MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Adaptive Probabilistic Decision-Based Energy Saving Strategy for the Next Generation Cellular Wireless Systems

Weihuang Fu, Zhifeng Tao, Jinyun Zhang, Dharma Agrawal

TR2010-051 July 2010

Abstract

As mobile stations (MSs) in the next generation cellular wireless systems will more frequently operate on multiple applications, such as web browsing, VoIP, online video etc., energy saving becomes more critical and face new challenges from the quality of service (QoS) requirements. A special operation state, called sleep mode, is designed for energy saving of MS, in which MS operates on continuous sleep cycles, where every sleep cycle is the sum of a listening window and a sleep window. This paper proposes an energy saving method that adaptively determines sleep cycles and shifts listening window. When MS is in sleep mode, the sleep cycles are extended by probabilistic decisions related to the traffic statistic attributes. We also introduce the energy saving strategy for the frames by mixing best effort and persistent allocation traffic. The frequency of sleep cycles is used as one of the parameters for QoS purpose. Different from conventional methods, listening window can be shifted for shorter response time in our method. Extensive simulation results validate the advantages of our method both in terms of energy saving and shorter response time.

IEEE ICC 2010

Adaptive Probabilistic Decision-based Energy Saving Strategy for the Next Generation Cellular Wireless Systems

Weihuang Fu[1], Zhifeng Tao[2], Jinyun Zhang[2], and Dharma P. Agrawal[1]
[1]Center for Distributed and Mobile Computing
Department of Computer Science, University of Cincinnati, Cincinnati, OH 45221, USA
[2]Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA
Email: fuwg@cs.uc.edu, tao@merl.com, jzhang@merl.com, and dpa@cs.uc.edu

***Abstract*—As mobile stations (MSs) in the next generation cellular wireless systems will more frequently operate on multiple applications, such as web browsing, VoIP, online video etc., energy saving becomes more critical and face new challenges from the quality of service (QoS) requirements. A special operation state, called sleep mode, is designed for energy saving of MS, in which MS operates on continuous sleep cycles, where every sleep cycle is the sum of a listening window and a sleep window. This paper proposes an energy saving method that adaptively determines sleep cycles and shifts listening window. When MS is in sleep mode, the sleep cycles are extended by probabilistic decisions related to the traffic statistic attributes. We also introduce the energy saving strategy for the frames by mixing best effort and persistent allocation traffic. The frequency of sleep cycles is also used as one of the parameters for QoS purpose. Different from conventional methods, listening window can be shifted for shorter response time in our method. Extensive simulation results validate the advantages of our method both in terms of energy saving and shorter response time.**

I. Introduction

Various communication/network applications, such as Web browsing, VoIP, online video, etc., have been emerging on handheld devices, i.e., mobile stations (MSs), as wider bandwidth support is expected from the next generation cellular wireless technology (e.g., IEEE802.16m [1], IEEE802.16j [2], LTE advanced [3]). Exacerbated by more frequent usage of handheld devices, power management with constraint energy supply from carry-on batteries, which has limited size and capacity, becomes a critical issue in developing these next generation cellular system standards.

In this paper, we discuss the problem of energy saving in the next generation cellular systems with considerations of traffic quality of service (QoS). As the topic of energy saving may use alternate names and terms in different wireless systems but with same functions, we illustrate our idea by standards in the IEEE802.16m system. However, the idea and mechanism of our method can be easily transferred and adopted in other wireless system.

In an IEEE802.16m system, sleep mode is one of three modes that an MS in connected state [1], in which an MS conducts pre-negotiated periods of absence from its serving base station (BS) air interface. The transmitting and receiving of frames by an MS consume relatively higher energy. In sleep mode, MS does not have to listen to the air-interface all the time, and the energy should be saved during the non-transmission periods.

When an MS is in its sleep mode, it operates on a series of continuous sleep cycles. The period of the sleep cycle is measured in frames in the IEEE802.16m. A sleep cycle consists of a sleep window and a listening window. During sleep window, the BS does not transmit frames to the MS and the MS can turn the radio power off for energy saving. During listening window, the MS is to receive the downlink transmissions from the BS as the same way in the state of normal operations. The default length of listening window can be a fixed value set during the initiation of sleep mode or during the sleep cycle update.

