Performance Analysis of Load Based Energy-Saving Strategies for Real Time Traffic in IEEE 802.16e

V. Goswami, C Misra
School of Computer Application,
KIIT University,
Bhubaneswar 751024,
INDIA
veena_goswami@yahoo.com, cmisra@yahoo.com

Abstract. In wireless and mobile telecommunication systems, the handheld mobile devices completely sustains on lightweight batteries for power supply. This mobile station dissipates lots of energy if it listens to base station continuously for intended messages. This paper presents a discrete time renewal input single and batch service queueing system with accessibility to the batches which is capable of calculating the power efficiency and packet access delay for the power saving type II. A cost model is developed to determine the optimum service rates which minimize the total expected cost. Various performance measures are provided based on this analysis. It is shown numerically that the proposed model can work efficiently to improve the system performance and establish the dependency relationships among the system performance measures and the system parameters.

Key-Words: IEEE 802.16e, downlink, sleep mode, accessible batch, queueing

1. Introduction

During the past decades, the rapid growth of high-speed multimedia services has created high demand for the mobile handheld devices including cellular phones for broadband access. WiMAX (Worldwide Interoperability for Microwave Access) is cost effective broadband access solution compared to other access network technology. IEEE 802.16 based technology, also known as WiMAX, is rapidly gaining popularity because of open standard, high performance, flexibility for supporting a number of users and sophisticated support for Quality-of-Service (QoS) at MAC layer for different categories of services [1]. With advantages of high transmission rate and predefined QoS model, IEEE 802.16 standard and its evolutions have been designed for fulfilling the requirements of very-high-data-rate applications, such as voice over IP (VoIP), video or audio streaming, as well as low-data-rate applications, such as web surfing.
IEEE 802.16e systems offer full-mobility for WiMAX and supports seamless handoff which renders switching between base stations in vehicular speeds without interrupting the connection [2]. The Mobile Stations (MSs) are fully dependent on light weight batteries for maintaining the connections with the serving base stations (BS). Therefore power saving is an important issue in recent years. The IEEE 802.16 standard family adopt power saving mechanism by discontinuing connection for a pre negotiated time with the BS station as shown in Figure1. The listening and sleep windows repeat every sleep cycle as long as the MS is in the sleep mode. It can be anticipated that if the mobile devices choose for longer sleep duration, energy can be saved but the delay in the buffer will be high. On the other hand, shorter sleep duration check traffic arrival frequently and unnecessarily energy will be consumed. Hence, there is a tradeoff between the power saving and the delay performance. The main issue is to maximize the power saving, while maintaining the required Quality of Service (QoS) on delay. IEEE 802.16e work-group has standardized those mechanisms that would augment battery lifetime in a WiMAX network without affecting the QoS performance.
Three types of power saving classes are defined primarily based on sleep mode operation which makes sleep scheduling more flexible. A fundamental difference of the three standard classes is the pattern of sleep, which determines the size of

ISSN 2067 – 4074

sleep windows succeeding the initial sleep window in the case of long idling. Type I class is recommended for non-real-time variable rate service or best effort (BE) service connections with an exponential increase of the sleep window size. Type II class is recommended for the Unsolicited Grant Service (UGS) and the Real Time Variable Rate (RT-VR) traffic with a constant sleep window size. Type III is recommended for multi-cast and management connections with the sleep window size controlled by the base station (BS).

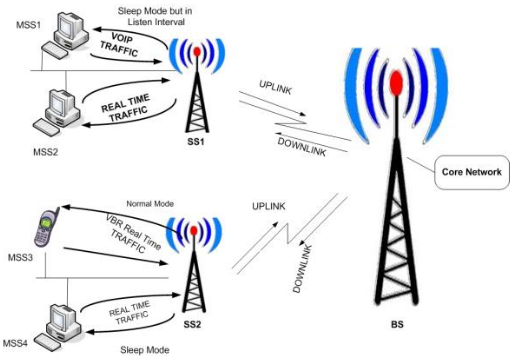

Figure 1. Infrastructure of IEEE802.16

We design a discrete-time based analytical model for delay sensitive real time traffic that pro-actively controlling the situation based on the number of packets waiting in the BS buffer. In this proposed model the power is saving and QoS both can be tradeoff effectively by providing single and batch service depending on the average numbers of packets waiting in the buffer. To minimize the delay in the buffer accessibility to the batches also incorporated. Discrete-time queueing systems are more accurate and effective in modeling frame-based systems to evaluate performance measures. Furthermore, the modeling and analysis of discrete-time queues is more complicated and quite different from the analysis applied for the corresponding continuous-time queueing models. One can obtain the continuous-time results from the discrete-time queues as a limiting case but the converse is not true. However, from theoretical and applied viewpoint both discrete- and continuous-time queueing systems have importance.
The rest of the paper is organized as follows. A brief overview of some related work is given in Section 2. An analytical model of the proposed downlink resource management framework to improve the power efficiency is presented in Section 3. Section 4 evaluates the performance metrics of the power saving class of type II for the proposed model. Section 5 discusses cost analysis and numerical results in the form of tables and graphs. Finally, Section 6 concludes the paper.

2. Related Work

Most of the performance analysis of the IEEE 802.16e sleep mode assumes downlink traffic. To investigate the energy consumption in IEEE 802.16e by considering the message delivery in both directions has been proposed in [3]. Chen et al. [4] presented two power saving class (PSC) management scheme in real time traffic scenario. They demonstrated their results using simulation and shown that resource utilization is better from power saving class I. Performance analysis of an energy saving mechanism in the IEEE 802.16e broadband wireless access network have been studied in [5,6]. They considered listening interval as a part of sleep window, but in reality listening interval consumes lots of energy as frequently it checks up buffering status.
The sleep mode operation as an $M/GI/1/N$ queue with multiple vacations has been discussed in [7]. A semi-Markov chain is proposed in [8] to analyze the performance of power saving class I under given traffic arrival patterns. A simulation-based performance evaluation of the sleep mode with various traffic types for efficient energy saving has been implemented in [9]. Recent study in [10], develops an analytical model based on the generalized traffic process. The results indicate that various traffic arrival processes generate significant discrepancy in the energy consumption and packet delay. Performance analysis of a renewal input bulk service queue with accessible and non-accessible batches in a continuous-time has been studied in [11]. Load based power saving scheme for power saving class III has been studied in [12]. The traffic load at BS is measured

when the buffered traffic reaches the threshold value; the size of the sleep window is estimated for mobile subscriber station. For efficiently managing the RT-VR traffic, this scheme is not suitable because in case of light traffic the buffered delay will be higher.

