

Research Article

Performance Analysis of the Sleep Mode in WiMAX 2 Networks with Multimedia Application

Shunfu Jin,^{1,2} Wuyi Yue,³ and Xiaofei Zhang⁴

¹ College of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China

² Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Yanshan University, Qinhuangdao 066004, China

³ Department of Intelligence and Informatics, Konan University, Kobe 658-8501, Japan

⁴ Affiliated Hospital, Hebei University of Engineering, Handan 056002, China

Correspondence should be addressed to Shunfu Jin; jsf@ysu.edu.cn

Received 5 August 2013; Revised 3 December 2013; Accepted 17 December 2013; Published 2 January 2014

Academic Editor: Alain Miranville

Copyright © 2014 Shunfu Jin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We consider the sleep mode with multimedia application in WiMAX 2 networks, where the real-time traffic includes the real-time and the best-effort traffic mixed. We present a queueing model with multiple heterogeneous vacations to characterize the system probability behavior in the networks with multimedia application. Taking into account the correlation of the real-time traffic, we assume the arrival process as a discrete-time Markovian arrival process (D-MAP) and analyze this queueing model by using the method of an embedded Markov chain. Then, we present the probability distribution for the number of data packets. Accordingly, we give the formulas for the performance measures in terms of the average response time of data packets, the energy saving ratio, and the standard deviation of the number of data packets. We also develop a cost function to determine the optimal length of the sleep cycle in order to maximize the energy saving ratio while satisfying the Quality of Service (QoS) constraint on the average response time of data packets. Finally, we provide numerical results to investigate the influence of the system parameters on the system performance.

1. Introduction

Wireless mesh networks are undergoing rapid development. They are intended to deliver wireless services for a large variety of applications in personal, local, campus, and metropolitan areas. Some protocol designs for wireless mesh networks have focused on power efficiency mechanisms.

IEEE 802.16 standard has been designed for fixed subscriber stations. As an enhancement of the IEEE 802.16 standard, IEEE 802.16e, called WiMAX [1], has improved the original standard supporting mobility so that the Mobile Station (MS) can move during services. Aiming at the next generation mobile WiMAX, called WiMAX 2, IEEE 802.16m [2] is currently being processed for standardization.

There have been many studies analyzing the performance of the sleep mode operations for Types I–III in WiMAX [3–7]. In [3], the authors evaluated and compared the sleep

mode operations for Type I and Type II power saving classes using the method of an embedded Markov chain. In order to avoid too frequent switching between the sleep state and the awake state, an enhanced power saving class Type III was provided in [4] and the system performances were analyzed for user initiated data packet arrivals. In [5], by increasing the unavailability interval, the authors proposed an enhanced power saving mechanism (ePSM), where both activated Type I and Type II power saving classes exist in an MS. The performance evaluation confirms that an ePSM can save more energy than conventional schemes.

Moreover, in [8], the authors used a two-phase Markovian Arrival Process to investigate the effect of the traffic parameters, such as the correlation parameter in the performance evaluation of a system for WiMAX. The optimal sleep mode strategy was obtained by minimizing the total average system cost.

However, an IEEE 802.16m amendment has been drafted to meet the requirements of WiMAX 2. IEEE 802.16m provides the enhanced performance required to support future advanced services and offers a novel sleep mode scheme which can reduce the power consumption and extend the lifetime of a battery-powered Mobile Station (MS) for a multimedia scenario.

WiMAX 2 has recently attracted a lot of research interest. In [9], the performance of the sleep mode in WiMAX 2 was mathematically analyzed, where the sleep cycle was supposedly fixed and the traffic indication (TRF-IND) was periodic. Moreover, the optimal TRF-IND interval was given to minimize the average power consumption of the MS while satisfying the Quality of Service (QoS) on the mean delay. In [10], the authors proposed a power-saving mechanism with binary exponential traffic indications in IEEE 802.16e/m. In [11], an efficient sleep mode operation was proposed by using a T_AMS timer. Also, the authors analyzed the proposed scheme by using an embedded Markov chain. The optimal parameters were given to minimize the power consumption while satisfying the QoS requirement on the average message delay.

In [12], taking into account that the listening window can be extended and the sleep cycle length can be adjusted, the authors conducted an analytical study on the power consumption and the average data packet delay to minimize the power consumption while satisfying a user-specified packet delay constraint. In [13], a concise analytical model for the sleep mode operation of WiMAX 2 was proposed. Considering both downlink and uplink traffics, the performance measures such as mean waiting time were derived and the performance comparisons between the WiMAX 2 and WiMAX were conducted with simulations. In [14], considering both downlink traffic and uplink traffic, a queueing model of two servers sharing vacations with close-down time and multiple vacations was built for WiMAX 2. The arrival process was assumed to follow a Bernoulli arrival process. By employing a two-dimensional embedded Markov chain, the authors derived the performance measures.

However, in the research on the sleep mode of WiMAX 2 mentioned above, the authors assumed that a sleep cycle was constituted under the condition that the time length of the extended listening window was to be equal to or less than a sleep cycle. The authors also assumed that the average remainder time of the extended listening window was equal to the difference between the time length of the sleep cycle and the average time length of the first sleep window. Moreover, in order to simplify the analysis procedure, the authors assumed uncorrelated traffic and modeled the incoming traffic using a Poisson process or a Bernoulli process. We know that, in multimedia application, two subsequent slot's states are correlated and the assumption of a Poisson process or a Bernoulli process is not reasonable.

In this paper, we present a comprehensive performance analysis by releasing those assumptions introduced in past research mentioned above for actual systems by considering the sleep mode of WiMAX 2 in a scenario, where the real-time traffic includes the real-time traffic and the best-effort traffic mixed. Considering the correlation coefficient between the

two subsequent time slots' states for the real-time traffic, we assume the arrival process to be a Discrete Time Markovian Arrival Process (D-MAP). The motivation for using D-MAP is that it can capture the digital nature of modern communication and model correlated traffic. Then, we build a multiple heterogeneous vacation queueing model to model the stochastic behavior of the sleep mode in the network system. Note that D-MAP can represent a variety of arrival processes, including the Bernoulli arrival process and the Markov Modulated Bernoulli process (MMBP). Therefore, the model presented in this paper is a generalization of the conventional models for analyzing the system performance of the sleep mode in WiMAX 2 with the real-time traffic including the real-time traffic and the best-effort traffic mixed.

By using the method of an embedded Markov chain, we present the probability distribution for the number of data packets. Considering the trade-off between the average response time of data packets and the energy saving ratio, we then develop a cost function to determine the optimal length of the sleep cycle in order to maximize the energy saving ratio while satisfying the QoS constraint on the average response time of data packets. With numerical results, we investigate the influence of the system parameters on the system performance. Finally, concerning the trade-off between different performance measures, we optimize the sleep cycle with the cost function.

