simulation

To examine the performance of the proposed scheme, both the VoIP and FTP traffic models were used as real-time traffic and non-real-time traffic for exhaustive simulations, respectively. Exhaustive simulations were conducted to compare the results of the proposed

Table 2 System simulation parameters

System	FTT	Frame duration	DL/UL		BW	BW per frame
TDD	1024	5 ms	80%	20%	7 MHz	75 kbits

Table 3 The real-time VoIP traffic model parameters

Service	Voice over IP (VoIP)
Voice CODEC	AMR (12.2 kbps)
Frame length	5 ms
Talk spurt length	Exponential, mean = 1,026 ms
Silence length	Exponential, mean = 1,171 ms
Silence suppression	Yes
Protocols	RTP/UDP/IP
Header compression	RTP/UDP/IP header compression
Speech activity	47.17%

algorithm and the traditional algorithm in different scenarios. The traditional algorithm is contrary to the proposed SbT concept: the BS allocates bandwidth to the MS on receiving the BR and the MS enters sleep mode when the number of consecutive frames for sleep is more than the SLEEP_THRESHOLD but awakens when uplink traffic generates. Each MS is assumed to establish either one real-time or non-real-time connection, and the channel condition is assumed to be unchanged. If the channel condition is considered, only the parameter BW_{frame} of the proposed schemes needs to be adjusted to vary according to the channel condition received by the

MS at the allocation time. Thus, although the channel condition will affect the system performance, the main design concept of the proposed scheme is unaffected. The parameters used for simulations are shown in Table 2.

The VoIP traffic is generated for the real-time connection, and its characteristics are specified in Table 3 [19]. Two types of maximum delay requirements are further assumed for the VoIP connections: Class I is 50 ms (10 frames) and Class II is 75 ms (15 frames). The packets of non-real-time connections arrive in the Poisson distribution and in exponentially distributed packet sizes. The mean packet inter-arrival time is 100 ms, and the mean packet size is 15 kbits. Moreover, two MRTR requirements exist for the non-real-time connection. The MRTR requirements of Class I and II are 700 and 1,500 bits per frame, respectively. The parameters α and β were assumed to be $\frac{K^{rt}_m}{2}$ and $\frac{K^{nrt}_m}{2}$, respectively, and the SLEEP_THRESHOLD was set to be three frames in length during the simulations.