An analytical model that shows the relationship between traffic load at BS and service rate of the traffic at MSs has been discussed in [13]. They used an $M/G/1$ queueing model to find a behavior of an BS and MS and incorporate a cost model to optimize the cost. A model consists of a central BS and an MS with real time connections has been proposed in [14]. Their successive scheduling approach minimizes power consumption of the MS. Listening interval spreading approach (LISA), which redistributes the idle periods and a number of interim listening periods in order to reduce the response time of interactive traffic has been discussed in [15]. They shown that the LISA scheme provides a power management solution for handling bursty traffic such as HTTP application, but this scheme is not suitable for handling light traffic.

In this paper, we propose an analytical model to maximize the energy saving for an MS with real time downlink traffic, following the notions of IEEE 802.16e sleep-mode operation. The proposed queueing model determines the length of each sleep window for the present traffic state with the consideration of tolerable delay and number of packets waiting in the buffer. An analytical traffic model with finite buffer discrete-time renewal input queueing system is carried out, where service time is geometrically distributed. We process small traffic in lower service rate in listening interval without interruption of sleep interval. But for heavy traffic, the downlink is done in batch processing with higher service rate. Moreover, for better response time, incoming download streams can be included in the buffer up to certain limit.

3. System Model

In this section, we discuss an analytical model of power saving for real time traffic of variable bit rate (RT-VBR). PSCs of type II are proposed and recommended for connections of UGS, RT-VBR service. Type II uses constant sleep window size instead of exponential sleep window size as depicted in Figure 2. The mobile sleep request (MOB SLP-REQ) message of type II only contains two parameters, that is, the size of initial sleep window and listen interval.

Downlink data intended for MS accumulated in the buffer is located at BS and based on the buffer size, it dynamically decides the type of the service provided that is, service is given individually in listening interval or in batch with different service rate as shown in Figure 3. Moreover, for better response time, incoming download streams can be included in the buffer up to certain limit. The performance evaluation is carried out. Numerical results demonstrate that the proposed model outperforms the IEEE 802.16e PSC II scheme and the inferred IEEE802.16e power-saving mechanism.

Under the 802.16e sleep-mode operation, an MS starts to sleep for a fixed amount of time, called sleep window, and wakes up in listen mode to find if the BS has any buffered downlink traffic destined to itself. If there is no such traffic, it switches to the sleep mode, and then checks with the BS again when it wakes up in the listen mode. In power saving class II, MS can communicate with BS and transfer small data amount at listening interval. The BS is at the best position to measure the traffic load and the related operational parameters can be negotiated between the MS and BS. We assume that the downlink traffic can be received in the MS during the listen interval and continues the processing of downlink traffic at the extended listen interval individually and at lower rate μ_1 and till the downlink data traffic is less than a control limit c. If the buffer size is equal or exceed to control limit c the downlink data will be processed at MS in batch and the processing of downlink traffic will be μ_2. It is further assumed that the late entries can join a batch in course of ongoing service as long as the number of packets in that batch is less than d (called maximum accessible limit).

The inter-arrival times of successive packets are independent and identically distributed (i.i.d.) random variables with probability mass function $a_i = P(A_n = i)$, $i \geq 1$, corresponding probability generating function $A(z) = \sum_{i=1}^{\infty} a_i z^i$ and mean inter-arrival packet time $1/\lambda = A^{(1)}(1)$, where $A^{(n)}(d)$ is the n -th derivative of $A(z)$ with respect to z at $z = d$. At every departure epoch, that is, before initiating service, the server may find the system in any one of the following three cases: (i) $0 \leq n \leq c$, (ii) $c+1 \leq n \leq d\text{-}1$ and (iii) $n \geq d$. In case (i), the server serves one packet at a time. In case (ii), the server takes the entire queue for batch service and admits the subsequent packets in the batch while the service is on, till the accessible limit d is reached, and such a batch is called an accessible batch (AB). In case (iii), it takes all the packets for the service and does not allow accessibility to the batch, that is, when the batch size is greater than or equal to d , the batch becomes non-

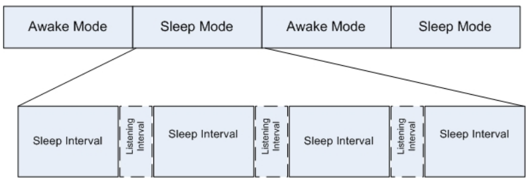

Figure 2. Awake and sleep mode in power saving class II.

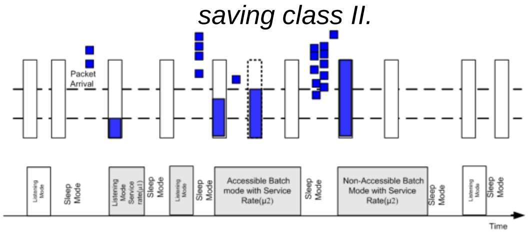

Figure 3. Effect of load based proposed power saving approach for real time traffic.

accessible (NAB) for late arriving packets. The service times S_1 in case of (i) are independent and geometrically distributed with mean μ_1^{-1}. The service times S_2 of batches in case (ii) and (iii) are independent and geometrically distributed with mean μ_2^{-1}.