The rest of this paper is organized as follows. In Section 2, we describe the system model for the multimedia application in which the real-time traffic includes the real-time and the best-effort traffic mixed. In Section 3, we analyze the system model in the steady state and give the probability distribution for the number of data packets. In Section 4, we obtain the formulas for the average response time of data packets, the energy saving ratio, and the higher moments as the standard deviation for the number of data packets. In addition, we construct a cost function to give the optimal design for the sleep cycle. In Section 5, we provide numerical results to illustrate the influence of the system parameters on the system performance. Finally, we draw our conclusions in Section 6.

2. System Model

In the sleep mode of WiMAX 2 networks, MS is provided with a set of sleep cycles no matter the system being at the awake period or the sleep period. Except for the initial sleep window, each subsequent sleep cycle consists of both a listening window and a sleep window. As specified in the standard of IEEE 802.16m, the real-time traffic includes the real-time traffic and the best-effort traffic mixed, and the length of the sleep cycle is fixed. Moreover, the value of the default listening window remains without change. Therefore, in this paper, we present the system model to describe the sleep mode in WiMAX 2 networks by setting a fixed time length for the sleep cycle [2].

During the listening window, data packets (if any) will be transmitted in the same way as in the awake period and the listening window will be extended if necessary. Therefore, in this paper, we regard the listening window with data packet

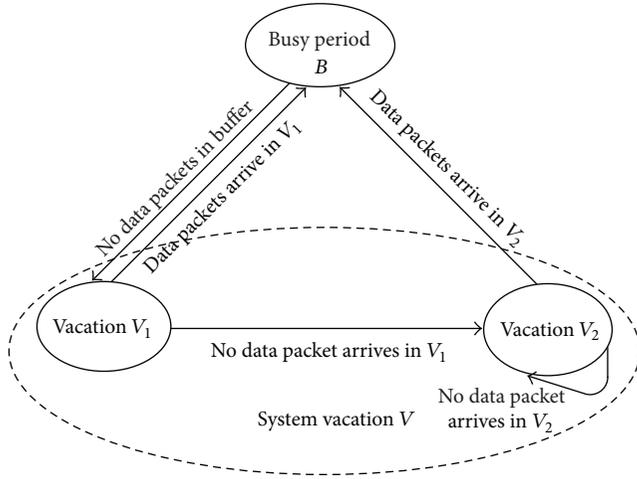


FIGURE 1: The state transition of the system model.

transmission as an extended listening window and the awake period following an extended listening window as a part of that extended listening window. Then the extended listening window is defined as a busy period B . Here, the extended listening window is perhaps followed by an awake period. A listening window will terminate on reaching the end of the current listening window including any extensions of the listening window, or reaching the end of the sleep cycle.

We assume the initial sleep window following an extended window to be a vacation period V_1 with the time length T_{V_1} and the subsequent sleep cycle without data packet transmission to be another vacation period V_2 with the time length T_{V_2} , respectively. We note that the extended listening window will end at any instant during a sleep cycle; in other words, the initial sleep window will begin at any instant of an extended listening window. For this, the time length T_{V_1} of the vacation period V_1 is supposed to be one half of the time length T_{V_2} of V_2 as $T_{V_1} = (1/2)T_{V_2}$, where T_{V_2} is in fact the time length of the sleep cycle that will be set as required by the system.

We also define a busy cycle R as a time period between the ending instants of two consecutive busy periods. We further define a system vacation V as a time period from the instant where a busy period ends to the instant where the next busy period begins. We can see that the first vacation V_1 and one or more subsequent vacations V_2 (if any) in a busy cycle R combine to produce a system vacation V .

With these assumptions stated above, we can model the system as a multiple heterogeneous vacation queueing model. The state transition of the system model is illustrated in Figure 1.

Figure 1 shows that:

- (1) if there is no data packet to be transmitted in the buffer of the serving Base Station (BS), the busy period B will be over and the vacation V_1 will begin;
- (2) if there is at least one data packet arrival within the time length T_{V_1} of the vacation V_1 , the MS will begin

a new busy period B when V_1 is over. Otherwise, the vacation V_2 will begin;

- (3) if there is at least one data packet arrival within the time length T_{V_2} of the vacation V_2 , the MS will enter a new busy period B when V_2 is over. Otherwise, the MS will begin another vacation V_2 again.

In this system model, we assume that the time axis is segmented into a sequence of equal intervals of unit duration, called slots. We also assume that data packets arrive only just before the end of a slot $t = \tau^-$, $\tau = 1, 2, \dots$ and depart only just after the end of a slot $t = \tau^+$, $\tau = 2, 3, \dots$. This is called a late arrival queueing system with delayed access. Various time epochs at which events occur are depicted in Figure 2.

The system consists of a single channel and a finite buffer size N . Data packets are supposed to be transmitted according to a First-Come, First-Served (FCFS) strategy.

We define the embedded Markov point as the instant at which a data packet is completely transmitted. We order these embedded points by r ($r = 1, 2, \dots$).

Moreover, we define the arrival phase by the stochastic state of the arrival process and define the state of the system by the number of data packets and the arrival phase at these embedded Markov points. Let L_r and J_r be the number of data packets and the arrival phase at the r th embedded point. Thus, $\{(L_r, J_r), r = 1, 2, \dots\}$ forms a discrete-time two-dimensional embedded Markov chain in a semi-Markov process with state space $\{(l, j), 0 \leq l \leq N, 1 \leq j \leq m\}$, where N is the system capacity and m is the number of phases in the underlying Markov Chain (UMC). N and m are the system parameters that we will give in the numerical results.

It should be noted that, in this paper, the embedded Markov points are chosen at the instant at which a data packet is completely transmitted and the condition of this system model being the stationary state is the number L of data packets in real-time traffic at the embedded Markov points which must be less than the system capacity N . With the assumptions and the conditions, in this system model, data packets would not be lost at the embedded Markov points.

Considering the correlation of the real-time traffic, the arrival process is assumed to be a D-MAP with representation $\{\mathbf{D}_r, r = 0, 1\}$, where \mathbf{D}_r is an $m \times m$ matrix and m is the number of phases in the UMC, where $m = 1, 2, \dots$. The element $[\mathbf{D}_0]_{ij}$ is the probability that there is a phase transition from i to j without any data packet arrivals, where $1 \leq i \leq m, 1 \leq j \leq m$. The element $[\mathbf{D}_1]_{ij}$ is the probability that there is a phase transition from i to j with a data packet arrival.

The matrix $\mathbf{D} = \mathbf{D}_0 + \mathbf{D}_1$ is the transition probability matrix of the UMC. The matrix \mathbf{D} is assumed to be irreducible.

Let $\tilde{\pi}$ be the stationary vector of size $1 \times m$ for the UMC; that is, $\tilde{\pi}$ is a solution for the equations as follows:

$$\tilde{\pi}\mathbf{D} = \tilde{\pi}, \quad \tilde{\pi}\mathbf{e} = 1, \quad (1)$$

where \mathbf{e} denotes a column one vector of size $m \times 1$.