The relationship between energy consumption and symbol duration was studied in [20] and [21] to determine circuit power and transmission power for overall energy efficiency. The authors indicate that the transmission power and the circuit power have different consumption models with respect to the transmission time per bit. In this paper, there is no variance for the transmission time
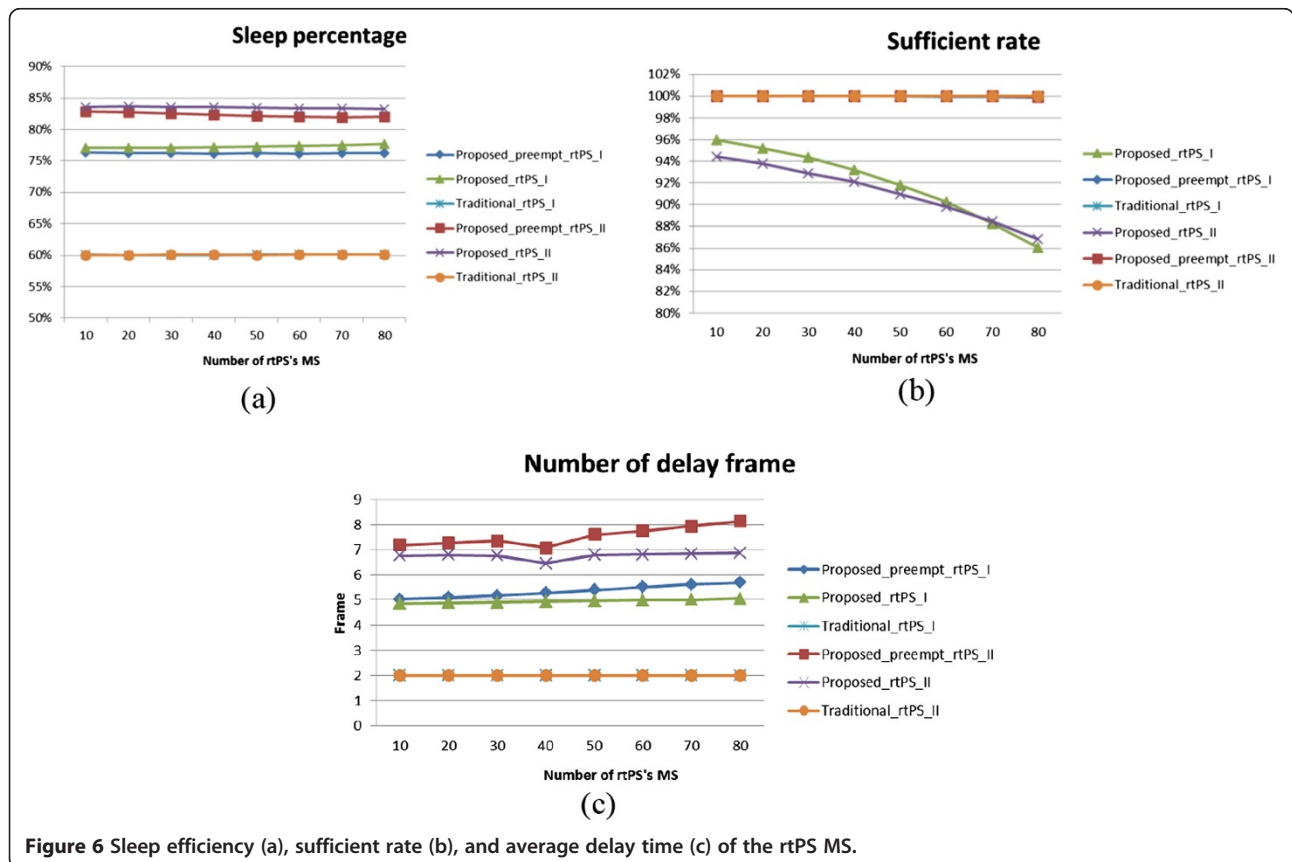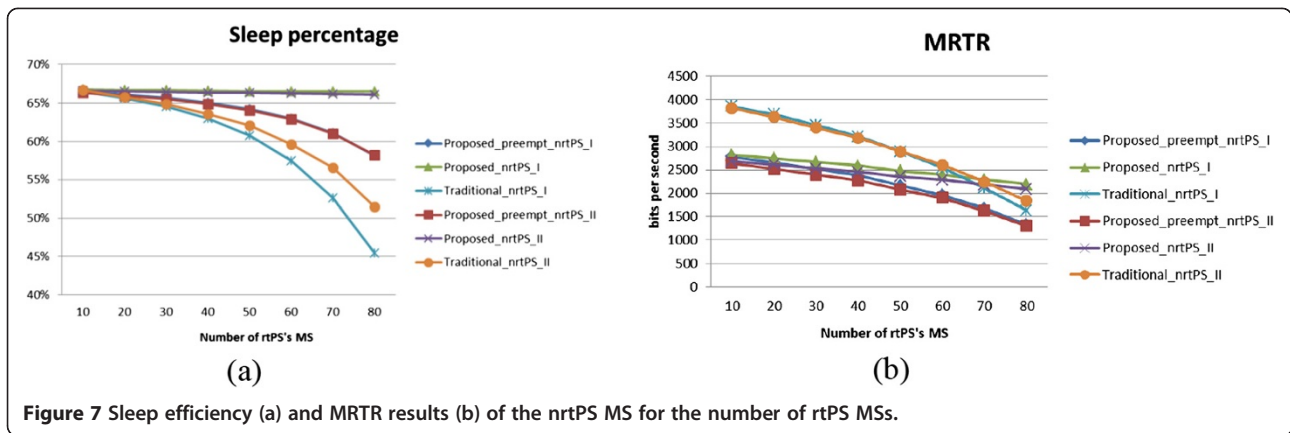

Figure 6 Sleep efficiency (a), sufficient rate (b), and average delay time (c) of the rtPS MS.