Let us assume that the time axis is marked with $0,1,2,\ldots,t,\ldots$ and slotted into intervals of unit length. Potential arrivals occur in $(t\text{-},t)$ and potential departures occur in $(t, t+)$. The state of the system just before a potential arrival (at t-) is described by the following random variables:

- $N_s(t\text{-})(N_q(t\text{-}))$ = number of customers present in the system (queue),
- $U(t\text{-})$ = remaining inter-arrival time for the next arrival,

$$\zeta(t\text{-}) = \begin{cases} 0, \text{server is idle or busy with single service}, \\ 1, \text{server is busy with an accessible batch}, \\ 2, \text{server is busy with a non-accessible batch}. \end{cases}$$

Let us define the joint probabilities for $u \geq 0$, as

$$\pi_{i,0}(u, t-)du = Pr\{N_s(t-) = i, U(t-) = u, \zeta(t-) = 0\},\ 0 \leq i \leq c,$$
$$\pi_{i,1}(u, t-)du = Pr\{N_s(t-) = i, U(t-) = u, \zeta(t-) = 1\},\ c+1 \leq i \leq d-1,$$
$$\pi_{i,2}(u, t-)du = Pr\{N_q(t-) = i, U(t-) = u, \zeta(t-) = 2\},\ i \geq 0.$$

At steady-state, the above probabilities are denoted by , $\pi_{i,0}(u)$, $\pi_{i,1}(u)$ and $\pi_{i,2}(u)$ and their probability generating functions are $\pi^*_{i,0}(z)$, $\pi^*_{i,1}(z)$and $\pi^*_{i,2}(z)$ respectively.

To obtain system length distribution at arbitrary epoch, we develop the difference equations using the remaining inter-arrival time as the supplementary variable. Relating the states of the system at two consecutive epochs $t-$ and $(t+1)-$and using the definitions and probabilities defined above, we obtain the following equations, in the steady-state, for $u \geq 1$,

$$\pi_{0,0}(u-1) = \pi_{0,0}(u) + \mu_1 \pi_{1,0}(u) + \mu_2 \pi_{0,2}(u) + \mu_2 \sum_{n=c+1}^{d-1} \pi_{n,1}(u), \quad (1)$$

$$\pi_{1,0}(u-1) = \bar{\mu_1}\pi_{1,0}(u) + a_u \pi_{0,0}(0) + \mu_1 \pi_{2,0}(u) + \mu_2 \pi_{1,2}(u) + \mu_2 a_u \sum_{n=c+1}^{d-1} \pi_{n,1}(0) + \mu_2 a_u \pi_{0,2}(0) + \mu_1 a_u \pi_{1,0}(0), \quad (2)$$

$$\pi_{n,0}(u-1) = \bar{\mu_1}\pi_{n,0}(u) + \mu_1 \pi_{n+1,0}(u) + \mu_2 \pi_{n,2}(u) + a_u \bar{\mu_1} \pi_{n-1,0}(0) + \mu_2 a_u \pi_{n-1,2}(0) + \mu_1 a_u \pi_{n,0}(0),\quad 2 \leq n \leq c-1, \quad (3)$$

$$\pi_{c,0}(u-1) = \bar{\mu_1}\pi_{c,0}(u) + \mu_2 \pi_{c,2}(u) + \mu_1 a_u \pi_{c,0}(0) + \bar{\mu_1} a_u \pi_{c-1,0}(0) + \mu_2 a_u \pi_{c-1,2}(0), \quad (4)$$

$$\pi_{c+1,1}(u-1) = \bar{\mu_2}\pi_{c+1,1}(u) + \bar{\mu_1} a_u \pi_{c,0}(0)$$

$$+\mu_2\pi_{c+1,2}(u)+\mu_2 a_u\pi_{c,2}(0), \quad (5)$$

$$\pi_{n,1}(u-1)=\bar{\mu_2}\pi_{n,1}(u)+\mu_2\pi_{n,2}(u) +\bar{\mu_2}a_u\pi_{n-1,1}(0)+\mu_2 a_u\pi_{n-1,2}(0), \quad c+2\le n\le d-1, \quad (6)$$

$$\pi_{0,2}(u-1)=\bar{\mu_2}\pi_{0,2}(u)+\mu_2 a_u\sum_{n=d-1}^{\infty}\pi_{n,2}(0) +\bar{\mu_2}a_u\pi_{d-1,1}(0)+\mu_2\sum_{n=d}^{\infty}\pi_{n,2}(u) \quad (7)$$

$$\pi_{n,2}(u-1)=\bar{\mu_2}\pi_{n,2}(u) +\bar{\mu_2}a_u\pi_{n-1,2}(0), n\ge 2. \quad (8)$$

Multiplying (1) to (8) by z^u and integrating with respect to u from 0 to ∞ yields

$$z\pi^*_{0,0}(z)=\pi^*_{0,0}(z)+\mu_1\pi^*_{1,0}(z)+\mu_2\sum_{n=c+1}^{d-1}\pi^*_{n,1}(z) +\mu_2\pi^*_{0,2}(z)-\pi_{0,0}(0)-\mu_1\pi_{1,0}(0) -\mu_2\sum_{n=c+1}^{d-1}\pi_{n,1}(0)-\mu_2\pi_{0,2}(0), \quad (9)$$

$$z\pi^*_{1,0}(z)=\bar{\mu_1}\pi^*_{1,0}(z)+A^*(z)\pi_{0,0}(0) +\mu_1\pi^*_{2,0}(z)+\mu_2\pi_{1,2}(z) +A^*(z)\mu_2\sum_{n=c+1}^{d-1}\pi_{n,1}(0)+A^*(z)\mu_2\pi_{0,2}(0) +A^*(z)\mu_1\pi_{1,0}(0)-\bar{\mu_1}\pi_{1,0}(0)-\mu_1\pi_{2,0}(0) -\mu_2(0)\pi_{1,2}(0) \quad (10)$$