The average arrival rate of the D-MAP is given as follows:

$$\lambda = \tilde{\pi}\mathbf{D}_1\mathbf{e}. \quad (2)$$

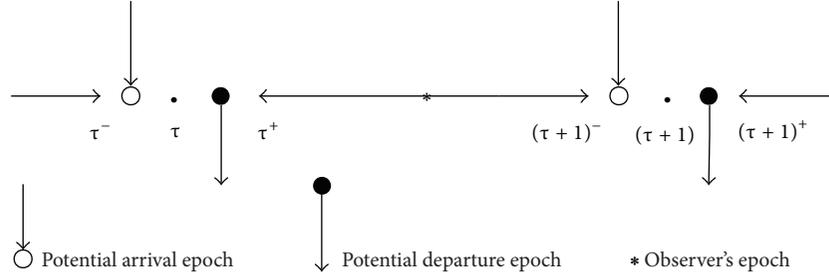


FIGURE 2: Various time epochs in this late arrival system with delayed access.

The matrix Probability Generating Function (PGF) of \mathbf{D} is defined as follows:

$$\mathbf{D}(z) = \mathbf{D}_0 + z\mathbf{D}_1. \quad (3)$$

The transmission time of a data packet is denoted by S . S is supposed to follow a general distribution. The probability distribution s_k , PGF $S(z)$, and the average value $E[S]$ of S are given as follows:

$$s_k = P\{S = k\}, \quad k = 1, 2, \dots, \\ S(z) = \sum_{k=1}^{\infty} s_k z^k, \quad E[S] = \sum_{k=1}^{\infty} k s_k. \quad (4)$$

3. Performance Analysis

3.1. Number of Data Packets. Let π_{lj} , $0 \leq l \leq N$, $1 \leq j \leq m$ be the probability distribution of the two-dimensional embedded Markov chain in a steady state. Then we have that

$$\pi_{lj} = \lim_{r \rightarrow \infty} P\{L_r = l, J_r = j\}. \quad (5)$$

Moreover, let $\boldsymbol{\pi}(l)$, $0 \leq l \leq N$ be the probability vector of data packet numbers for the embedded Markov chain in a steady state. Therefore $\boldsymbol{\pi}(l)$, $0 \leq l \leq N$ can be given as follows:

$$\boldsymbol{\pi}(l) = (\pi_{l1}, \pi_{l2}, \pi_{l3}, \dots, \pi_{lm}). \quad (6)$$

In order to compute $\boldsymbol{\pi}(l)$, $0 \leq l \leq N$, some notations are introduced as follows.

Let $\mathbf{A}_S(k, n)$ be a matrix of size $m \times m$, whose element $[\mathbf{A}_S(k, n)]_{ij}$ is the conditional probability that there are n data packet arrivals during the transmission time (lasting k slots) of a data packet, given that the transmission started with arrival phase i and ended with arrival phase j .

Let $\mathbf{A}_S(n)$ denote a matrix of size $m \times m$ representing that there are n data packet arrivals within the transmission time S of a data packet. $\mathbf{A}_S(n)$ can be given as follows:

$$\mathbf{A}_S(n) = \sum_{k=1}^{\infty} s_k \mathbf{A}_S(k, n), \quad 0 \leq n. \quad (7)$$

Let $\mathbf{A}_{V_1}(T_C, n)$ be a matrix of size $m \times m$, whose element $[\mathbf{A}_{V_1}(T_C, n)]_{ij}$ is the conditional probability that there are n

data packet arrivals during the vacation V_1 (lasting T_C slots), given that the arrival process transfer is from phase i at the beginning instant of the vacation V_1 to phase j at the end instant of the vacation V_1 .

Let $\mathbf{A}_{V_2}(T_{V_2}, n)$ be a matrix of size $m \times m$, whose element $[\mathbf{A}_{V_2}(T_{V_2}, n)]_{ij}$ is the conditional probability that there are n data packet arrivals during the vacation V_2 (lasting T_{V_2} slots), given that the arrival process transfer is from phase i at the beginning instant of the vacation V_2 to phase j at the end instant of the vacation V_2 .

Considering the multiple vacations in this paper, let $\mathbf{A}_V(n)$ be a matrix of size $m \times m$ with element $[\mathbf{A}_V(n)]_{ij}$ representing the probability that there are n data packet arrivals within the system vacation, given that the system vacation V started with the arrival process of phase i and ended with the arrival process of phase j .

We calculate $\mathbf{A}_V(n)$ in three possible cases as follows. (1) For the case of $N > T_{V_2}$, if $1 \leq n \leq T_C$,

$$\mathbf{A}_V(n) = \mathbf{A}_{V_1}(T_C, n) + \mathbf{A}_{V_1}(T_C, 0) \\ \times (\mathbf{I} - \mathbf{A}_{V_2}(T_{V_2}, 0))^{-1} \mathbf{A}_{V_2}(T_{V_2}, n); \quad (8)$$

if $T_C + 1 \leq n \leq T_{V_2}$,

$$\mathbf{A}_V(n) = \mathbf{A}_{V_1}(T_C, 0) (\mathbf{I} - \mathbf{A}_{V_2}(T_{V_2}, 0))^{-1} \times \mathbf{A}_{V_2}(T_{V_2}, n); \quad (9)$$

if $n > 2T_C$,

$$\mathbf{A}_V(n) = \mathbf{0}. \quad (10)$$

(2) For the case of $T_C \leq N \leq T_{V_2}$, if $1 \leq n \leq T_C$,

$$\mathbf{A}_V(n) = \mathbf{A}_{V_1}(T_C, n) + \mathbf{A}_{V_1}(T_C, 0) \\ \times (\mathbf{I} - \mathbf{A}_{V_2}(T_{V_2}, 0))^{-1} \mathbf{A}_{V_2}(T_{V_2}, n); \quad (11)$$

if $T_C + 1 \leq n \leq T_{V_2}$,

$$\mathbf{A}_V(n) = \mathbf{A}_{V_1}(T_C, 0) (\mathbf{I} - \mathbf{A}_{V_2}(T_{V_2}, 0))^{-1} \times \mathbf{A}_{V_2}(T_{V_2}, n); \quad (12)$$

if $n = N$,

$$\begin{aligned} \mathbf{A}_V(n) &= \mathbf{A}_{V_1}(T_C, 0) (\mathbf{I} - \mathbf{A}_{V_2}(T_{V_2}, 0))^{-1} \\ &\quad \times (\mathbf{A}_{V_2}(T_{V_2}, N) + \mathbf{A}_{V_2}(T_{V_2}, N + 1) \\ &\quad + \dots + \mathbf{A}_{V_2}(T_{V_2}, T_{V_2})); \end{aligned} \quad (13)$$

if $n > N$,

$$\mathbf{A}_V(n) = \mathbf{0}. \quad (14)$$

(3) For the case of $T_C > N$, if $n < N$,

$$\begin{aligned} \mathbf{A}_V(n) &= \mathbf{A}_{V_1}(T_C, n) + \mathbf{A}_{V_1}(T_C, 0) \\ &\quad \times (\mathbf{I} - \mathbf{A}_{V_2}(T_{V_2}, 0))^{-1} \mathbf{A}_{V_2}(T_{V_2}, n); \end{aligned} \quad (15)$$

