Adaptive Traffic Indication Algorithm for Energy Efficiency in IEEE 802.16e Systems

Jenhui Chen* and Yu-Lin Li
Dept. Computer Science and Information Engineering
Chang Gung University, Kweishan, Taoyuan, Taiwan, R.O.C.
Email: jhchen@mail.cgu.edu.tw

***Abstract*—The efficiency of power saving mechanism on wireless communications will influence the time the mobile stations (MSs) can operate. Due to the characteristics of centralized control in WiMAX system, the sleeping period of each subscriber is dominated by a base station (BS) based on their service types, traffic loads, and expected sleeping periods. The power saving mechanism uses an exponential backoff sleeping window manner to determine the sleeping period of each MS. In recent researches, some of them optimize the sleeping period by estimating the packet inter-arrival time for improving the energy efficient. However, those mechanisms do not reflect the relationship between the traffic load and available bandwidth. That is, according to the available and priorities of connections, the lower priority connections can not receive data immediately and waste energy on the waiting time. Thus, in this paper, we propose an adaptive traffic indication algorithm (ATIA) to let MSS do the extend sleep on bandwidth unavailable condition, and illustrate an adaptively adjusting sleeping window scheme for delay versus energy consumption. Simulation results show that ATIA increase the degree of power saving with comparison to IEEE 802.16e; and further, it shows that ATIA is capable of combining with other power saving mechanisms and improve performance.**

***Index Terms*—algorithm, power saving, scheduling, traffic, wireless**

I. Introduction

The IEEE 802.16 working group [1], [2] is originally organized to develop standards of fixed broadband wireless access (BWA) technology that provides high-speed, high-bandwidth, and high-capacity (HHH) broadband wireless access services. In recent researches [3], [4], [5], [6], [7], [8], some of them analyzed the performance of power saving in IEEE 802.16e. For example, Lei and Tsang [3] analyzed the performance of Type I and Type II and found out a decision to switch sleep mode for optimal energy management, but those mechanism will lose the advantage from the classification of power saving class. Thus, BS can not use the different sleep type to set the priority of being awaken. Han and Choi [4] found out the effect of idle interval and the final sleeping interval by performance analysis, but they discuss those operational parameter without bandwidth condition. Lee and Bahk [6] think about the handover occur in power saving would cause consume of cell searching, and design the adaptive sleep window mechanism. In their mechanism, they model the MAC state transition by four states: Adaptive, Power saving, Idle, and Doze. But they use the connection identification (CID) assignment to represent the data connection is going that is non-reality. Choi, Lee, and Cho [7] designed hybrid mechanism for voice over IP (VoIP) service but such mechanism will cause more and more packet drop. Another problem is the condition on all silence is too different that will let their mechanism lack practicality. Kim, Choi, and Kang [8] adaptive change initial and final sleeping window base on last sleeping state and correctly measure the point of awaken, but it will become no use on condition that bandwidth is unavailable.

Comparing to above researches, they always take sleep operation as an independent mechanism. So they always concentrate on mechanism optimization by adjusting sleeping interval or change sleeping process. Based on IEEE 802.16e standard, after receiving the positive traffic indication message, MS would breakup power saving operation and stay idle until the MS be scheduled. In other word, if it does not consider about the bandwidth allocation, it would become so ideality in real condition. Consequently, we design a mechanism to make an accurate decision for doing traffic indication to optimize energy management.

In this paper, we design a traffic indication algorithm combining power saving with scheduler operation. In IEEE 802.16 systems, scheduler defined five priority types for different services; the priority affects how to allocate bandwidth. We force on the Type I services that are recommended for connections of BE and NRT type. For BE and NRT services, the low priority will cause that scheduler cannot allocate enough bandwidth to satisfy those services in time. Overall, our mechanism design for reduce both the time of that BS interrupt MSs's sleep time and the energy consume of MSs.