Figure 7 Sleep efficiency (a) and MRTR results (b) of the nrtPS MS for the number of rtPS MSs.
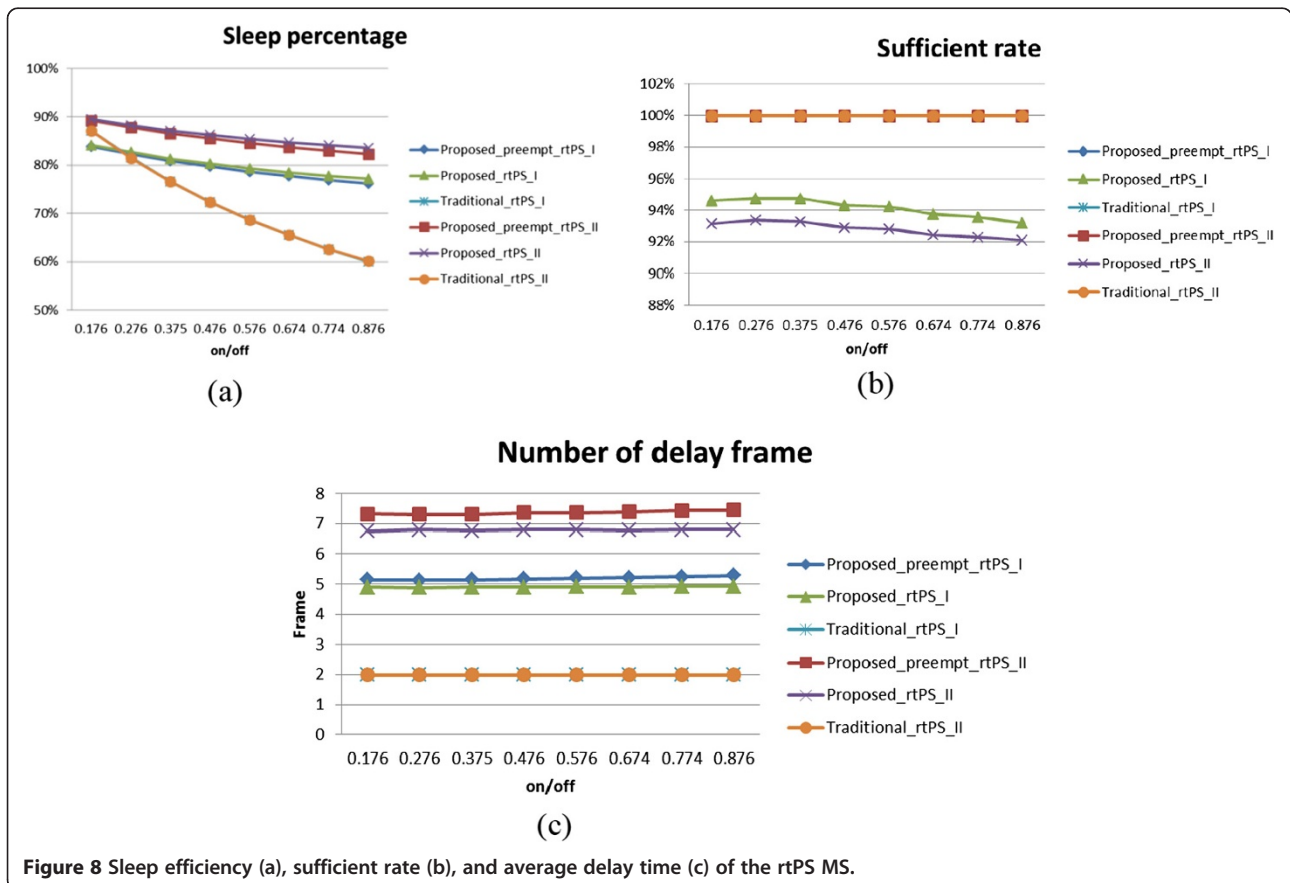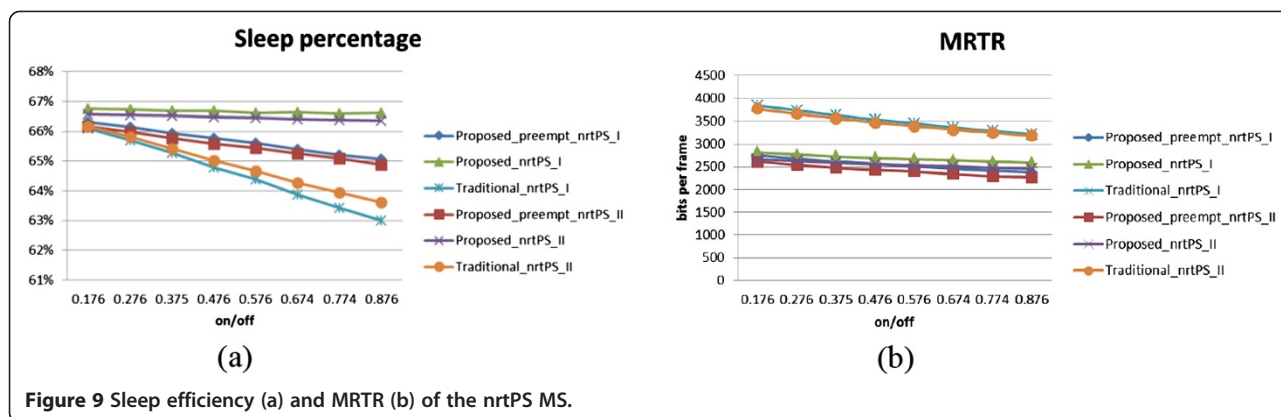
per bit in the transmission system; hence, we consider the sleep efficiency in term of the percentage of frames that MS sleeps. In addition, the following performance index are also compared: average delay time (in terms of number of frames), sufficient rate of rtPS connections, and average MRTR of nrtPS connections. The sufficient rate is defined as the ratio of the actual bandwidth allocated for the MS to its requested bandwidth. The MRTR and the sufficient rate were examined to determine whether the proposed scheme can allocate sufficient radio resources