In previous wireless systems, binary truncate exponent (BTE) algorithm [4][5][6] and its variations [7][8] have been widely implemented. Many wireless cellular system including IEEE802.16e [9] use BTE to determine the length of sleep cycles. After MS enters in sleep mode, BTE algorithm will double the sleep cycles every time, until a pre-negotiated maximum sleep cycle is reached or the sleep mode is terminated. In the sleep mode operating with BTE algorithm, the MS wakes up at the beginning of every sleep cycle for signals from the BS or data transmission.

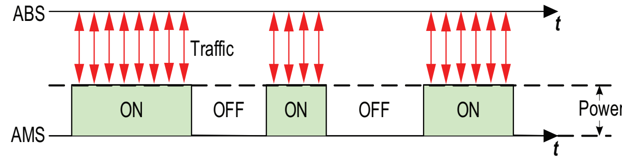
The applications running on MSs can include best effort (BE) traffic and/or persistent allocation (PA) traffic. In IEEE 802.16m, MAC link is connection-based and the data exchange is performed by scheduling algorithm. BE traffic is scheduled by BS when transmission is possible. The examples of BE traffic are web browsing, email traffic, etc. PA is used for the traffic with periodic pattern and relatively fixed payload size to reduce the resource map overhead of connections. The examples of PA traffic are voice traffic, video traffic, etc. Our investigation shows that there exists a space for improvement between the performance of BTE algorithm and the optimal solution. And, BTE algorithm is not suitable for BE and PA mixed traffic due to the absence of energy saving consideration to PA traffic.

In this paper, we propose an adaptive energy saving strategy in sleep mode to approach the optimal energy saving. Different from conventional energy saving that only considers BE traffic,

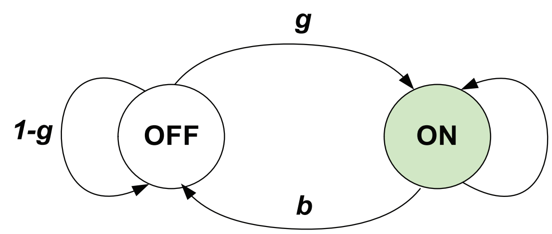
Weihuang Fu worked on this study while visiting Mitsubishi Electric Research Lab, Cambridge, MA.

our scheme supports BE traffic, PA traffic, and their mixed traffic. The advantages of our method lie in the following aspects: (i) The sleep cycle is extended based on probabilistic decision instead of binary truncation. (ii) The set of extension factors in our method is adaptively derived from the cumulative distribution function (CDF) of the packet arrivals, so that the probabilities of packet capture in sleep window are nearly equal. (iii) Sleep mode is able to adaptively deal with mixed traffic pattern by shifting or interpolating listening window in a sleep cycle. (vi) Our method is able to deal with mixed traffic, which is not effectively addressed by prior sleep mode.

The remainder of this paper is organized as follows: The problem formulation and modeling of energy saving mechanism is in Section II. In Section III, we analyze the traffic attribute and propose the energy saving strategies for BE and PA traffic. Section IV provides the performance evaluation with extensive simulations. Finally, the paper is concluded in Section V.



(a) ON-OFF model



(b) State transit diagram

Fig. 1: Traffic state model

II. Problem Formulation and Modeling

In this section, we model the energy saving mechanism and introduce the sleep mode in IEEE802.16m to illustrate severity of the problem. Our development of the sleep model operation scheme is based on the characteristics of these traffic to obtain near-optimal energy saving with shorter response time.