if $n = N$,

$$\begin{aligned} \mathbf{A}_V(n) &= (\mathbf{A}_{V_1}(T_C, N) + \mathbf{A}_{V_1}(T_C, N + 1) + \dots + \mathbf{A}_{V_1}(T_C, T_C)) \\ &\quad + \mathbf{A}_{V_1}(T_C, 0) \times (\mathbf{I} - \mathbf{A}_{V_2}(T_{V_2}, 0))^{-1} \\ &\quad \times (\mathbf{A}_{V_2}(T_{V_2}, N) + \mathbf{A}_{V_2}(T_{V_2}, N + 1) \\ &\quad + \dots + \mathbf{A}_{V_2}(T_{V_2}, T_{V_2})); \end{aligned} \quad (16)$$

if $n > N$,

$$\mathbf{A}_V(n) = \mathbf{0}. \quad (17)$$

Among which, \mathbf{I} and $\mathbf{0}$ are the identity and zero matrices of size $m \times m$, respectively.

It may be noted here that in order to know $\mathbf{A}_S(n)$ and $\mathbf{A}_V(n)$, we need to compute $\mathbf{A}_S(k, n)$, $\mathbf{A}_{V_1}(T_C, n)$, and $\mathbf{A}_{V_2}(T_{V_2}, n)$ efficiently.

Following the arguments of Chaudhry and Gupta [15], it can be seen that $\mathbf{A}_S(k, n)$, $\mathbf{A}_{V_1}(T_C, n)$, and $\mathbf{A}_{V_2}(T_{V_2}, n)$ satisfy the following equations:

$$\begin{aligned} \mathbf{A}_S(k, n) &= \mathbf{D}_0 \mathbf{A}_S(k - 1, n) + \mathbf{D}_1 \mathbf{A}_S(k - 1, n - 1), \\ &1 \leq k, 0 \leq n, \end{aligned}$$

$$\begin{aligned} \mathbf{A}_{V_1}(T_C, n) &= \mathbf{D}_0 \mathbf{A}_{V_1}(T_C - 1, n) + \mathbf{D}_1 \mathbf{A}_{V_1}(T_C - 1, n - 1), \\ &0 \leq n \leq T_C, \end{aligned}$$

$$\begin{aligned} \mathbf{A}_{V_2}(T_{V_2}, n) &= \mathbf{D}_0 \mathbf{A}_{V_2}(T_{V_2} - 1, n) + \mathbf{D}_1 \mathbf{A}_{V_2}(T_{V_2} - 1, n - 1), \\ &0 \leq n \leq 2T_C. \end{aligned} \quad (18)$$

Particularly,

$$\mathbf{A}_S(0, 0) = \mathbf{A}_{V_1}(0, 0) = \mathbf{I},$$

$$\mathbf{A}_S(k, -1) = \mathbf{A}_{V_1}(T_C, -1) = \mathbf{A}_{V_2}(T_{V_2}, -1) = \mathbf{0}, \quad (19)$$

$$\mathbf{A}_S(k, n) = \mathbf{A}_{V_1}(T_C, n) = \mathbf{A}_{V_2}(T_{V_2}, n) = \mathbf{0},$$

where $0 \leq k < n, 0 \leq T_C < T_{V_2} < n$.

The transition probability matrix \mathbf{P} of the two-dimensional Markov process $\{(L(r), J(r)), r \geq 0\}$ is given as follows:

$$\mathbf{P} = \begin{pmatrix} \mathbf{F}_0 & \mathbf{F}_1 & \dots & \mathbf{F}_{N-1} & \mathbf{F}_N \\ \mathbf{A}_S(0) & \mathbf{A}_S(1) & \dots & \mathbf{A}_S(N-1) & \widehat{\mathbf{A}}_S(N) \\ \mathbf{0} & \mathbf{A}_S(0) & \dots & \mathbf{A}_S(N-2) & \widehat{\mathbf{A}}_S(N-1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_S(1) & \widehat{\mathbf{A}}_S(2) \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_S(0) & \widehat{\mathbf{A}}_S(1) \end{pmatrix}, \quad (20)$$

where elements of the first row are obtained as follows:

$$\mathbf{F}_j = \sum_{k=1}^{j+1} \mathbf{A}_V(k) \mathbf{A}_S(j+1-k), \quad 0 \leq j \leq N-2,$$

$$\mathbf{F}_{N-1} = \sum_{k=1}^{N-1} \mathbf{A}_V(k) \mathbf{A}_S(N-k) + \mathbf{A}_V(N) \mathbf{A}_S(0), \quad (21)$$

$$\mathbf{F}_N = \sum_{k=1}^{N-1} \mathbf{A}_V(k) \widehat{\mathbf{A}}_S(N+1-k) + \mathbf{A}_V(N) \widehat{\mathbf{A}}_S(1).$$

The last element for each row, except for the first row, is given as follows:

$$\widehat{\mathbf{A}}_S(n) = \sum_{l=n}^{\infty} \mathbf{A}_S(l). \quad (22)$$

Let $\boldsymbol{\pi} = (\boldsymbol{\pi}(0), \boldsymbol{\pi}(1), \dots, \boldsymbol{\pi}(N))$ be the stationary probability vector of size $1 \times (N+1)m$ for the transition probability matrix \mathbf{P} . Then we can obtain $\boldsymbol{\pi}(l)$ by solving the equations as follows:

$$\boldsymbol{\pi} \mathbf{P} = \boldsymbol{\pi}, \quad \boldsymbol{\pi} \mathbf{e} = 1. \quad (23)$$

Let L be the number of data packets at the embedded Markov points. The average value $E[L]$ of L can be given as follows:

$$E[L] = \sum_{l=0}^N l \boldsymbol{\pi}(l) \mathbf{e}. \quad (24)$$

3.2. Busy Cycle. Let T_B be the time length of a busy period B and $E[T_B]$ be the corresponding average value of T_B . By using the analysis results of the classical Geom/G/1 queueing model [16], $E[T_B]$ can be given as follows:

$$E[T_B] = \frac{E[S]}{\boldsymbol{\pi}(0) \mathbf{e}}. \quad (25)$$

Let N_V be the number of sleep windows in a busy cycle R . Then we can get the probability distribution of N_V as follows:

$$P\{N_V = x\} = \begin{cases} 1 - \bar{\lambda}^{-T_C}, & x = 1, \\ \bar{\lambda}^{-(2x-3)T_C} (1 - \bar{\lambda}^{-T_{V_2}}), & x \geq 2, \end{cases} \quad (26)$$

where $\bar{\lambda} = 1 - \lambda = 1 - \bar{\boldsymbol{\pi}} \mathbf{D}_1 \mathbf{e}$.