in each allocated frame for the nrtPS traffic. MRTR ρ is defined as

$$\rho = \frac{N_\mathrm{p} \times P_\mathrm{s}}{N_\mathrm{f}} \qquad (6)$$

where N_p is number of sent packets, P_s and N_f denote the packet size and the number of used frames, respectively.

Figure 8 Sleep efficiency (a), sufficient rate (b), and average delay time (c) of the rtPS MS.

Figure 9 Sleep efficiency (a) and MRTR (b) of the nrtPS MS.

The average sleep efficiency is defined as the average percentage of the number of sleep frames to the number of total frames per connection during the simulations.

5.1 Fixed number of nrtPS MS

Figure 6 shows the sleep percentage, sufficient rate, and average delay results of the rtPS MS for the different numbers of rtPS MS. The number of nrtPS connections is fixed at 10. The simulation results show that the proposed scheme has better sleep efficiencies for both connections compared to the traditional scheme. The results demonstrate that the proposed scheme can improve the sleep efficiency by up to 20% to 25%. The rtPS Class II connection has a better sleep efficiency than the rtPS Class I connection. The main reason is that Class II has a longer delay tolerance than Class I, and therefore has a higher probability to arrange longer sleep periods, which additionally indicates that sufficient rates of the non-preemption schemes decrease as the number of rtPS MS increases. The main reason for this occurrence is that the rtPS cannot preempt the reserved bandwidth for the nrtPS in non-preemptive cases and the sufficient rate decreases as the traffic load increases because of the increase in rtPS connections. This phenomenon does not exist in the preemptive and traditional schemes. Figure 6(c) shows the performances of the average delays, which indicates that the proposed scheme has longer delay times than the traditional scheme. The average delay time of the rtPS Class I and Class II are 5 and 7.5 frames, as shown in Figure 6(c), respectively. However, they are still under the required maximum delay tolerance (10 frames and 15 frames, respectively) because the proposed scheme adopts the just enough QoS concept to conserve energy.