A. Traffic Model

Application traffic between advanced MS (AMS) and advanced BS (ABS) can be profiled by an ON-OFF model [10]. For example, a typical phone conversation can be marked by a series of periods of active talking interleaved by silence/listening period, as shown in Fig. 1(a). ABS schedules the traffic transmission for certain duration, denoted by arrows in the figure. The scheduling of traffic exchange does not always exist. During some time intervals, traffic does not exist between ABS and AMS. The period for the presence of traffic between AMS and ABS is called ON state and the period that traffic not there is the OFF state. So, the power supply of AMS radio module can be temporarily turned off or be reduced to a sleep power level during the OFF periods, and resume the normal power level during the ON periods. The process can be seen as a random process that transits between ON and OFF states, as the two state Markov chain shown in Fig. 1(b). Since the length of frame in IEEE802.16m is fixed, the state transit happens only at the end of every frame. Mathematically, the conditional probability that the AMS transits from ON state to OFF state can be denoted by b and the conditional probability that the AMS which is in the ON state will continue remain in the ON state is given by $1 - b$. Similarly, the conditional probability that the AMS which is in OFF state changes to the ON state at next frame is denoted by g and the conditional probability that the AMS will remain in the OFF state is given by $1 - g$. The values of the parameters highly depend on the traffic attributes.

We illustrate some traffic attributes in ON-OFF model as follows, including BE traffic and PA traffic.

- *Web browsing traffic*: The attributes of web browsing traffic, i.e., HTTP traffic, governed by structure of webpages and human browsing behavior. It can be approximated by ON-OFF model. In the ON state, traffic is transferred from the website to the AMS; while the user is reading the webpage, it remains on the OFF state. When the user finish the reading and open another webpage, the state will be switched back to the ON state.
- *Voice traffic*: A whole phone conversation are comprised by active talking and silence. Six state model has been introduced in [11]. If we reduce those states that involve any talk, to the ON state and reduce the mutuall silence state to OFF state, a simplified two state ON-OFF model is obtained.
- *FTP traffic*: FTP traffic can be well characterized by ON-OFF model, where the files are in the transfer in ON state and user is in reading during the time interval of OFF state.
- *Email traffic*: When email traffic is modeled by ON-OFF model, we see that: in the ON state, the email can be sent to the server or received from the server; in the OFF state, user is writing or reading the email locally.

There are many other types of traffic can be modeled by ON-OFF model, which can refer to [10].

B. Sleep Mode

Sleep mode is one of the functions that AMS uses in power management to conduct pre-negotiated air interface power-off intervals. The length of sleep cycle is measured in frames in recent IEEE802.16m system description document. Normally, a sleep cycle is comprised of a listening window and a sleep window, which may be measured in units of subframe. The operation of sleep mode in IEEE802.16m is as follows [1]:

The initiation of sleep mode can be activated or entered either by AMS or ABS, and the parameters of sleep cycle are negotiated between the AMS and the ABS. The ABS makes the final decision regarding the AMS request and instructs

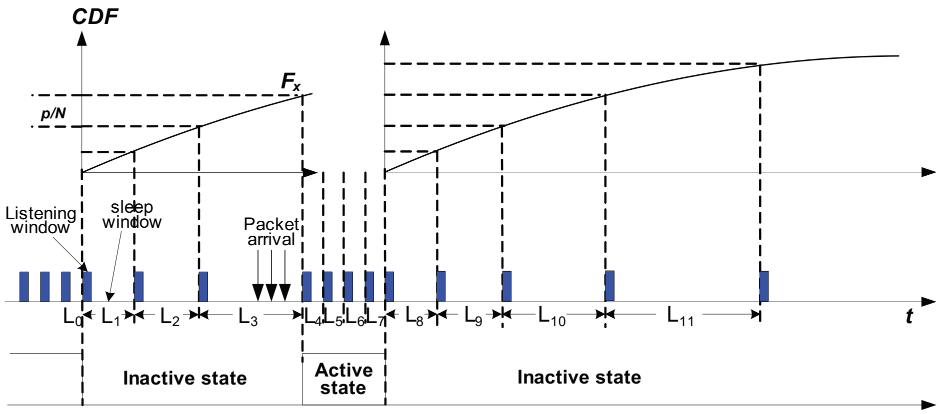


Fig. 2: Equal Probability Division

the AMS to enter sleep mode. The negotiation of sleep parameters is performed by using MAC management messages MOB_SLP-REQ and MOB_SLP-RSP. The AMS can initiate the negotiation by sending a MOB_SLP-REQ message and receives a MOB_SLP-RSP message. Alternatively, the ABS can also send an unsolicited MOB_SLP-RSP message to the AMS.