II. System Model

A. Network Configuration

The system model is based on the IEEE 802.16e standard with power saving class I and II. Multiple MSs are connected to the BS over wireless channels, where multiple connections can be supported by each MS. These connections consist of m RT-VR services and n NRT services, denoted as C_{rt} and C_{nrt}. For simplicity, each MS only occupies one connection for C_{rt} and one for C_{nrt}, respectively. All connections communicate with the BS using time division multiplexing/time-division multiple access (TDM/TDMA) mechanism for data

*This work was supported in part by the National Science Council, Taiwan, R.O.C., under Contract NSC97-2221-E-182-035-MY2.

CHINACOM 2010, August 25-27, Beijing, China

DOI 10.4108/chinacom.2010.47

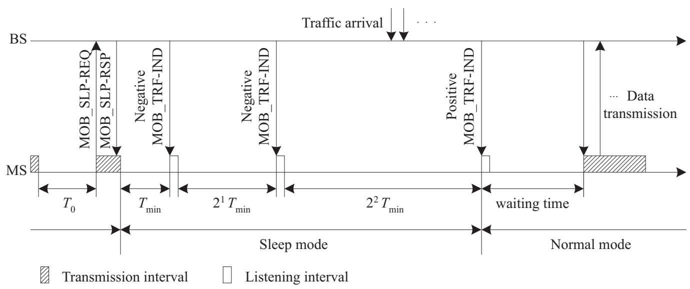

Fig. 1. The type I power saving operation in IEEE 802.16e standard.

transmissions. This paper will focus on the downlink, although the results can be extended to the uplink as well. A buffer is implemented at the BS for each connection and operates in a first-input-first-output (FIFO) manner.

The traffic of all connections is treated as a random variable, and is assumed as Poisson with mean arrival rate λ_R for RT-VR traffic and λ_N for NRT traffic. It satisfies that $\Lambda = \lambda_R + \lambda_N$ is the total traffic arrival rate of the system. The data lengths of RT-VR and NRT services, denoted as ℓ_R and ℓ_N, are generally distributed with exponential distribution.

B. Type I Power Saving Mechanism

The type I power saving mechanism (PSM), shown in Fig. 1, which is recommended for low priority connections such as NRT or BE types, follows the exponential sleeping window size mechanism for adjusting the duration of sleeping intervals. At first, each MS will enter the power saving (PS) mode if there is no traffic arriving for it during a time period T_0. The duration of T_0 is a system parameter and is preferred as 32 frames long in [5]. For efficiency reasons, the sleeping interval employs a discrete-time frame scale. The time immediately following an idle T_0 is framed and an MS is waked up for checking traffic arrival only after at the end of each sleeping interval T_i for $1 \leq i \leq w$ and $\forall i \in \mathbb{N}$, where w is the maximum number of sleeping states. The value T_i is called sleeping window, and depends on the probability of packet arrival. Initially, T_i is set equal to a value $W_{\min}$ called minimum sleeping window. After each successful sleep, T_i is doubled up until the maximum value $W_{\max}$ when $2^i W_{\min} \geq W_{\max}$.

The value of parameters $W_{\min}$ and w are coordinated by MOB_SLP-REQ and MOB_SLP-RSP messages. Once there is no traffic during a time period exceeds T_0, the MS will transform its state from the operational state S_0 into the PS mode for conserving the energy as shown in Fig. 2. The PS mode will be continued unless a packet arrives during any T_i for $1 \leq i \leq w$. Any data arrives at this interval, the BS will buffer these data and sends a positive MOB_TRF-IND message to wake up the MS at the upcoming listening interval denoted as T_L.

Let S_S denote the set of power saving states and $S_S = \{S_1, S_2, \ldots, S_w\}$. All elements of S_S follows the sequence

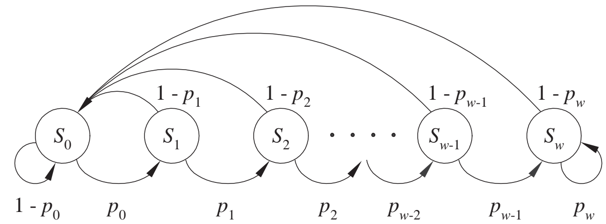

Fig. 2. The state transition diagram of PSM.