$$z\pi^*_{n,0}(z)=\bar{\mu_1}\pi^*_{n,0}(z)+\mu_1\pi^*_{n+1,0}(z) +\mu_2\pi^*_{n,2}(z)+\mu_2 A^*(z)\pi_{n-1,2}(0) +\mu_1 A^*(z)\pi_{n,0}(0)+\bar{\mu_1}\pi_{n-1,0}(0)A^*(z) -\bar{\mu_1}\pi_{n,0}(0)-\mu_1\pi_{n+1,0}(0) -\mu_2\pi_{n,2}(0),\ 2\le n\le c-1, \quad (11)$$

$$z\pi^*_{c,0}(z)=\bar{\mu_1}\pi^*_{c,0}(z)+\bar{\mu_1}A^*(z)\pi_{c-1,0}(0) +\mu_2\pi^*_{c,2}(z)+\mu_1 A^*(z)\pi_{c,0}(0) +\mu_2 A^*(z)\pi_{c-1,2}(0)-\bar{\mu_1}\pi_{c,0}(0) -\mu_2\pi_{c,2}(0), \quad (12)$$

$$z\pi^*_{c+1,1}(z)=\bar{\mu_2}\pi^*_{c+1,1}(z)+\bar{\mu_1}A(z)\pi_{c,0}(0) +\mu_2\pi^*_{c+1,2}(z)+\mu_2 A^*(z)\pi_{c,2}(0) -\bar{\mu_2}\pi_{c+1,1}(0)-\mu_2\pi_{c+1,2}(0), \quad (13)$$

$$z\pi^*_{n,1}(z)=\bar{\mu_2}\pi^*_{n,1}(z)+\bar{\mu_2}A^*(z)\pi_{n-1,1}(0) +\mu_2\pi^*_{n,2}(z)-\mu_2\pi_{n,2}(0) +\mu_2 A^*(z)\pi_{n-1,2}(0)-\bar{\mu_2}\pi_{n,1}(0), \quad c+2\le n\le d-1, \quad (14)$$

$$z\pi^*_{0,2}(z)=\bar{\mu_2}\left(\pi^*_{0,2}(z)-\pi_{0,2}(0)+A^*(z)\pi_{d-1,1}(0)\right) +\mu_2\sum_{n=d}^{\infty}\left(\pi^*_{n,2}(z)-\pi_{n,2}(0)\right) +\mu_2 A^*(z)\sum_{n=d-1}^{\infty}\pi_{n,2}(0), \quad (15)$$

$$z\pi^*_{n,2}(z)=\bar{\mu_2}\pi^*_{n,2}(z)+\bar{\mu_2}A^*(z)\pi_{n-1,2}(0) -\bar{\mu_2}\pi_{n,2}(0),\ n\ge 1. \quad (16)$$

One may note that by adding equations (9) - (16), taking the limit as $z\to 1$, using the normalization condition $\sum_{i=0}^{c}\pi_{i,0}+\sum_{i=c+1}^{d-1}\pi_{i,1}+\sum_{i=0}^{\infty}\pi_{i,2}=1$ and after simplification we get

$$\sum_{i=0}^{c}\pi_{i,0}(0)+\sum_{i=c+1}^{d-1}\pi_{i,1}(0) +\sum_{i=0}^{\infty}\pi_{i,2}(0)=\lambda. \quad (17)$$

The left hand side (LHS) quantity is the mean number of entrances into the system per unit time and is equal to mean arrival rate λ given in right hand side (RHS).

3.1 Steady-state distribution at pre-arrival epochs

Let $\pi^-_{i,0}$ and $\pi^-_{i,1}$ represent the probability that there are i customers present in the system prior to an arrival epoch of a customer when the server is idle or busy with single customer and busy with an accessible batch, respectively. Also, let $\pi^-_{i,2}$ represent the probability that there are i customers present in the queue prior to an arrival epoch of a customer with non-accessible batch.

To obtain the steady-state distribution of the number of customers in the queue/system at pre-arrival epochs, we first connect pre-arrival epoch probabilities $\pi^-_{i,0}$ $(0\le i\le c)$, $\pi^-_{i,1}$ $(c+1\le i\le d-1)$ and $\pi^-_{i,2}$ $(i\ge 0)$ with the rates, $\pi^-_{i,0}$ $(0\le i\le c)$, $\pi^-_{i,1}$ $(c+1\le i\le d-1)$ and $\pi^-_{i,2}$ $(i\ge 0)$. These are given by

$$\pi^-_{i,j}=\frac{\pi_{i,j}(0)}{\sum_{i=0}^{c}\pi_{i,0}(0)+\sum_{i=c+1}^{d-1}\pi_{i,1}(0)+\sum_{i=0}^{\infty}\pi_{i,2}(0)} =\frac{1}{\lambda}\pi_{i,j}(0), \quad (18)$$

where λ is given by (17).

We first evaluate, $\pi_{i,0}(0)$ $(0\le i\le c)$, $\pi_{i,1}(0)$ $(c+1\le i\le d-1)$ and $\pi_{i,2}(0)$ $(i\ge 0)$

from (9) to (16) in the following manner. Setting $\theta = \mu_2$ in (16), we obtain

$$\pi_{n,2}(0) = \alpha^n \pi_{0,2}(0),\ n \geq 0. \quad (19)$$

From (16), using (19) and simplifying, we get

$$\pi^*_{n,2}(z) = \frac{\bar{\mu}_2 \alpha^{n-1}(A^*(z) - \alpha)\pi_{0,2}(0)}{z - \bar{\mu}_2},\ n \geq 1. \quad (20)$$

Differentiating (16) with respect to z and using $\alpha^{(j)} = A^{*(j)}(\bar{\mu}_2)$, we obtain

$$\pi^*_{n,2}(\mu_2) = \bar{\mu}_2 \alpha^{(1)} \alpha^{n-1} \pi_{0,2}(0),\ \geq 1, \quad ((21)$$