Let T_V be the time length in slots of the system vacation period V . The average value $E[T_V]$ of T_V is then given as follows:

$$E[T_V] = T_C \left(1 - \bar{\lambda}^{T_C}\right) + T_C \bar{\lambda}^{T_C} \frac{3 - \bar{\lambda}^{T_{V_2}}}{1 - \bar{\lambda}^{T_{V_2}}}. \quad (27)$$

Let T_R be the time length of the busy cycle R . The average value $E[T_R]$ of T_R is then given as follows:

$$E[T_R] = E[T_B] + E[T_V]. \quad (28)$$

Combining (25) and (28), the probability of the MS being awake is given as follows:

$$b = \frac{E[T_B]}{E[T_R]}. \quad (29)$$

4. Performance Measures and Optimal Design

We define the response time σ of a data packet as being the time period in slots that has elapsed from the arrival instant to the end of the transmission of that data packet. This is a very important performance measure for evaluating the user QoS for the system. By applying Little's law [17], the average response time σ of data packets is obtained as follows:

$$\sigma = \frac{E[L]}{\lambda} = \frac{\sum_{l=0}^N l \pi(l) \mathbf{e}}{\lambda}. \quad (30)$$

The energy saving ratio γ is defined as the amount of energy conserved per slot, which is one of the most important performance measures for evaluating the energy saving efficiency of the sleep mode in IEEE 802.16m. The energy saving ratio γ is given as follows:

$$\gamma = C_1 (1 - b) - \frac{C_2}{T_{V_2}}, \quad (31)$$

where C_1 is the energy saved per slot for the system being in the sleep period and C_2 is the additional energy consumption per slot due to the traffic indications introduced in IEEE 802.16m.

We know that the design and development of the system require not only average performance measures but also higher moments, because the output stream from one system often forms the input stream to another.

It is well known that the diffusion approximation of the queueing length in the system buffer is important for evaluating the system performance [18]. The diffusion degree of the queueing length can reflect the delay jitter of data packets. Here, we give the standard deviation β of the number of data packets in the system buffer as follows:

$$\beta = \sqrt{E[L^2] - (E[L])^2}, \quad (32)$$

where $E[L]$ is given by (24) and $E[L^2]$ is the second moment for the number of data packets at the embedded Markov points given as follows:

$$E[L^2] = \sqrt{\sum_{l=0}^N l^2 \pi(l) \mathbf{e}}. \quad (33)$$

Obviously, a longer sleep cycle can enhance the energy saving ratio while increasing the average response time of data packets. On the other hand, a shorter sleep cycle can decrease the average response time of data packets while reducing the energy saving ratio. Therefore, it is important to determine an optimal value for the sleep cycle in order to maximize the energy saving ratio while satisfying the QoS constraint on the average response time of data packets.

To this end, we introduce a cost function $F(T_{V_2})$, which is proportional to the average response time σ and is inversely proportional to the energy saving ratio γ . Thus, the cost function is given as follows:

$$F(T_{V_2}) = f_1 \sigma + \frac{f_2}{\gamma}, \quad (34)$$

where f_1 is the factor of the average response time of data packets to the cost and f_2 is the factor of the energy saving ratio to the cost.

5. Numerical Results

In this section, we provide numerical results from analysis and simulation to illustrate the system performance for the sleep mode in WiMAX 2. The simulation results match well with the results from the analysis. As an example, we set system parameters as follows: $N = 15$, $E[S] = 1$, and $\lambda = 0.2$.

Since D-MAP can represent a variety of arrival processes, we decide on a special case of D-MAP, namely, a Bernoulli modulated ON-OFF traffic model, as an example of the system input. The ON-OFF traffic model refers to the arrival process modulated by a two-state Markov chain. When the system is in the ON state, the arrival process follows a Bernoulli process with the parameter ω . If the system is in the OFF state, no arrival occurs.

We assume that the system is in the ON state with the probability ψ . The correlation coefficient between the two subsequent slot's states is $1 - 1/M$, where M is the correlation parameter for the average time lengths for the system being in the ON and OFF states, respectively. The mean time length for the system being in the ON state is $M/(1 - \psi)$, and correspondingly M/ψ is the average time length for the system being in the OFF state.

Therefore, the matrix PGF $\mathbf{D}(z)$ for the traffic model in this paper can be given as follows:

$$\mathbf{D}(z) = \begin{pmatrix} 1 - \omega + \omega z & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 - \frac{1 - \psi}{M} & \frac{1 - \psi}{M} \\ \frac{\psi}{M} & 1 - \frac{\psi}{M} \end{pmatrix}. \quad (35)$$

Letting $z = 0$ in (35), we can derive the formula of \mathbf{D}_0 as follows:

$$\mathbf{D}_0 = \begin{pmatrix} (1 - \omega) \left(1 - \frac{1 - \psi}{M}\right) & (1 - \omega) \frac{1 - \psi}{M} \\ \frac{\psi}{M} & 1 - \frac{\psi}{M} \end{pmatrix}. \quad (36)$$

TABLE 1: System parameters.

Parameter	Value
One slot	5 ms
Mean transmission time of a packet	5 ms
Probability being at the ON state	0.8
Power conservation during sleep period	100 mW
Energy consumption for sending a TRF-IND	15 mJ

Letting $z = 1$ in (35), we can obtain the formula of \mathbf{D} as follows:

$$\mathbf{D} = \begin{pmatrix} 1 - \frac{1-\psi}{M} & \frac{1-\psi}{M} \\ \frac{\psi}{M} & 1 - \frac{\psi}{M} \end{pmatrix}. \quad (37)$$

Combining (36) and (37), we can obtain the formula of \mathbf{D}_1 as follows:

$$\mathbf{D}_1 = \mathbf{D} - \mathbf{D}_0 = \begin{pmatrix} \omega \left(1 - \frac{1-\psi}{M}\right) & \omega \frac{1-\psi}{M} \\ 0 & 0 \end{pmatrix}. \quad (38)$$

Meanwhile, we assume that the transmission time S of a data packet follows a geometric distribution with the parameter μ ($0 \leq \mu \leq 1$). Then the PGF $S(z)$ and the average value $E[S]$ of S are given as follows:

$$S(z) = \sum_{k=1}^{\infty} \mu^{k-1} \mu z^k = \frac{\mu z}{1 - \mu z}, \quad E[S] = \sum_{k=1}^{\infty} k \mu^{k-1} \mu = \frac{1}{\mu}. \quad (39)$$

Now, we will discuss the influence of the sleep cycle on the system performance measures for different correlation parameters M with numerical results. Obviously, the smaller the value of correlation parameters M is, the less the correlation degree of the real-time traffic is. Particularly, $M = 1$ means that there is only the best-effort traffic in the system and $M = \infty$ means that there is only the real-time traffic when the traffic is mixed. The ratio of the best-effort traffic to the real-time traffic in the mixed traffic is given as $1/M$. When we let B_Q denote the quantity of the best-effort traffic, R_Q denotes the quantity of the real-time traffic, and M_Q denotes the quantity of the mixed traffic, we can have

$$B_Q = M_Q \times \frac{1}{M}, \quad R_Q = M_Q \times \left(1 - \frac{1}{M}\right). \quad (40)$$

By referencing [9, 10], we set the system parameters for numerical results in Table 1.