order by $S_1 \prec S_2 \prec \ldots \prec S_w$. The transition probability from S_i to S_{i+1}, which is denoted by p_i, is the probability that there is no packet arriving during $T_i = t_{i+1} - t_i$ where t_i denotes the time the MS entering S_i and is given by

$$p_i = P[S_{i+1}|S_i] = e^{-\lambda_N T_i}, \quad 0 \leq i \leq w, \tag{1}$$

and T_i expresses different values as

$$T_i = \begin{cases} T_0, & i = 0 \\ 2^{i-1} W_{\min} + T_L, & 1 \leq i \leq w \end{cases}. \tag{2}$$

We note that when $i = 0$, it indicates that the MS is in S_0 and T_0 is a system parameter.

Based on Markov process as fig. 2, the steady state probability of each S_i, denoted by $P(S_i)$, only depends on its previous state S_{i-1} and is obtained by

$$P(S_i) = \begin{cases} \sum_{j=1}^{w} P(S_j)(1-p_j), & i = 0 \\ P(S_{i-1}) p_{i-1}, & 1 \leq i < w \\ \dfrac{P(S_{w-1}) p_{w-1}}{1 - p_w}, & i = w \end{cases}. \tag{3}$$

According to the axiom of probability theory, it satisfies

$$\sum_{i=0}^{w} P(S_i) = 1. \tag{4}$$

Substituting Eq. (3) into Eq. (4) leads to

$$P(S_0) + \sum_{i=1}^{w-1} P(S_{i-1}) p_{i-1} + \frac{P(S_{w-1}) p_{w-1}}{1 - p_w} = 1. \tag{5}$$

Solving Eq. (5) for $P(S_0)$, we have

$$P(S_0) + \sum_{i=1}^{w-1} \left(P(S_0) \prod_{j=0}^{i-1} p_j \right) + \frac{P(S_0) \prod_{i=0}^{w-1} p_i}{1 - p_w} = 1 \tag{6}$$

and

$$P(S_0) \left(1 + \sum_{i=1}^{w-1} \prod_{j=0}^{i-1} p_j + \frac{\prod_{i=0}^{w-1} p_i}{1 - p_w} \right) = 1. \tag{7}$$

Obtained $P(S_0)$ from Eq. (7), then

$$\begin{aligned} P(S_0) &= \left(1 + \sum_{i=1}^{w-1} \prod_{j=0}^{i-1} p_j + \frac{\prod_{i=0}^{w-1} p_i}{1 - p_w} \right)^{-1} \\ &= \left(1 + \sum_{i=1}^{w-1} \prod_{j=0}^{i-1} e^{-\lambda_N T_j} + \frac{\prod_{i=0}^{w-1} e^{-\lambda_N T_i}}{1 - e^{-\lambda_N T_w}} \right)^{-1} \end{aligned} \tag{8}$$

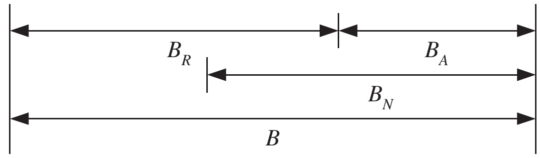

Fig. 3. The bandwidth allocation between B_N and B_R.

Each $P(S_i)$ can be obtained by $P(S_0)$ from Eq. (3).

C. MAC Resource with AMC Scheme

Efficient bandwidth utilization for a prescribed packet error rate (PER) performance at the physical layer (PHY) can be accomplished with adaptive modulation and coding (AMC) schemes, which match transmission parameters to the time-varying wireless channel conditions adaptively. Assume there are l different AMC modulation levels supported in the IEEE 802.16e PHY. The achievable modulation depends on its corresponding received SNR value E_b/N_0 and denoted as μ_i, where i indicates the i-th AMC modulation level and $1 \leq i \leq l$. Let s_i be the minimum required E_b/N_0 for achieving the i-th AMC modulation level. Then, the achieved μ_i in the i-th AMC level satisfies $s_i \leq E_b/N_0 < s_{i+1}$ where $s_{i+1} = \infty$ for i using the highest level of AMC modulation. Let $\mu_{i,j}$ denote the achievable modulation for the i-th MS with the j-th AMC level and be given by

$$\mu_{i,j} \triangleq \{\mu_j | \ s_j \leq x_i < s_{j+1}\} \tag{9}$$

where x_i represents the received E_b/N_0 from i-th MS via the channel quality information channel (CQICH).