Setting $z = \bar{\mu}_2$ in (15) and (14), using (21) and simplifying, we get

$$\pi_{i,1}(0) = h_i \pi_{0,2}(0),\ c+1 \leq i \leq d-1, \quad (22)$$

where

$$h_i = \left(1 - \frac{\bar{\mu}_2 \alpha^{(1)} \alpha^{d-1}}{1-\alpha}\right)\alpha^{i-d} - (d-i-1)\mu_2 \alpha^{(1)} \alpha^{i-1},$$

Putting $z = \bar{\mu}_2$ in (13), using (21) and (22), we get

$$\pi_{c,0}(0) = h_c \pi_{0,2}(0), \quad (23)$$

where $h_c = \frac{\bar{\mu}_2}{\bar{\mu}_1 \alpha}(h_{c+1} - \mu_2 \alpha^{(1)} \alpha^c)$.. Setting $z = \bar{\mu}_1$ in (12), using (23) and (20), we get

$$\pi_{c-1,0}(0) = \frac{\pi_{0,2}(0)}{\bar{\mu}_1 \psi}\left[(\bar{\mu}_1 - \mu_1 \psi)h_c + \mu_2 \alpha^c - \frac{\mu_2 \bar{\mu}_2 \alpha^{c-1}(\psi - \alpha)}{\mu_2 - \mu_1}\right], \quad (24)$$

where $\psi = A^*(\bar{\mu}_1)$ and $\psi^{(j)} = A^{*(j)}(\bar{\mu}_1)$.
Substituting $z = \bar{\mu}_1$ in (11), using (21) and (22), we get

$$\pi_{i,0}(0) = \frac{1}{\bar{\mu}_1 \psi}[\bar{\mu}_1 \pi_{i+1,0}(0) + \mu_1 \pi_{i+2,0}(0) - \mu_1 \pi^*_{i+2,0}(\bar{\mu}_1) - \mu_1 \psi \pi_{i+1,0}(0) + \mu_2 \pi_{i+1,2}(0) - \mu_2 \pi^*_{i+1,2}(\bar{\mu}_1) - \mu_2 \psi \pi_{i,2}],\ i = c-2, c-3, \ldots, 1, \quad (25)$$

$$\pi_{0,0}(0) = \frac{1}{\psi}[\bar{\mu}_1 \pi_{1,0}(0) + \mu_1 \pi_{2,0}(0) - \mu_1 \pi^*_{2,0}(\bar{\mu}_1) - \mu_1 \psi \pi_{1,0}(0) + \mu_2 \pi_{1,2}(0) - \mu_2 \pi^*_{1,2}(\bar{\mu}_1) - \mu_2 \psi \pi_{0,2} - \mu_2 \psi \sum_{n=c+1}^{d-1} \pi_{n,1}(0)], \quad (26)$$

where $\pi^*_{i,0}(\mu_1)$ and $\pi^*_{i,2}(\mu_1)$, $i = c, c-1, \ldots, 1$, appearing in (25) can be obtained from (11), (12) and (16) in the following manner.
We obtain $\pi^*_{i,2}(\mu_1)\ i \geq 1$ by differentiating (16) with respect to z and setting $z = \bar{\mu}_1$.

$$\pi^{*(j)}_{i,2}(\bar{\mu}_1) = \frac{\bar{\mu}_2 \psi^{(j)} \alpha^{i-1} \pi_{0,2}(0) - j\pi^{*(j-1)}_{i,2}(\bar{\mu}_1)}{\mu_2 - \mu_1}\quad i \geq 1. \quad (27)$$

Differentiating (11) and (12) with respect to z, $j\,(= N-1)$ times and setting $z = \bar{\mu}_1$, we have

$$\pi^{*(j)}_{c,0}(\bar{\mu}_1) = \frac{1}{j+1}[\bar{\mu}_1 \psi^{(j+1)} \pi_{c-1,0}(0) + \mu_2 \pi^{*(j+1)}_{c,2}(\bar{\mu}_1) + \mu_1 \psi^{(j+1)} \pi_{c,0}(0) + \mu_2 \psi^{(j+1)} \pi_{c-1,2}(0)],$$

$$\pi^{*(j)}_{i,0}(\bar{\mu}_1) = \frac{1}{j+1}[\mu_1 \psi^{(j+1)} \pi_{i-1,0}(0) + \mu_2 \pi^{*(j+1)}_{i,2}(\bar{\mu}_1) + \mu_1 \pi^{*(j+1)}_{i+1,0}(\bar{\mu}_1) + \bar{\mu}_1 \psi^{(j+1)} \pi_{i-1,0}(0) + \mu_1 \psi^{(j+1)} \pi_{i,0}(0)],$$
$$i = c-1, c-2, \ldots, 2.$$

It may be noted that to evaluate $\pi^-_{i,j}$ from (18) we do not require the value of $\pi_{0,2}(0)$, since it cancels out in the numerator and denominator of RHS of (18).

3.2 Steady state distribution at arbitrary epochs

To obtain the queue/system length distribution at arbitrary epochs we develop relations between distributions of number of customers in the queue/system at pre-arrival and arbitrary epochs. This is discussed in the following theorem.

Theorem 1. The arbitrary epoch probabilities are given by

$$\pi_{n,2} = \frac{\bar{\mu}_2 \lambda (1-\alpha)\alpha^{n-1}}{\mu_2}\pi^-_{0,2},\ n \geq 1,$$

$$\pi_{0,2} = \frac{\lambda \pi^-_{0,2}}{\mu_2}[\bar{\mu}_2(h_{d-1} - 1) + \alpha^{d-1}],$$

$$\pi_{n,1} = \frac{\lambda \pi^-_{0,2}}{\mu_2}[\bar{\mu}_2(h_{n-1} - h_n) + (1-\alpha)\alpha^{n-1}],$$
$$c+2 \leq n \leq d-1,$$

$$\pi_{c+1,1} = \frac{\lambda \pi^-_{0,2}}{\mu 2}[(\bar{\mu}_1 h_c - \bar{\mu}_2 h_{c+1}) + (1-\alpha)\alpha^c],$$