The average response time σ of data packets versus the sleep cycle T_{V_2} with different correlation parameters M is depicted in Figure 3.

It is illustrated that for the same correlation parameter M , the longer the time length T_{V_2} of the sleep cycle is, the greater the average response time σ of data packets is. The reason is that when the time length of the sleep cycle is longer, the system is more likely to be in the sleep period, so the average

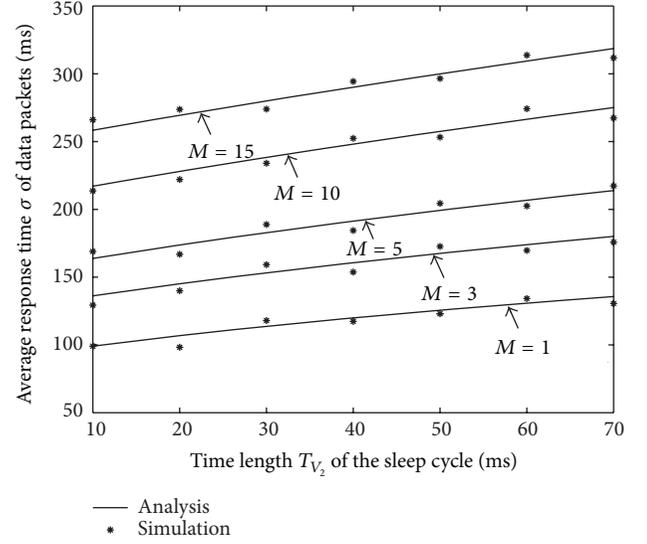


FIGURE 3: Average response time σ of data packets versus sleep cycle T_{V_2} .

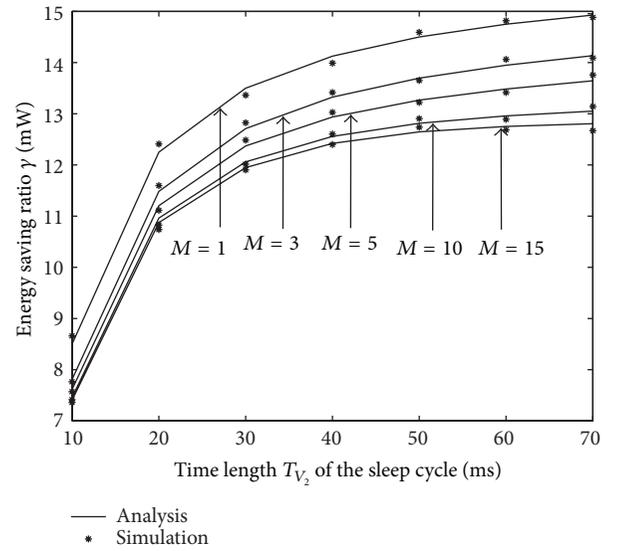


FIGURE 4: Energy saving ratio γ versus sleep cycle T_{V_2} .

response time of data packets is longer. On the other hand, for all the time lengths T_{V_2} of the sleep cycle, the larger the correlation parameter M is, the longer the average response time σ of data packets will be. That is because the greater the correlation parameter is, the denser the data packet arrivals are and the longer the queueing length is, so the longer the average response time of data packets will be. Moreover, we observe that if the best-effort traffic only is considered in multimedia application, the average response time of data packets will be undervalued.

The energy saving ratio γ versus the time length T_{V_2} of the sleep cycle with different correlation parameters M is demonstrated in Figure 4.

It can be observed that for the same correlation parameter M , the energy saving ratio γ will increase along with the

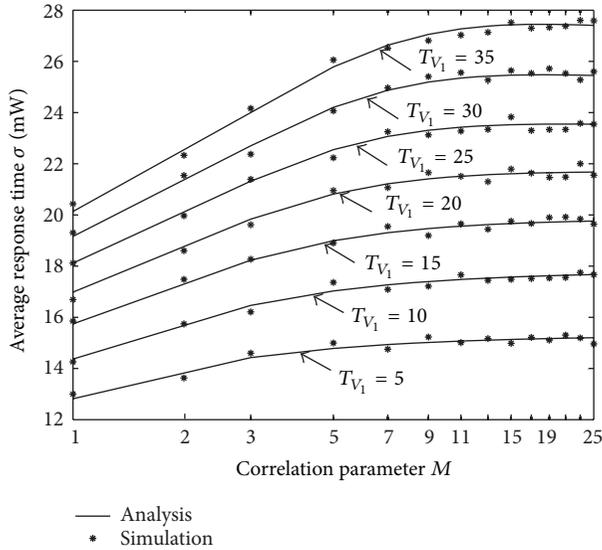


FIGURE 5: Average response time σ of data packets versus correlation parameter M .

time length T_{V_2} of the sleep cycle. The reason is that when the time length of the sleep cycle is longer, the system is more likely to be in a sleep period. Energy will be saved in the sleep period, so the energy saving ratio will be greater. It can also be found that for all the time lengths T_{V_2} of the sleep cycle, the energy saving ratio γ will decrease as the correlation parameter M decreases. This is because the bigger the correlation parameter M is, the denser the data packet arrivals are, the busier the MS is, and the less energy will be saved. Moreover, we can find that if the best-effort traffic is only considered for a multimedia scenario, the energy saving effect will be overevaluated.

In Figures 5 and 6, we show how the average response time σ of data packets and the energy saving ratio γ change versus the correlation parameter M from 1 to 25 with different time lengths T_{V_1} of the initial sleep window from 5 to 35.

Moreover, we observe that for the same correlation parameter M , with a smaller value for the time length T_{V_1} of the initial sleep window, such as $T_{V_1} = 5$, the average response time σ and the energy saving ratio γ have smaller results. But when the time length T_{V_1} of the initial sleep window becomes a higher value, such as $T_{V_1} = 35$, results of the average response time σ of data packets and the energy saving ratio γ are greater. The reason is that when the time length T_{V_1} of the initial sleep window is longer, the system is more likely to be in the sleep period, so the longer the response time of data packets is and the greater the energy saving ratio will be.