According to the achievable μ_i, each available orthogonal frequency-division multiple access (OFDMA) slot can convey the corresponding bit rate varies from time by time in downlink subframe. Let b_i denote one OFDMA slot can transmits the number of bit in μ_i, and can be represented as

$$b_i = n_s N_b(i) c_r \tag{10}$$

where n_s is the number of used data subcarrier in an OFDMA-slot, $N_b(i)$ is the bit volume (bits/subcarrier) in μ_i, and c_r is the selected-correspondingly coding rate. According to the IEEE 802.16e PHY with PUSC, one cluster has 24 data subcarriers, and one OFDMA slot occupies 2 symbols with partitioning into 2 cluster. Hence, there are 48 data subcarriers within one OFDMA slot, $N_b(i)$ depending on the modulation levels has the different capacities of transition bit (2 bits/subcarrier for QPSK modulation, 4 bits/subcarrier for 16-QAM modulation, and 6 bits/subcarrier for 64-QAM modulation), and c_r has 3 types of cording rate: 1/2, 2/3, and 3/4.

III. Adaptive Traffic Indication Algorithm

The above mentioned PSM uses the MOB_TRF-IND message to wake up an MS if it is in the PS mode and there is a traffic for it. Once an MS wakes up from PS mode, the BS has to allocate bandwidth to the MS for data transmission. However, when an MS wake up, the energy will be wasted if there is no available bandwidth for that MS. To avoid this drawback, we design an adaptive traffic indication (ATIA) algorithm combined with packet scheduling scheme to reduce the energy consumption.

First, ATIA separates the operational state S_0 into two states S_N and S_E. S_N represents the normal operational state and S_E represents the extended PS state. When an MS is in S_N, it can receive data from the BS if there is a bandwidth for it. Otherwise, the MS will be notified to enter S_E for extended power saving.

A. The Available Bandwidth Allocation

Consider m RT-VR and n NRT connections in the system and the priority of C_{rt} is higher than that of C_{nrt}. ATIA follows the priority order to allocate available bandwidth for connection usage. Therefore, the available bandwidth $B_A = B - B_R$, shown as Fig. 3, can be used for allocation to C_{nrt}, where B is the total bandwidth of the system measured in OFDMA slots and $B_R = \sum_{i=0}^{m-1} \lambda_R(i)\ell_R(i)/b_i$ is the total required bandwidth of C_{rt} per each subframe.

If B_A is enough for C_{nrt}, i.e., $B_A \geq B_N$, where $B_N = \sum_{i=0}^{n-1} \lambda_N(i)\ell_N(i)/b_i$ represents the total required bandwidth of C_{nrt}, all C_{nrt} will be served in time in the current subframe. Otherwise, the bandwidth is not enough for supporting all C_{nrt} to be served in the current subframe, i.e., $B_A < B_N$, ATIA will determine some C_{nrt} prior to serve.

B. The Selection Criteria

The criterion of connection selection for bandwidth allocation follows the priority of each connection by its corresponding queueing length $Q_i(t)$ and waiting times in S_E denoted as ω. The priority order of each connection W_i is formulated as

$$W_i = Q_i(t)/\omega \tag{11}$$

where $\omega = 1, 2, \ldots, \infty$ and $\omega \in \mathbb{N}$. The value of $Q_i(t)$ is time-vary and can be calculated following Little's Formula [9] by

$$Q_i(t) = \lambda_N(i)\delta + Q_i(t-1) \tag{12}$$

where δ represents the duration time of a subframe. The ATIA simply selects the lowest value of W_i as the highest priority connection

$$i = \arg\min_i W_i. \tag{13}$$

Therefore, if two connections C_i and C_j have $W_i < W_j$, then C_i is first selected for serving until all B_A is exhausted.

Before a C_i enters S_E, it may stays in either S_S or S_N as shown in Fig. 4. There are two circumstances that a C_i will enter S_E.