During the period of sleep mode, AMS may increase its sleep cycle under several conditions:

- When the AMS does not receive any packets from its downlink during a listening window, or
- When the AMS receives a management message indicating that the traffic will not be scheduled in the following frames.

The termination of sleep mode can be made either by the AMS or the ABS. If the AMS initiates the termination, it will send a MOB_SLP-REQ message with deactivation request, which ends the sleep mode by confirmation from the ABS with responding a MOB_SLP-RSP message. The ABS can also send an unsolicited MOB_SLP-RSP message to deactivate sleep mode directly.

III. Analysis and Scheme

A. Analysis of Traffic Arrival

In active state, packets have short inter-arrival intervals. As a result, AMS will have higher probability to receive packets during its listening window of sleep mode. In contrast, AMS has lower probability to receive packets during its listening window during inactive state of downlink traffic. Since the duration of inactive state can be identified by ON-OFF traffic model, the power saving should be optimized with intelligent adaptation of sleep cycle length.

Fig. 2 illustrates our proposed sleep cycle adaption. In active state, listening window appear every L_0 time length, where L_0 denotes the normal sleep cycle duration and is initiated at the beginning of sleep mode. When traffic enters inactive state, AMS prolongs the sleep cycle to save more energy.

Assume there are N listening windows when the traffic is in inactive state. The optimized design is to let each listening window have the same probability of capturing packets arrival. In other words, the probabilities that packets arrive during each of these N listening windows should be the same. Given this assumption, the length of sleep cycle can be determined by the reverse mapping from cumulative distribution function (CDF) F_x of traffic arrival in inactive state. As shown in Fig. 2, sleep cycle duration L_1, L_2, and L_3 have the same corresponding probability of packet arrivals in CDF F_x and their listening windows have the same probability to capture these packet arrivals. For example, when packets arrive during the sleep window of L_3, the AMS will receive the packets in the next listening window of L_4. The sleep mode will thus transit from the inactive state to the active state. When the sleep mode transits from the active state to the inactive state, the same procedures will be repeated again. While listening window of L_8 does not receive traffic, the next sleep cycle L_9 is extended per the described method.

Given the CDF of the inactive state duration, F_x, our scheme divides the probability between 0 and p ($p < 1$) into N equal segments and sets the starting points of each sleep cycle accordingly. Fig. 2 illustrates the idea that CDF F_x is equally divided into N non-overlap segments between probability 0 and p so that the probabilities that each sleep window can capture packet arrivals are equal. Given CDF F_x, the ith sleep cycle length $T_{slp}(i)$ can be computed as:

$$T_{slp}(i) = \begin{cases} F_x^{-1}(p\frac{i}{N}) - F_x^{-1}(p\frac{i-1}{N}), & i = 1, ..., N \\ T_{slp}(i-1), & i > N. \end{cases} \quad (1)$$

Since the sleep cycle is expressed in units of frame, $F_x(t)$ can also be easily changed to discrete CDF $F_k(i)$. The corresponding sleep cycle $L_{slp}(i)$ in frame is expressed by:

$$L_{slp}(i) = \begin{cases} F_k^{-1}(p\frac{i}{N}) - F_k^{-1}(p\frac{i-1}{N}), & i = 1, ..., N \\ L_{slp}(i-1), & i > N. \end{cases} \quad (2)$$