The MRTR was examined to determine whether the proposed scheme can allocate sufficient radio resources in each allocated frame for the nrtPS traffic but still conserve energy. Figure 7 shows the sleep percentage and MRTR results of the nrtPS MS for the number of rtPS MSs. Figure 7a shows that the proposed scheme has better

sleep percentages than those of the traditional scheme, especially when the number of MSs increases. Moreover, it shows that the MRTR of the proposed scheme is less than that of the traditional scheme, but they still satisfy the assumed MRTR requirements of 700 and 1,500 bits per frame, respectively. The results satisfy the just enough QoS concept used in the proposed scheme.

5.2 Effect of rtPS on/off ratio

Figure 8 shows the results of the sleep percentages, sufficiency rates, and average delay results of the rtPS MS. Figure 9 provides the results of the sleep percentages and MRTR of the nrtPS MS. The numbers of the rtPS and nrtPS connections in this simulation are 40 and 10, respectively. In Figure 8, the sleep efficiency decreases as the on/off ratio increases because of the longer "on" duration. The sufficient rate of the non-preemptive schemes is less than 100% as explained in the previous simulation. Although the proposed schemes have higher delay times than the traditional scheme, the desired performance is still satisfied. The results shown in Figure 9 indicate that the sleep efficiencies of the proposed schemes perform better than those of the traditional scheme and still satisfy the required MRTR.

5.3 Sleep threshold effect for MS

The traditional and proposed schemes define a threshold to prevent MSs from frequently switching modes in the

Table 4 The sleep efficiency of rtPS Class I with respect to the sleep threshold

Sleep threshold	Traditional	Proposed (non-preempt)	Proposed (preempt)
3	60.03%	76.89%	76.51%
4	58.33%	75.51%	75.43%
5	57.00%	74.46%	74.29%
6	55.94%	73.29%	73.08%
7	55.08%	71.67%	71.41%

Table 5 The sleep efficiency of rtPS Class II with respect to the sleep threshold

Sleep threshold	Traditional	Proposed (non-preempt)	Proposed (preempt)
3	60.14%	83.75%	83.26%
4	58.45%	83.14%	82.91%
5	57.12%	82.55%	82.52%
6	55.94%	81.46%	81.44%
7	55.21%	80.56%	80.44%
8	54.52%	79.05%	78.96%
9	53.93%	77.29%	77.12%
10	53.44%	68.35%	68.26%
11	53.01%	59.06%	59.05%
12	52.63%	57.19%	57.18%

simulation; thus, MSs do not switch to sleep modes when the number of consecutive frames for sleep is less than the defined sleep threshold. A simulation was conducted by varying the threshold to examine sleep efficiency. The numbers of rtPS and nrtPS connections in this simulation are 40 and 6, respectively. The sleep percentages of rtPS Class I and II MSs are shown in Tables 4 and 5. The sleep percentages decrease as the threshold increases because the MS will be unable to enter sleep mode if the number of consecutive frames for sleep is less than the specified threshold following radio resource allocation. The traditional scheme schedules the radio resource to the MS as soon as possible on receiving the BR. Thereafter, the BS always allocates the nearest frames, which have residual radio resources. They are allocated to the MS without considering the provisioning of the just enough delay tolerance to achieve longer consecutive frames for sleep. Therefore, the sleep percentage of the traditional scheme is less than that of the proposed scheme, as shown in the tables. The delay tolerances of rtPS Classes I and II in the simulation are

assumed to be 10 and 15 frames, respectively. Equation 2 of Section 3 indicated that the MS must awaken two frames earlier to receive the UL-MAP to know when to and which frame to transmit data to. The maximum sleep periods of rtPS Classes I and II are 7 and 12 frames, respectively. Therefore, the Class I rtPS MS is unable to enter sleep mode when the sleep threshold is greater than 7. The Class II rtPS MS is unable to enter sleep mode if the threshold is greater than 12.

The performance of the average delay is provided in Figure 10. It shows that the average delay increases as the threshold value increases, but the average delays of the Class I and Class II rtPS MSs decrease sharply for the threshold values, which equal 8 and 13, respectively. The main reason is that the MS mostly remains awake for such large threshold values and the data packets have a tendency to be delivered in shorter delays.