- C1: The C_i is waked up from S_S and there is no available bandwidth for the C_i. The sleeping time duration in S_E denoted as T_e is a system parameter and is analyzed later.
- C2: The C_i is in S_N and there is no available bandwidth for the C_i, e.g., $B_A = 0$. In this case, only higher priority connections will be selected out to be served.

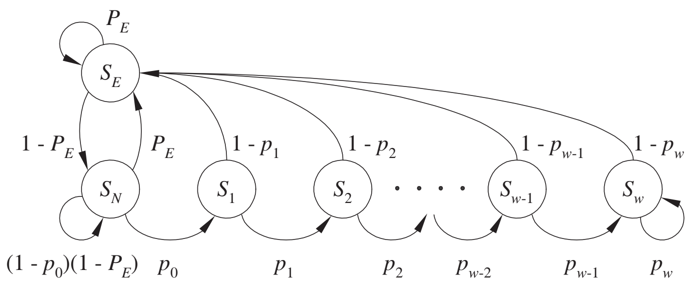

Fig. 4. The state transition diagram between S_N and S_E.

C. Performance Analysis

The probabilities of S_E and S_N denoted as $P(S_E)$ and $P(S_N)$ satisfy

$$P(S_N) + P(S_E) = P(S_0) \tag{14}$$

and $P(S_E)$ is equal to

$$P(S_E) = P(S_N)P_E + P(S_E)P_E \tag{15}$$

where $P_E = p_A \cdot p_E$ is the conditional probability that nodes will be selected into S_E when $B_A < B_N$. Let p_A be the probability when $B_A < B_N(T_e)$ where $B_N(T_e)$ represents the total required bandwidth during T_e and p_E be the probability of C_i being selected to be served. Assume the guaranteed bandwidth of each allowed C_{nrt} in the system is B_g, then the number of C_{rt} and C_{nrt} will demand the bandwidth after a duration T_e denoted as k_R and k_N, respectively, is given by

$$k_R = m(1 - e^{-\lambda_R T_e}) \tag{16}$$

and

$$k_N = nP(S_0)(1 - e^{-\lambda_N T_e}). \tag{17}$$

Thus the condition of no enough B_A for C_{nrt} satisfies

$$\frac{B - k_R \ell_R}{B_g} < k_N = nP(S_0)(1 - e^{-\lambda_N T_e}) \tag{18}$$

and solving Eq. (18) for λ_N we have

$$\lambda_N > \ln\left(\frac{B - k_R \ell_R - nP(S_0)B_g}{nP(S_0)B_g}\right) \Big/ T_e. \tag{19}$$

According to (19), the probability of no enough bandwidth for all C_{nrt}, p_A, is calculated as

$$p_A = P[k_N > \lambda_N T_e] = 1 - \sum_{x=0}^{\lambda_N T_e} \frac{(\lambda_N T_e)^x}{x!} e^{-\lambda_N T_e}. \tag{20}$$

As in the condition, p_E is the probability of turn to S_E and can be expressed as

$$p_E = 1 - \frac{B - k_R \ell_R}{k_N B_g}. \tag{21}$$

Then, we can substitute Eq. (15) into Eq. (14) and represent it as

$$P(S_N) + \frac{P(S_N)P_E}{1 - P_E} = P(S_0) \tag{22}$$

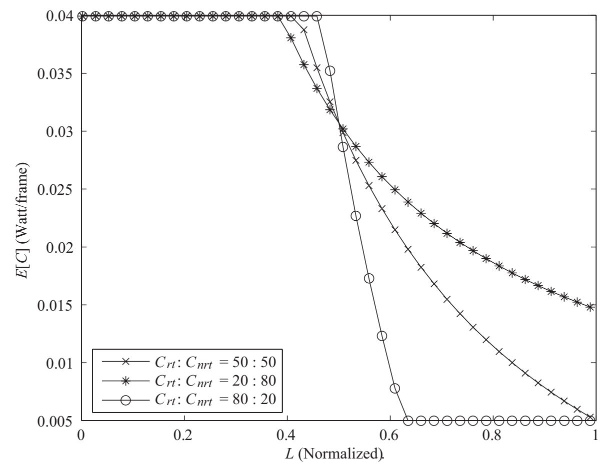

Fig. 5. Energy consumption vs. real-time to non-real-time traffic ratios.