The relationship between the length of current sleep cycle

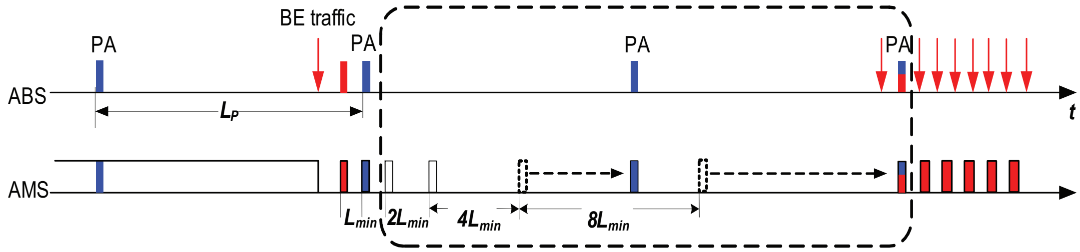


Fig. 3: $L_{min} < L_P$

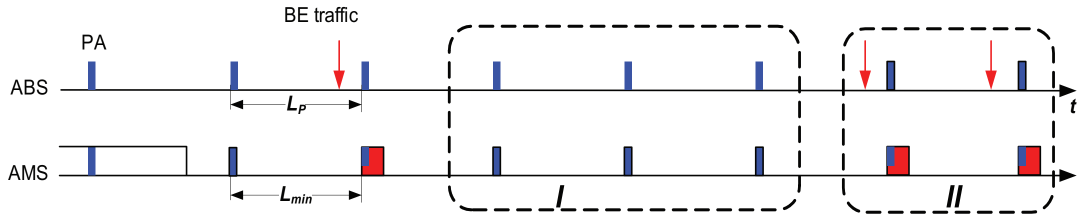


Fig. 4: $L_{min} \geq L_P$

and the length of previous sleep cycle can be given by:

$$L_{slp}(i) = \begin{cases} F_k^{-1}(\frac{p}{N}), & i = 1 \\ \frac{F_k^{-1}(p\frac{i}{N}) - F_k^{-1}(p\frac{i-1}{N})}{F_k^{-1}(p\frac{i-1}{N}) - F_k^{-1}(p\frac{i-2}{N})} L_{slp}(i-1), & i = 2, ..., N \\ L_{slp}(i-1), & i > N. \end{cases} \quad (3)$$

When the AMS enters inactive state of sleep mode, it computes the sleep cycle by:

$$L_{slp}(i) = \min(f_{ext}(i) \times L_{slp}(i-1), \text{Final Sleep Cycle}), \quad (4)$$

where

$$f_{ext}(i) = \frac{F_k^{-1}(p\frac{i}{N}) - F_k^{-1}(p\frac{i-1}{N})}{F_k^{-1}(p\frac{i-1}{N}) - F_k^{-1}(p\frac{i-2}{N})}. \quad (5)$$

The extension factors can be computed by the ABS and sent to the AMS at the initiation of sleep mode or during the sleep cycle update. They can also be computed by ABS and AMS separately by negotiating parameters, p, N, and F_x^{-1}. It should be noted that doubling of sleep cycle method, i.e., BTE algorithm, can be a special case of our method where $f_{ext}(i) = 2$, which is given by:

$$L_{slp}(i) = \min(2 \times L_{slp}(i-1), \text{Final Sleep Cycle}). \quad (6)$$

B. *Energy Saving for Mixed Traffic*

The listening window in a sleep cycle can be shifted for mixed traffic. Mixed traffic consists of periodic traffic and non-periodic traffic, where periodic traffic is scheduled by persistent allocation. There are two cases to consider for mixed traffic: $L_{min} < L_P$ and $L_{min} \geq L_P$, where L_{min} denotes the initial sleep cycle of BE traffic and L_P denotes the arrival interval of PA traffic in units of frame.

1) $L_{min} < L_P$: PA traffic and BE traffic are considered as the example shown in Fig. 3.

Sleep Cycle Extension: When no BE data is received during the listening window, the sleep cycle is extended to save energy. When only PA data is received during the listening window, the sleep cycle will not be extended. Sleep cycle extension follows:

$$L_{slp}(i) = \min(f_{ext}(i) \times L_{slp}(i-1), L_P, \text{Final Sleep Cycle}). \quad (7)$$

When AMS is receiving traffic via persistent allocation, the maximum sleep cycle size will be no larger than L_P. If BE data is received during listening window, the sleep cycle is reset to L_{min}.

Listening Window Shift: AMS should awake for PA frames. There are two situations based on the relative location of this frame. In the first situation, PA frame will be in the listening window of current sleep cycle, which does not need any adjustment for the listening window. The PA frame can be received during the listening window.