We also find that the differences among values of the energy saving ratio γ are bigger when the time length T_{V_1} is smaller. This is because for a smaller value of the time length T_{V_1} , there will be more sleep windows in a busy cycle; therefore a slight increase in the time length T_{V_1} will result in a huge increase in the value of the energy saving ratio γ . On the other hand, when the time length T_{V_1} is greater, the system is more likely to transfer from a sleep state to an awake state and the fewer sleep windows will be in a busy cycle. Therefore, an

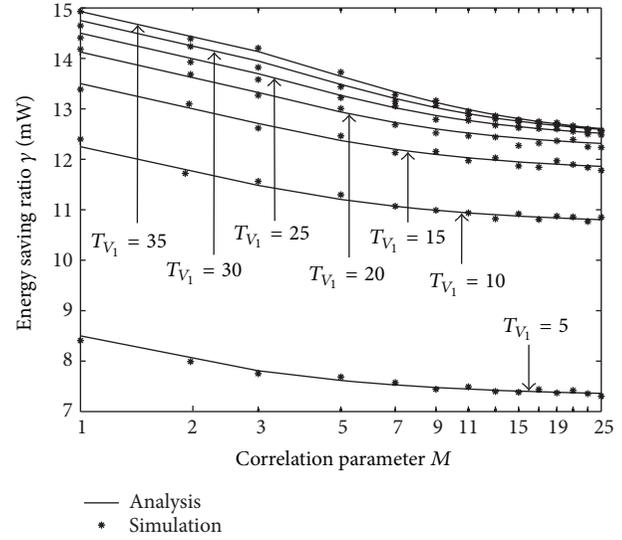


FIGURE 6: Energy saving ratio γ versus correlation parameter M .

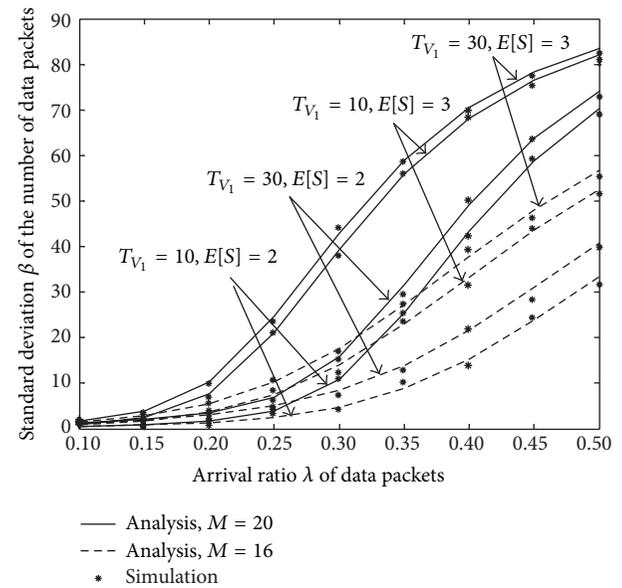


FIGURE 7: Standard deviation β of the number of data packets versus arrival ratio λ of data packets.

increase in the time length T_{V_1} will result in a gradual increase in the energy saving ratio γ .

Considering the special case for $M = 1$, namely, there is only the best-effort traffic, for all the time length T_{V_1} , the average response times σ of data packets are shorter and the energy saving ratios γ are bigger. That is to say, if the correlation of the real-time traffic is neglected, the system performance of the sleep mode with multimedia application in WiMAX 2 networks will be overoptimistically.

The change trend of the standard deviation β of the number of data packets versus the arrival ratio λ of data packets is depicted in Figure 7.

In Figure 7, we find that for all the time lengths T_{V_1} of the initial sleep cycle, the average transmission time $E[S]$

of data packets and the correlation parameter M , the higher the arrival ratio λ of data packets is, the larger the standard deviation β of the number of data packets will be. This is because the higher the arrival ratio λ of data packets is, the more data packets will arrive within a certain time period. Note that the minimal number of data packets is 1; a greater average number of data packets in the buffer during the extended listening period means a maximal number of data packets and thus the standard deviation β of the number of data packets will be greater.

We notice that, for the same time length T_{V_1} of the initial sleep cycle, the correlation parameter M , and the arrival ratio λ of data packets, the standard deviation β of the number of data packets will increase with an increase in the average transmission time $E[S]$ of data packets. The reason is that the longer the average transmission time $E[S]$ of data packets is, the maximal value of the queuing length in the system will be during the extended listening period. So the standard deviation β of the number of data packets will be greater.

In Figure 7, we also find that, for the same time length T_{V_1} of the initial sleep cycle, the same average transmission time $E[S]$ of data packets, and the same arrival ratio λ of data packets, the larger the correlation parameter M is, the greater the standard deviation β of the number of data packets will be. For example, the standard deviation β of the number of data packets for the case of $M = 20$ is much greater than that for the case of $M = 16$, for each case of T_{V_1} . This is because in the $M = 20$ case, the quantity of the real-time traffic is much greater than that of the best-effort traffic in the mixed traffic, so it is more likely that some of the real-time traffic happens to be waiting in the buffer. This results in the standard deviation β of the number of data packets being greater.

In Figure 7, we also observe that, for the same correlation parameter M , the same average transmission time $E[S]$ of data packets, and the same arrival ratio λ of data packets, the longer the time length T_{V_1} is, the greater the standard deviation β of the number of data packets will be. This is because the longer the time length T_{V_1} is, the greater the maximal number of data packets arriving within the time length T_{V_1} will be, so that the number range of the data packets will become wide at the beginning instant of the data packets' transmission. It causes the standard deviation β of the data packets' number to be greater.

Figure 7 clearly shows that the standard deviation β of the number of data packets is heavily dependent on the correlation of the real-time traffic. If we underestimate the correlation degree of the real-time traffic in multimedia application, as in the system models presented in the available literature, the standard deviation β of the number of data packets will be undervaluated. It means that the results from analyzing the system model will be considerably separated from actual systems.

To maximize the energy saving ratio γ while satisfying the QoS constraint on the average response time σ of data packets, we investigate the changing trend for the cost function $F(T_{V_2})$ of the system given by (34) versus the time length T_{V_2} of the sleep cycle with different correlation parameters M . Taking $f_1 = 0.3$ and $f_2 = 0.5$ as an example, the results are shown in Figure 8.

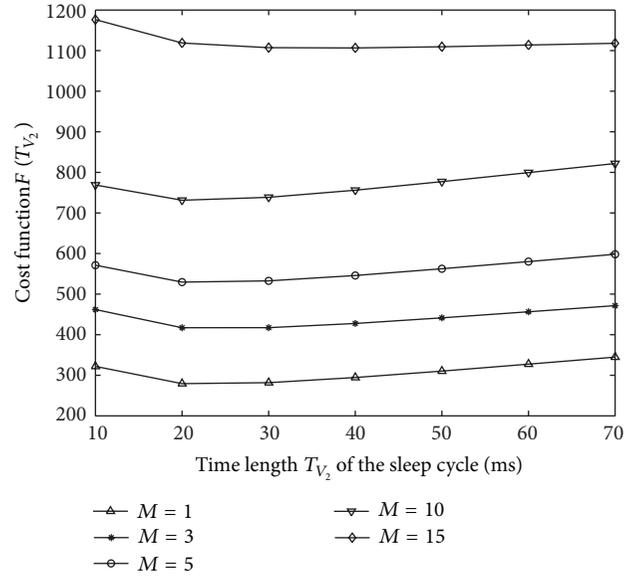


FIGURE 8: Cost function $F(T_{V_2})$ versus sleep cycle T_{V_2} .