$$\pi_{c,0} = \frac{\mu_2}{\mu_1}\pi^-_{c,2} + \frac{\lambda}{\mu_1}\{\bar{\mu}_1(\pi^-_{c-1,0} - \pi^-_{c,0}) + \mu_1 \pi^-_{c,0} + \mu_2(\pi^-_{c-1,2} - \pi^-_{c,2})\},$$

$$\pi_{n,0} = \pi_{n+1,0} + \frac{\lambda}{\mu_1}\{\mu_2(\pi^-_{n-1,2} - \pi^-_{n,2}) + \frac{\mu_2}{\mu_1}\pi_{n,2} + \bar{\mu}_1(\pi^-_{n-1,0} - \pi^-_{n,0}) + \mu_1(\pi^-_{n,0} - \pi^-_{n+1,0})\},$$
$2 \le n \le c-1,$

$$\pi_{1,0} = \pi_{2,0} + \frac{\mu_2}{\mu_1}\pi_{1,2} + \frac{\lambda}{\mu_1}\{\pi^-_{0,0} - \bar{\mu}_1\pi^-_{1,0} + \mu_2(\pi^-_{0,2} - \pi^-_{1,2}) + \mu_2\sum_{n=c+1}^{d-1}\pi^-_{n,1} + \mu_1(\pi^-_{1,0} - \pi^-_{2,0})\}.$$

Proof: Setting $z=1$ in (10)-(16), dividing both sides by λ and using (18), we obtain the result of the theorem. Using the normalization condition, we obtain $\pi_{0,0} = 1 - \sum_{i=1}^{c}\pi_{i,0} - \sum_{i=c+1}^{d-1}\pi_{i,1} - \sum_{i=0}^{\infty}\pi_{i,2}$..

4. Performance measures

In this section, we discuss some important operating characteristics in queueing system. In the case of infinite buffer, the average number of customers in the queue (L_q) is given by

$$L_q \equiv \sum_{i=2}^{c}(i-1)\pi_{i,0} + \sum_{i=0}^{\infty} i\,\pi_{i,2}.$$

The average waiting time in the queue (W_q) can be obtained using Little's rule as $W_q = L_q/\lambda$.

4.1 Expected lengths of the idle period, the busy period and the busy cycle

Let the expected length of the busy period, the idle period and the busy cycle be denoted by $E[B]$, $E[I]$ and $E[BC]$, respectively. A busy period starts when the server becomes busy and ends when it goes idle. The queue alternates between idle and busy periods. In fact, this is an alternating renewal process. The long-run proportion of time that the server is idle equals

$$\pi_{0,0} = \frac{E[I]}{E[I]+E[B]}.$$

Hence, the expected busy period is given by
$E[B] = \frac{1-\pi_{0,0}}{\lambda\pi_{0,0}}$, $E[I] = 1/\lambda$.

A busy cycle is the time between two successive departures leaving an empty system or equivalently, the sum of a busy period and an adjacent idle period. The expected length of busy cycle (BC) is $E[BC] = 1/\lambda\pi_{0,0}$.

5. Cost analysis and numerical results

The performance measures derived can now be used to optimize the performance of the system. We develop the total expected cost function per unit time for the discussed queueing system, assuming the decision variables as μ_1 and μ_2. Let
C_h=holding cost per unit time for each customer present in the system.
C_1=cost per unit time for keeping the server idle. The cost is incurred by keeping the server off without providing service to the customers.
C_2=cost incurred per unit time for keeping the server busy.
C_3=start-up cost for turning the server on.
The total expected cost per unit time, $F(\mu_1, \mu_2)$, is given by

$$F(\mu_1,\mu_2) = C_hL_q + C_1\frac{E[I]}{E[BC]} + C_2\frac{E[B]}{E[BC]} + C_3\frac{1}{E[BC]} = C_hL_q + C_1\pi_{0,0} + C_2(1-\pi_{0,0}) + C_3\lambda\pi_{0,0}.$$

Table 1: System performance measures under the optimal operating conditions for various values of $\boldsymbol{\lambda}$ *and* $\boldsymbol{c}$

c	λ	μ_1^*, μ_2^*	$F(\mu_1^*, \mu_2^*)$	$E[B]$	L_q	W_q
5.0	0.2	0.50, 0.84	15.73040	3.32003	0.12858	0.64288
	0.4	0.50, 0.55	31.18790	7.34159	0.72601	1.81501
	0.6	0.50, 0.55	42.54040	16.12450	1.37056	2.28426
	0.8	0.10, 0.55	46.65910	39.02870	1.53424	1.91780
10.0	0.2	0.50, 0.55	15.79980	3.33331	0.13332	0.66659
	0.4	0.50, 0.55	38.41640	9.38937	1.30650	3.26625
	0.6	0.50, 0.55	66.34720	30.83570	3.35800	5.59666
	0.8	0.10, 0.55	75.13220	70.86440	3.91612	4.89516

ISSN 2067 – 4074

Table 2. System performance measures under the optimal operating conditions for various values of c and d

c	d	μ_1^*, μ_2^*	$F(\mu_1^*, \mu_2^*)$	$E[B]$	L_q	W_q
5.0	8	0.50, 0.51	31.19550	7.35745	0.72643	1.81608
	10	0.50, 0.51	31.18380	7.34926	0.72556	1.81390
	12	0.50, 0.51	31.18290	7.34863	0.72549	1.81374
	15	0.50, 0.51	31.18280	7.34858	0.72549	1.81373
7.0	8	0.50, 0.95	34.83330	8.38520	1.01761	2.54402
	10	0.50, 0.51	34.83030	8.48930	1.01627	2.54068
	12	0.50, 0.51	34.82440	8.48496	1.01582	2.53956
	15	0.50, 0.95	34.82390	8.48461	1.01579	2.53947

Our objective is to determine the optimum value of the μ_1 and μ_2, say μ_1^* and μ_2^*, so as to minimize the total expected cost per unit time. The above cost function is non-linear, so it is difficult to derive the explicit analytical expression. Thus, a direct search method has been applied to find the optimum values of cost.