In the other situation, PA frame falls in the sleeping window of current sleep cycle. Without adjustment, the PA frame will be missed. So, the sleep cycle should be able to shift the listening window to embrace the PA frame. This offset is in term of number of frames. As the example shown in Fig. 3, the listening window is shifted to the PA frame location for both listening and PA frame receiving purposes.

2) $L_{min} \geq L_P$: The operation should be synchronized to save energy so that other traffic can be scheduled at the same frame used by the PA. When sleep cycle performs extension, it is also constrained by L_P, as Eq. (7) shows. In this case, L_{min} is equal to L_P. So, sleep cycle should be repeated with the same period and is synchronized with PA frames. In the example shown in Fig. 4, when BE traffic arrives, the data will be transmitted together with PA data to AMS during the listening window. If there is no BE traffic in arriving, as shown by region I of Fig. 4, the sleep cycle performs extension algorithm. However, due to $L_{min} = L_P$, the extended sleep cycle is still with length L_{min}. When BE traffic keeps arriving, it will be sent during the listening windows including PA

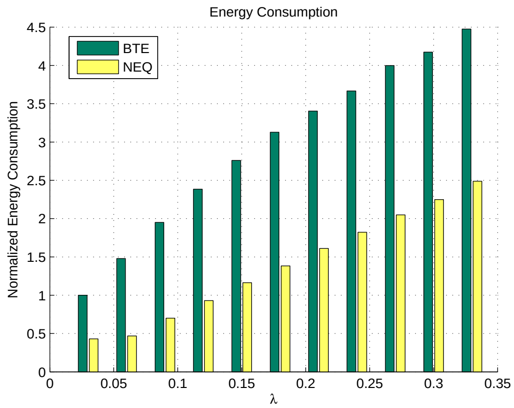


Fig. 5: Energy Consumption

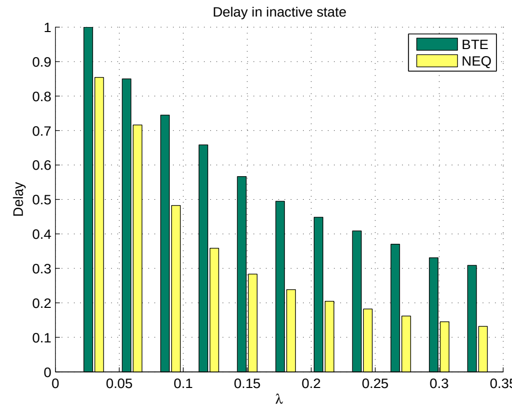


Fig. 6: Delay in Active State

frames, the sleep window may be extended to receive high volume BE traffic, as illustrated by region II of Fig. 4.

It is noted that persistent allocation may occur before or after the initiation of sleep mode. If sleep mode is set before the PA, the sleep mode should be reset or updated in the case $L_{min} > L_P$. By resetting or updating the default sleep cycle, L_{min} can be set to be less than or equal to L_P, which operates as the procedures described in the two cases above. This case also applies for "periodic-only traffic".

When persistent allocation is used during the sleep mode, the final sleep cycle should be changed to be equal to L_P. In each sleep cycle, it needs to check the expected location of PA frame. If the persistent allocation is terminated during sleep mode, the sleep cycle extension follows the calculation indicated in Eq. (4).

IV. Performance Evaluation

In this section, We compare the performance of energy consumption and packet delay in BTE (i.e., sleep window doubling scheme) with that of our proposed N-equal probability (NEQ) algorithms by using simulation. The Figs. 5 and 6 show the performance of both algorithms during inactive state. When packet is formed at the sender, it may be scheduled and transmitted to the receiver at the next available listening window. The sleep mode will transit from inactive state to active state. In the simulation, the frame time length is set to 5ms and the length of listening window is set to one frame length. For the simulation of BTE algorithm, the initial sleep cycle is set to four times of frame length and the final sleep cycle length is set to 512 times of the initial sleep cycle. For the NEQ algorithm, we set $p = 0.5$. $N = 8$ if the traffic arrival rate is greater than 0.05, and $N = 20$ otherwise.