where $P(S_N)$ can be expressed as

$$P(S_N) = P(S_0)(1 - P_E). \tag{23}$$

The average energy consumption of ATIA denoted as $E[C]$ can be calculated as $E[S_0] + E[S]$, where $E[S_0]$ and $E[S]$ are the average energy consumption in the state S_0 and S_S, respectively. Therefore, $E[S_0]$ can be expressed as

$$\begin{aligned} E[S_0] \quad = \quad & P(S_0)\Big[P(S_N)\Big(E_T T_e \lambda_N / B_g + E_L T_0\Big) \\ & + P(S_E) T_e E_S\Big] \Big/ T_{S_0}, \end{aligned} \tag{24}$$

where T_{S_0} is the mean time interval of S_0 and can be obtained by

$$T_{S_0} = P(S_N)\left(T_e \lambda_N / B_g + T_0\right) + P(S_E)T_e \tag{25}$$

and E_T and E_L are the energy consumption in transmission and listen states. $E[S]$ represents the average energy consumption and can be expressed as

$$E[S] = (1 - P(S_0)) \frac{\sum_{i=1}^{w} P(S_i)(T_i E_S + T_L E_L)}{\sum_{i=1}^{w} P(S_i)(T_i + T_L)}. \tag{26}$$

Finally, by summing Eq. (24) and (26), the total energy consumption can be obtained from Eq. (III-C).

D. Analysis Result

First, we observe energy consumption on the condition that affect by different ratio of bandwidth requirement between C_{rt} and C_{nrt} and show it as Fig. 5. When the ratio as 20:80, C_{nrt} has more connections can be selected to do extend sleep, then it will decrease the energy consumption first. Contrary to high ratio of C_{nrt}, the high ratio of C_{rt} will decrease the number of connection to do extend sleep on low traffic load. But when total traffic increase, p_A will increase more quickly than B_N has high ratio.

$$E[C] = \frac{P(S_0)\Big[P(S_N)\Big(E_T T_e \lambda_N / B_g + E_L T_0\Big) + P(S_E) T_e E_S\Big]}{P(S_N)\left(T_e \lambda_N / B_g + T_0\right) + P(S_E) T_e} + (1 - P(S_0)) \frac{\sum_{i=1}^{w} P(S_i)(T_i E_S + T_L E_L)}{\sum_{i=1}^{w} P(S_i)(T_i + T_L)}. \tag{27}$$

TABLE I
PARAMETERS USED IN OUR SIMULATION.

Parameter	Value
System Bandwidth	20 MHz
Sub-Channels	30
Frame Duration	5 ms
Idle time	32
Initial sleeping window	1
Maximum sleeping window	64
Simulation Time	300 s
E_S, E_L, E_I, E_T	50,170,750,1000 mW

IV. SIMULATION

The system parameters are shown in Table I [5]. In simulation, we observe the different traffic load condition to compare our performance under OFDMA system. The NRT service has the constant packet size of 1024kb and the delay bounded 500 ms. The traffic arrival follows Poisson distribution with $\lambda = 0.02$ per frame. We exponentially set the video service that the required QoS delay within 150ms and the data packet length from 5120 to 15360kb. The packet arrival rate follows Poisson distribution with $\lambda = 0.02$. Because we force on observe the power saving on NRT services, we do not support ATIA for RT-VB services. With scheduler, we first process the EDF [10] on RT-VB service if the bandwidth is enough then to do round robbin on NRT service.

Fig. 6 illustrates the comparison of average energy consumption per frame. When the total traffic load is equal as 45%, the bandwidth start not enough support all of the MS at all times. In our scheme, it also start turn to do selection indication. And show the energy consumption decrease progressively at that point. The more RT-VB service user, the bandwidth become more inefficiency.