A numerical illustration is provided for geometric arrivals by considering the following cost parameters: $C_h = 12$, $C_1 = 10$, $C_2 = 20$, $C_3 = 35$, case 1: $d = 15$ and case 2: $\lambda = 0.4$. The optimal values μ_1* and μ_2^*, the minimum expected cost $F(\mu_1^*, \mu_2^*)$ and the values of system performance measures $E[C]$, $E[B]$, $E[I]$, L_q and W_q are shown in Table 1 for case 1 and in Table 2 for case 2, respectively.

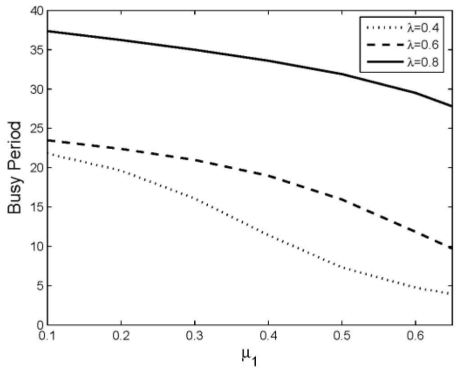

Figure 4. Impact of μ_1 on $E[B]$*

From Tables 1 - 2, we observe that (i) $F(\mu_1^*, \mu_2^*)$ increase as λ increases; (ii) the optimum value μ_1* decreases as λ increases; (iii) the optimum value μ_2^* increases as λ increases; (iv) the values of system performance measures increase as λ increases.

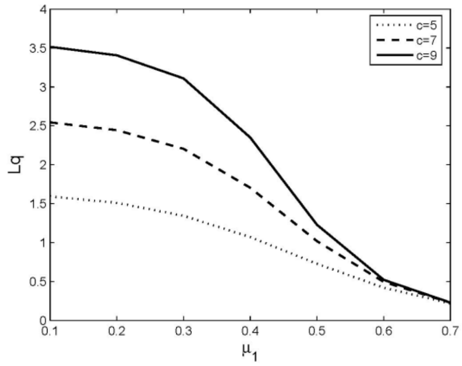

Figure 5. Effect of μ_1 on the average queue length

Figure 4 shows the impact of service rate for processing downstream traffic μ_1 on the

expected busy period ($E[B]$) The inter-arrival time is assumed to be geometric with $c = 5, d = 10$ and $\mu_2 = 0.7$. It can be observed that the expected busy period $E[B]$ decreases as service rate μ_1 increases for all values of λ. For smaller values of arrival rate or light traffic the expected busy period is comparatively decreasing more than intense traffic for higher values of service rate for processing downstream traffic μ_1.

Figure 5 shows the effect of service rate for processing downstream traffic μ_1 on the average queue length (L_q) for various values of control limit c. The inter-arrival time is assumed to be geometric with parameters $\lambda = 0.4$, $d = 15$ and $\mu_2 = 0.75$. It can be seen that average queue length decreases initially with increase of service rate for processing downstream traffic μ_1 until the average queue length starts to saturate. The service rate for processing downstream traffic in batch mode μ_2 is kept unchanged and average queue length decrease with the increase of control limit c.

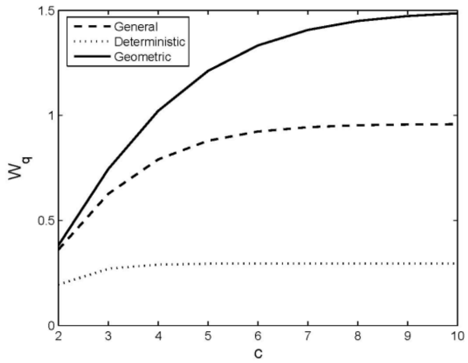

Figure 6. Impact of c on W_q

Figure 6 depicts the effect of control limit c on the average waiting-time in the queue (W_q) for various inter-arrival time distributions. It is observed that the average waiting-time in the queue increases as c increases but finally becomes constant with further increase of c. Further, the W_q in case of deterministic and arbitrary distributions is higher than the one obtained in case of geometric inter-arrival time distributions. The effect of d on average waiting time in the queue W_q for various service rates μ_2 is shown in Figure 7. It can be observed that as d increases, the corresponding average waiting time in the queue decreases monotonically until the waiting time in the queue saturates. But for fixed d, the average waiting time in the queue increases as service rate μ_2 increases.

Figure 8 illustrates the effect of average waiting time in the queue on control limit c and arrival rate, respectively. The inter-arrival time distribution is assumed to be geometric with $d = 10$, λ =0.5, $\mu_1 = 0.1$ and μ_2=0.85. The queueing delay increases linearly when both c and λ increase. This fact is in conformity with the results in Table 1. Therefore, we can define an admissible region in terms of the c and λ so that an acceptable queueing delay can be guaranteed.

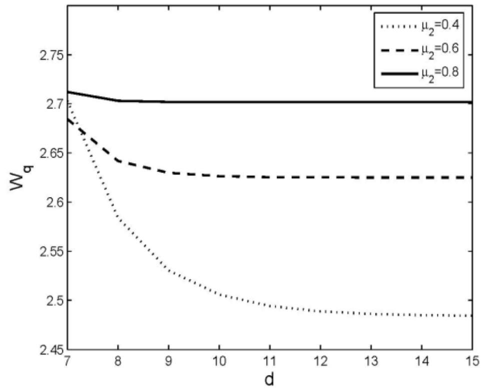

Figure 7. Impact of d on W_q

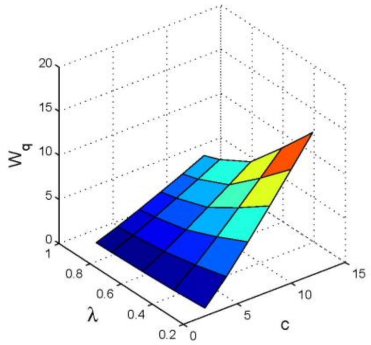

Figure 8. Effect of λ and c on W_q

The impact of μ_1 and μ_2 on cost function is shown in Figure 9. The inter-arrival time distribution is assumed to be geometric with $c = 5$, $d = 10$ and $\lambda = 0.5$. The cost function decreases as service rate μ_1 increases. But the influence of service rate μ_2 is comparatively modest. Therefore, we can infer that the cost function depends on the choice of the service rates μ_1 and μ_2 to a greater extent.