TABLE 2: Optimal time length T_{V_2} of the sleep cycle.

Correlation parameter	T_{V_2}	Minimal cost $F(T_{V_2})$
$M = 1$	20	279.4107
$M = 3$	20	417.1222
$M = 5$	20	529.5695
$M = 10$	20	731.3965
$M = 15$	30	1106.5756

From Figure 8, we see that two stages are experienced by the cost function $F(T_{V_2})$. In the first stage, the cost function $F(T_{V_2})$ will decrease along with an increase in the time length T_{V_2} of the sleep cycle. During this stage, the longer the time length of the sleep cycle is, the longer the MS will stay in the sleep period and the more energy will be saved, so the less the cost function will be. In the second stage, the cost function $F(T_{V_2})$ will increase along with the time length T_{V_2} of the sleep cycle. During this period, the longer the time length of the sleep cycle is, the longer the MS will stay in the sleep period, the longer the average response time σ of data packets is, and thus the greater the cost function will be.

Conclusively, there is a minimal cost $F(T_{V_2})$ when the time length T_{V_2} of the sleep cycle is set to an optimal value. The optimal time length of the sleep cycle and the minimal cost for the different correlation parameters are summarized in Table 2.

6. Conclusions

How to improve the energy saving mechanism in battery powered mobile stations is one of the most pressing questions facing researchers of communication networks. As an enhancement of WiMAX, WiMAX 2 can save more energy. In this paper, concerning the real-time traffic including

the real-time and the best-effort traffic mixed, by setting the initial sleep window as one half of the fixed sleep cycle, we were able to build a multiple heterogeneous vacation queueing model. Considering the correlation of the data packets shown in real-time traffic with multimedia application, the arrival process was assumed to be D-MAP. The steady-state probability distribution for the queueing model was derived by using the method of an embedded Markov chain. For the performance measures, the average response time of data packets, the energy saving ratio, and the standard deviation of the number of data packets were given to evaluate the system performance with different correlation parameters. Moreover, an optimal design for the sleep cycle was given to maximize the energy efficiency with the constraint of user QoS in terms of the average response time of data packets. Finally, by using numerical results obtained from analysis and simulation, the influence of system parameters on the system performance was studied and the trade-off between different performance measures were investigated.

The research in this paper provides a theoretical basis for improving the power saving schemes in WiMAX 2 and has potential applications for solving other energy-related problems in modern communication networks.

Conflict of Interests

The authors declare that they do not have any commercial or associative interest that represents a conflict of interests in connection with the work submitted.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (no. 11201408) and Natural Science Foundation of Hebei Province (no. F2012203093), China, and was also supported in part by MEXT, Japan.

References

- [1] M. Jusoh, M. F. Jamlos, M. R. Kamarudin, and T. Sabapathy, "A reconfigurable WiMAX antenna for directional and broadside application," *International Journal of Antennas and Propagation*, vol. 2013, Article ID 405943, 8 pages, 2013.
- [2] IEEE, *IEEE 802.16m-2011—Standard for Local and Metropolitan Area Networks Part 16: Air Interface For Broadband Wireless Access Systems Amendment 3: Advanced Air Interface*, 2011.
- [3] L. Kong and D. H. K. Tsang, "Performance study of power saving classes of type I and II in IEEE 802.16e," in *Proceedings of the Annual IEEE Conference on Local Computer Networks (LCN'06)*, pp. 20–27, 2006.
- [4] S. Jin and W. Yue, "Performance analysis and evaluation for power saving class type III in IEEE 802.16e network," *Journal of Industrial and Management Optimization*, vol. 6, no. 3, pp. 691–708, 2010.
- [5] K. Lee and Y. Mun, "Enhanced power saving mechanism for Type I and Type II power saving classes in IEEE 802.16e," *IEICE Transactions on Communications*, vol. E94-B, no. 9, pp. 2642–2645, 2011.
- [6] H. Li and T. Yang, "Steady-state queue size distribution of discrete-time PH/Geo/1 retrial queues," *Mathematical and Computer Modelling*, vol. 30, no. 3-4, pp. 51–63, 1999.
- [7] Z. Niu, Y. Zhu, and V. Benetis, "A phase-type based Markov chain model for IEEE 802.16e sleep mode and its performance analysis," in *Proceedings of International Teletraffic Congress*, pp. 791–802, 2007.
- [8] Z. Saffer and M. Telek, "Analysis of BMAP vacation queue and its application to IEEE 802.16e sleep mode," *Journal of Industrial and Management Optimization*, vol. 6, no. 3, pp. 661–690, 2010.
- [9] E. Hwang, K. J. Kim, J. J. Son, and B. D. Choi, "The power-saving mechanism with periodic traffic indications in the IEEE 802.16e/m," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 1, pp. 319–334, 2010.
- [10] E. Hwang, K. J. Kim, J. J. Son, and B. D. Choi, "The power saving mechanism with binary exponential traffic indications in the IEEE 802.16e/m," *Queueing Systems*, vol. 62, no. 3, pp. 197–227, 2009.
- [11] S. Baek and B. D. Choi, "Performance of an efficient sleep mode operation for IEEE 802.16M," *Journal of Industrial and Management Optimization*, vol. 7, no. 3, pp. 623–639, 2011.
- [12] S. Jin, X. Chen, D. Qiao, and S. Choi, "Adaptive sleep mode management in IEEE 802.16m wireless metropolitan area networks," *Computer Networks*, vol. 55, no. 16, pp. 3774–3783, 2011.
- [13] C.-Y. Chen, C.-H. Hsu, and K.-T. Feng, "Performance analysis and comparison of sleep mode operation for IEEE 802.16m advanced broadband wireless networks," in *Proceedings of the IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC'10)*, pp. 1425–1430, 2010.
- [14] Z. Huo and X. Liu, "Performance analysis of sleep-mode with uplink/downlink traffics in IEEE 802.16m," in *Proceedings of Web Information Systems and Mining (WISM'11)*, pp. 426–433, 2011.
- [15] M. L. Chaudhry and U. C. Gupta, "Analysis of a finite-buffer bulk-service queue with discrete-Markovian arrival process: D-MAP/ $G^{a,b}/1/N$," *Naval Research Logistics*, vol. 50, no. 4, pp. 345–363, 2003.
- [16] S. Jin, Y. Zhao, W. Yue, and Z. Saffer, "Performance analysis and optimization of an adaptive admission control scheme in cognitive radio networks," *Mathematical Problems in Engineering*, vol. 2013, Article ID 727310, 10 pages, 2013.
- [17] F. Zhang and Z. Zhu, "A discrete-time retrial queue with vacations and two types of break-downs," *Journal of Applied Mathematics*, vol. 2013, Article ID 834731, 11 pages, 2013.
- [18] S. Havedanloo and H. R. Karimi, "Improving the performance metric of wireless sensor networks with clustering Markov chain model and multilevel fusion," *Mathematical Problems in Engineering*, vol. 2013, Article ID 783543, 11 pages, 2013.