6. Conclusion

The uplink scheduling scheme is proposed from an energy-saving viewpoint in this study. Contrary to the traditional concept, the novel SbT concept is introduced in the proposed ESC-US scheme. The proposed scheme analyzes the sleep efficiency with the QoS constraints to determine whether the MS is suitable to enter sleep mode and calculates the proper sleep period in a systematic manner if the MS can enter sleep mode. Both rtPS and nrtPS connections are considered in the proposed scheme. The main achievement of this study was improving the energy-saving efficiency under the desired QoS constraint. The proposed scheme arranged the radio resources to satisfy its QoS in a just enough manner to maximize the sleep efficiency. Several performance indices, such as the average delay, sleep efficiency, sufficient rate, and MRTR, were examined through exhaustive simulations. The performance of the proposed scheme is compared to the traditional uplink scheduling scheme. The simulation results demonstrate that the proposed SbT concept is
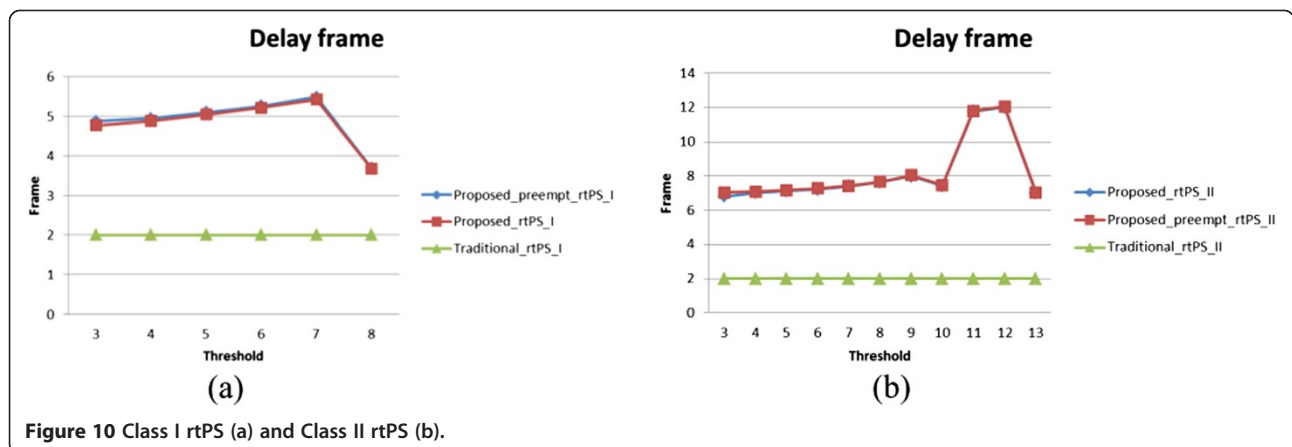

Figure 10 Class I rtPS (a) and Class II rtPS (b).

meaningful and the energy-saving performance is superior to the traditional scheme. The simulation results clearly show the tradeoff between energy savings and performance. The proposed SbT with the just enough QoS concept provides an alternative method for considering the tradeoff issue. Additionally, the approach can be adopted for LTE network by extending the proposed concept to be applied for the relative power-saving parameters such as the decision of DRX cycles. Although the ESC-based uplink scheduling algorithm is proposed in this study and downlink scheduling is proposed in [18], the integration of the downlink and uplink energy-saving centric scheduling is more complex. Further study for the harmonic operation between these two independent algorithms is required to maximize the overall system performance.

Acknowledgement

Author details

[1]Department of Communication Engineering, National Central University, Jhongli City, Taoyuan County 01, Taiwan. [2]Department of Computer Science and Information Engineering, Minghsin University of Science and Technology, Xinfeng Hsinchu 30401, Taiwan.