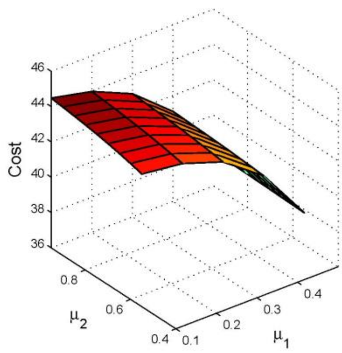

Figure 9. Impact of μ_1 and μ_2 on cost

6. Conclusion

In this paper, we have designed a load based energy saving mechanism for real time traffic that can adaptively control the system. Based on the operating management of power saving class II an analytical discrete-time queueing model is developed to evaluate the average overall delay and the power consumption of the future IEEE 802.16e wireless network for variable bit rate real time traffic. The model allows discrete time renewal input traffic and FCFS service discipline. The performance of the proposed scheme and the probabilistic characteristics of the lengths of an awake interval and a sleep interval is also provided. Using the supplementary variable technique, we have developed a recursive method to find the steady-state queue/system length distributions at pre-arrival and arbitrary epochs.

Various performance measures such as the sleep interval ratio, the average power consumption, and the mean delay of a message are evaluated to find the optimal system parameters while satisfying the QoS on delay bound. We constructed a cost function to show the optimal parameter value for proposed model. The longer sleep windows lead to higher energy efficiency, but they increase the overall message delay. When the overall message delay is limited, it is possible to determine optimal sleep mode parameters with our analytical model, which maximize the energy efficiency. We have shown a dependence of achievable energy efficiency on maximum tolerable delay.

The analytical model and the performance analysis presented in this paper provide a theoretical improvement and has potential applications in power saving mechanism. Moreover, the model used in this paper can be applied to analyze more complex models such as a non-renewal input single and batch service infinite (finite) buffer queueing system with accessibility to the batches that are left for future investigation. The proposed scheme in this paper improves the existing scheme in view of saving resources and energy.

References

[1] The Institute of Electrical and Electronics Engineers, Inc. IEEE standard for local and metropolitan area networks Part 16 (2004) Air interface for fixed broadband wireless access systems 2004.

[2] The Institute of Electrical and Electronics Engineers, Inc. (2006) IEEE standard for local and metropolitan area networks-Part 16: Air interface for fixed and mobile broadband wireless access systems, 2006.

[3] Y. Zhang and M. Fujise, Energy Management in the IEEE 802.16e MAC, *IEEE Communications Letters*, Vol. 10, 2006, pp.311–313.

[4] J. J. Chen, S. L. Wu and S. W. Wang, Power Saving Class Management for Energy Saving in IEEE 802.16e Wireless Networks, *Tenth International Conference on Mobile Data Management: Systems,* Services and Middleware,2009, pp.490–495.

[5] Y. Xiao, Energy Saving Mechanism in the IEEE 802.16e Wireless MAN, *IEEE Communication Letters*, Vol. 9, 2005, pp. 595–597.

[6] Y. Xiao, Performance Analysis of an Energy Saving Mechanism in the IEEE 802.16e Wireless MAN, *IEEE CCNC*, Vol. 1, 2006, pp. 406–410.

[7] J. B. Seo, S. Q. Lee, N. H. Park, H. W. Lee and C. H. Cho, Performance Analysis of Sleep Mode Operation in IEEE 802.16e, *IEEE VTC'2004*, Vol. 2, 2004, pp. 1169–1173.

[8] K. Han and S. Choi, Performance Analysis of Sleep Mode Operation in IEEE 802.16e Mobile Broadband Wireless Access Systems, *The 12th Asia-Pacific Conference on Communications*,2006, pp. 1141–1145.

[9] J. Jang, K. Han and S. Choi, Adaptive Power Saving Strategies for IEEE 802.16e Mobile Broadband Wireless Access, *IEEE VTC'2006-Spring*, 2006, pp. 1–5.

[10] Y. Zhang, Performance Modeling of Energy Management Mechanism in IEEE 802.16e Mobile WiMAX, *IEEE Wireless Communications and Networking Conference*, 2007, pp. 3205–3209.

[11] V. Goswami and P. Vijaya Laxmi, Performance analysis of a renewal input bulk service queue with accessible and non-accessible batches, *Intenational Journal of Quality Technology and Quatitative Management*, Vol. 8, 2011, pp. 87–100.

[12] C. C. Yang, , Y. T. Mai, J. Y. Chen, Y. S. Shen and Y. C. Kuo, LBPS: Load based power saving in IEEE802.16e Network,

Computer and Electrical Engineering, Vol. 38, 2012, pp. 891–905.

[13] V. Mancuso, and S. Alouf Analysis of power saving with continuous connectivity, *Computer Networks*, Vol. 56, No. 10, 2012, pp. 2481–2493.

[14] W. H. Liao and W. M. Yen, Power-saving scheduling with a QoS guaranteeina mobile WiMAX system, *Journal of Network and Computer Application*, Vol. 32, 2009, pp.1144–1152.

[15] S. T. Sheu, Y. C. Cheng and L. W. Chen, Listening Interval Spreading Approach (LISA) for Handling Burst Traffic in IEEE 802.16e Wireless Metropolitan Area Networks, *Proceedings of IEEE 68th Vehicular Technology Conference*, 2008, pp. 1-5.