The distribution of inter-arrival duration of traffic follows exponential distribution. For exponential distribution, the combination still follows exponential distribution but having new arrival rate which is the sum of the arrival rates of each traffic. So, we investigate the performance with traffic arrival rate in a wide region. In our simulation, energy consumption is defined as the energy consumed during listening windows. Delay, i.e., response time, is defined as the time interval that is from the AMS entering the sleep mode to its termination of sleep mode made by packet arrival.

The energy consumptions by listening windows of BTE and NEQ schemes are compared in Fig. 5. The energy consumed by NEQ is less than that consumed by BTE under different packet arrival rate. NEQ saves about 40 to 60 percent more energy for the given range of arrival rates in the figure. For web browsing traffic, the packet inter-arrival rate is $\lambda = 0.033$. We can see from the figure that the energy consumption of NEQ is only around 40 percent of BTE. For VoIP traffic, the packet inter-arrival rate is $\lambda = 0.3$. From Fig. 5, we see that the energy consumption is a half of conventional method. For other traffic type or mixed traffic, the efficiency of energy saving can be checked by selecting the appropriate point of mixed arrival rate on the figure.

Fig. 6 shows the delay between the packet arrival and the first available listening window after the packet arrival. The delay of NEQ algorithm is about 40 to 60 percent of that of BTE in the figure. Given web browsing traffic ($\lambda = 0.033$), the delay caused by sleep cycle extension is reduced by approximate 85 percent. More importantly, the delay is reduced more than 50 percent for VoIP traffic ($\lambda = 0.3$). We also can check the delay decrease of other types of traffic or mixed traffic by the packet inter-arrival rate.

V. Conclusion

Our proposed method for energy saving in sleep mode can significantly reduce the energy consumption and frame delay caused by sleep cycle extension. This is because our method is based on the statistical attributes of various traffic which is able to approach the optimal solution by statistical technique. We illustrate our idea and mechanism by the sleep mode operation in IEEE802.16m and show the simulation results in validating the performance. Such mechanism can be also utilized in other mobile communication systems to save MS's energy and prolong the battery life.

REFERENCES

[1] "The draft IEEE 802.16m system description document," in *IEEE 802.16 Broadband Wireless Access Working Group*.

[2] "Part16: Air interface for fixed and mobile broadband wireless access systems -multihop relay specification," in *IEEE 802.16 Broadband Wireless Access Working Group*.

[3] E. Dahlman, H. Ekstrom, A. Furuskar, Y. Jading, J. Karlsson, M. Lundevall, and S. Parkvall, "The 3G long-term evolution-radio interface concepts and performance evaluation," in *IEEE 63rd Vehicular Technology Conference, 2006. VTC 2006-Spring*, vol. 1, 2006.

[4] Y. Xiao, "Energy saving mechanism in the IEEE 802.16 e wireless MAN," *IEEE Communications Letters*, vol. 9, no. 7, 2005.

[5] L. Kong and D. Tsang, "performance study of power saving classes of type I and II in IEEE 802.16 e," in *Local Computer Networks, Proceedings 2006 31st IEEE Conference on*, 2006.

[6] Y. Zhang, "Performance modeling of energy management mechanism in ieee 802.16 e mobile wimax," in *Wireless Communications and Networking Conference, 2007. WCNC 2007. IEEE*.

[7] H. Choi and D. Cho, "Hybrid Energy-Saving Algorithm Considering Silent Periods of VoIP Traffic for Mobile WiMAX," in *IEEE International Conference on Communications, 2007. ICC'07*, 2007.

[8] E. Hwang, K. Kim, J. Son, and B. Choi, "The power saving mechanism with binary exponential traffic indications in the IEEE 802.16 e/m," *Queueing Systems*.

[9] "Ieee802.16e, part16: Air interface for fixed and mobile broadband wireless access systems -amendment for physical and medium access control layer for combined fixed and mobile operation in licensed bands," in *IEEE 802.16 Broadband Wireless Access Working Group*.

[10] "Project 802.16m evaluation methodology document (EMD)," in *IEEE 802.16 Broadband Wireless Access Working Group*.

[11] P. Brandy, "Model for Generating On-Off Speech Patterns in Two-Way Conversation," *The Journal of the Acoustical Society of America*, vol. 46, 1969.