References

1. IEEE Standard for Local and metropolitan area networks Part 16, *Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1, IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005*, 2006
2. 3GPP TS 36.321 V1. 0.0 Access, Evolved Universal Terrestrial Radio, *Medium Access Control (MAC) protocol specification (Release 8)*, 2008
3. K Lahiri, A Raghunathan, S Dey, Battery-driven system design: a new frontier in low power design, in *Proceedings of the 7th Asia and South Pacific and the 15th International Conference on VLSI Design, Bangalore*, 2002, p. 261
4. G Miao, N Himayat, Y Li, A Swami, Cross layer optimization for energy efficient wireless communications: a survey. Wirel. Commun. Mob. Comput. **9**, 529–542 (2009)
5. G Miao, N Himayat, GY Li, S Talwar, Low-complexity energy-efficient scheduling for uplink OFDMA. Commun. IEEE. Trans. **60**, 112–120 (2012)
6. Y Xiao, Energy saving mechanism in the IEEE 802.16e wireless MAN. IEEE Commun. Lett. **9**, 595–597 (2005)
7. Y Zhang, M Fujise, Energy management in the IEEE 802.16e MAC. IEEE Commun. Lett. **10**(4), 311–313 (2006)
8. JB Seo, SQ Lee, NH Park, HW Lee, CH Cho, Performance analysis of sleep mode operation in IEEE 802.16e, in *Proc. IEEE 60th Vehicular Technology Conference, vol. 2*, 2004, pp. 1169–1173
9. K Han, S Choi, Performance analysis of sleep mode operation in IEEE 802.16e mobile broadband wireless access systems, in *IEEE 63rd Vehicular Technology Conference, vol. 3, Melbourne*, 2006, pp. 1141–1145
10. J Lee, D Cho, An optimal power-saving class II for VoIP traffic and its performance evaluations in IEEE 802.16e. Computer. Comm. **31**(14), 3204–3208 (2008)
11. TC Chen, JC Chen, YY Chen, Maximizing unavailability interval for energy saving in IEEE 802.16e wireless MANs. IEEE Trans. Mobile. Comput. **8**(4), 475–487 (2009)
12. L Kong, DH Tsang, Performance study of power saving classes of types I and II in IEEE 802.16e, in *Proc. 31st IEEE Conf. Local Computer Networks (LCN'06), Tampa*, 2006, pp. 20–27
13. Y Zhang, Y Xiao, MVC Leung, Energy management analysis and enhancement in IEEE 802.16e wireless MAN. IEEE Trans. Veh. Technol. **58**(7), 3738–3752 (2009)
14. J Shi, G Fang, Y Sun, J Zhou, Z Li, E Dutkiewicz, WLC17-5: improving mobile station energy efficiency in IEEE 802.16e WMAN by burst scheduling, in *Global Telecommunications Conference (GLOBECOM'06), San Francisco*, 2006, pp. 1–5
15. SC Huang, RH Jan, C Chen, Energy efficient scheduling with QoS guarantee for IEEE 802.16e broadband wireless access networks, in *Proceedings of the 2007 International Conference on Wireless Communications and Mobile Computing, New York*, 2007, pp. 547–552
16. SL Tsao, YL Chen, Energy-efficient packet scheduling algorithms for real-time communications in a mobile WiMAX system. Comput. Commun. **31**(10), 2350–2359 (2008)
17. SR YANG, CC KAO, An energy-efficient scheduling algorithm for IEEE 802.16 e broadband wireless access systems, in *Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing: Connecting the World Wirelessly; New; New York*, 2009, pp. 532–536
18. SC Ke, YW Chen, HA Fang, An energy saving centric downlink scheduling scheme for WiMAX networks. Int. J. Commun. Syst. (2012). doi:10.1002/dac.2486
19. R Jain, *WiMAX System Evaluation Methodology v2. 1* (WiMAX Forum, 2008)
20. G Miao, N Himayat, YG Li, A Swami, Cross-layer optimization for energy-efficient wireless communications: a survey. Wireless. Commun. Mob. Computing. **9**(4), 529–542 (2009)
21. D Feng, C Jiang, G Lim, LJ Cimini, G Feng, GY Li, A survey of energy-efficient wireless communications. IEEE Commun. Surveys. Tuts. **15**(1), 167–178 (2013)