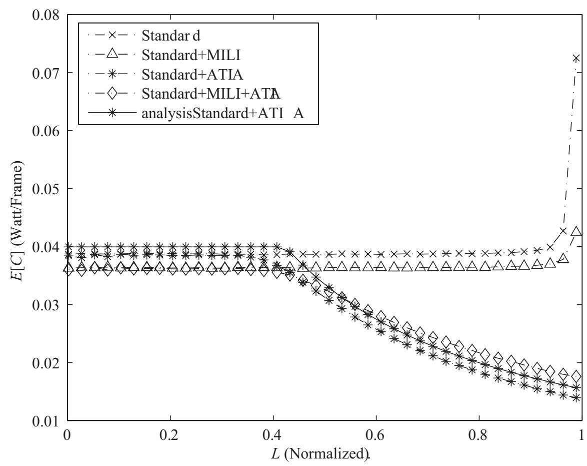

Fig. 6. Simulation of energy consumption among different bandwidth condition.

V. CONCLUSION

This paper addresses a problem of an MS will waste its battery energy if there is no available bandwidth for it after its wake up. To avoid this problem, this paper proposes the ATIA to accurately handle all MSs to receive their data without unwanted waiting in the operation mode. This manner uses an extended sleep mode to keep energies of MSs from wasting in idle mode. To avoid causing the starvation and delay of receiving packets, ATIA adopts an adaptive sleeping window to take care packet drop probability and delay. In simulation, we combine our scheme with standard power saving mechanism in round robbin scheduler to present our scheme efficiently reduce energy consumption. There is another advantage of ATIA that it allows more MSs into the system since ATIA reduces the transmission delay as the traffic load is high.

REFERENCES

[1] IEEE 802.16 Working Group, "IEEE Standard for Local and Metropolitan Area Networks–Part 16: Air Interface for Fixed Broadband Wireless Access Systems," IEEE Std. 802.16-2004, October 2004.

[2] IEEE Std 802.16e-2005, "Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems," February 2006.

[3] L. Kong and D.H.K. Tsang, "Performance Study of Power Saving Classes of Type I and II in IEEE 802.16e," in *Proc. IEEE Local Computer Networks*, pp.20–27, Tampa, Florida, November 2006.

[4] H. Kwanghun and C. Sunghyun, "Performance Analysis of Sleep Mode Operation in IEEE 802.16e Mobile Broadband Wireless Access Systems," in *Proc. IEEE VTC-Spring 2006*, pp.1141–1145, Melbourne, Australia, May 2006.

[5] Z. Yan, "Performance Modeling of Energy Management Mechanism in IEEE 802.16e Mobile WiMAX," in *Proc. IEEE WCNC 2007*, pp.3205–3209, Hong Kong, March 2007.

[6] N.-H. Lee and S. Bahk, "MAC Sleep Mode Control Considering Downlink Traffic Pattern and Mobility," in *Proc. IEEE VTC-Spring 2005*, pp.2076–2080, Stockholm, Sweden, June 2005.

[7] H.-H. Choi, J.-R. Lee, and D.-H. Cho, "Hybrid Power Saving Mechanism for VoIP Services with Silence Suppression in IEEE 802.16e Systems," *IEEE Commun. Letters*, pp.455–457, May 2007.

[8] M.G. Kim, J.Y. Choi, and M. Kang, "Adaptive Power Saving Mechanism Considering the Request Period of Each Initiation of Awakening in the IEEE 802.16e System," *IEEE Commun. Letters*, vol. 12, no. 2, February 2008.

[9] W. Stallings, *High-Speed Networks and Internets: Performance and Quality of Service*, Prentice Hall, 2002.

[10] L. Georgiadis, R. Guerin, and A. Parekh, "Optimal Multiplexing on a Single Link: Delay and Buffer Requirements," *IEEE Trans. Information Theory*, vol. 43, no. 5, pp.1518–1535, September 1997.

[11] S.H. ALi, K. Lee, and V.C.M. Leung, "Dynamic Resource Allocation in OFDMA wireless Metropolitan Area Networks," *IEEE Wireless Commun. Mag.*, vol. 14, no. 1, pp.6–13, February 2007.