PAPER

An efficient power saving mechanism for delay-guaranteed services in IEEE 802.16e*

Yunju PARK[†], *Nonmember and* **Gang Uk HWANG**[†a)], *Member*

SUMMARY As the IEEE 802.16e Wireless Metropolitan Access Network (WMAN) supports the mobility of a mobile station (MS), increasing MS power efficiency has become an important issue. In this paper, we analyze the sleep-mode operation for an efficient power saving mechanism for delay-guaranteed services in the IEEE 802.16e WMAN and observe the effects of the operating parameters related to this operation. For the analysis we use the $M/GI/1/K$ queueing system with multiple vacations, exhaustive services and setup times. In the analysis, we consider the power consumption during the wake-mode period as well as the sleep-mode period. As a performance measure for the power consumption, we propose the power consumption per unit time per effective arrival which considers the power consumption and the packet blocking probability simultaneously. In addition, since we consider delay-guaranteed services, the average packet response delay is also considered as a performance measure. Based on the performance measures, we obtain the optimal sleep-mode operation which minimizes the power consumption per unit time per effective arrival with a given delay requirement. Numerical studies are also provided to investigate the system performance and to show how to achieve our objective.

key words: *sleep-mode operation, power saving, IEEE 802.16e, power consumption, average packet response delay*

1. Introduction

As wireless internet services are rapidly expanding, it is an important task to provide high speed and high quality wireless services. To meet these demands, in the Wireless Metropolitan Access Network (WMAN), mobility and power management of a Mobile Station (MS) become important issues, and to solve the issues the IEEE 802.16 standard [1, 2] has been extended to the IEEE 802.16e standard where the handover process and the sleep mode operation are included [3].

In the IEEE 802.16e standard [3], the sleep-mode operation has three types of the Power Saving Classes. Power Saving Class is a group of connections which have common demand properties. Power Saving Class of type I is recommended for the connections of Best Effort (BE) and Non-Real-Time Variable Rate (NRT-VR) services. Power Saving Class of type II is recommended for the connections of Unsolicited grant service (UGS) and Real-Time Variable Rate (RT-VR) services. Power Saving Class of type III is recommended for multicast connections and for management operations.

Regarding the analysis of the sleep-mode operation in IEEE 802.16e, Nga *et al.* [4] analyzed a numerical model to determine the operating parameters in the sleep-mode operation and proposed a delay guaranteed energy saving algorithm to minimize energy consumption with a given MAC (Medium Access Control) SDU (Service Data Unit) response delay. Xiao [5] considered a sleep-mode scheme for the power saving mechanism and analyzed the effects of the operating parameters. Seo *et al.* [6] used the $M/GI/1/K$ queueing system with multiple vacations and considered the dropping probability of packets and the mean waiting times of packets as the performance measures. Jang *et al.* [7] simulated the sleep-mode operation of the Power Saving Class of type I and II, and found the optimal values of operating parameters to satisfy different QoS requirements. Kim *et al.* [8] introduced an efficient power management mechanism which takes into account the remaining energy. Kim *et al.* [9] modeled the sleep mode operation in the IEEE 802.16e MAC and evaluated the effect of operating parameters on the performance of power management by considering the average interarrival time of MAC frames. They used simulation to evaluate the performance. Xu *et al.* [10] discussed a novel adaptive sleep-mode scheme which considered quick responses to the packet arrival events. Dong *et al.* [11] modelled the sleep-mode operation where two-type-returning from sleep mode is considered. Most studies modelled the sleep-mode period, but they focused on the analysis of the sleep-mode period only and did not consider the wake-mode period and the queueing effect of the system. However, in our paper, we analyze the packet response delay and the power consumption during the wake-mode period as well as the sleep-mode period.

The operation of IEEE 802.16e system is based on frame units of 5 ms, and the sleep-mode operation in IEEE 802.16e is performed between one base station (BS) and one MS. In this paper, we focus on downlink transmission from a BS to an MS, and assume that there is a finite size buffer in the BS to develop a system model close to the real one. The finite size buffer in the BS accommodated packets addressed to the MS

[†] The authors are with the Department of Mathematical Sciences and Telecommunication Engineering Program in Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

[†] This paper was presented in part at the 65th IEEE Vehicular Technology Conference VTC2007-spring, Dublin, Ireland, April 22-25, 2007.

a) E-mail: guhwang@kaist.edu

* This research was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (No. R01-2007-000-20053-0).

with the Power Saving Class of type I packets. The detailed sleep-mode operation of this type will be given in section 2. In an IEEE 802.16e network, the resources are shared by all MS in the network, and some amount of resources are assigned to our MS at each frame. So, the downlink transmission system can be modelled by a single server queueing system. The amount of resources allocated to our MS determines the packet transmission time. The amount of resources allocated to our MS is not fixed in general. In addition, the mobility and the wireless channel condition may affect the packet service process through packet transmission errors. Such effects can be approximately taken into consideration by using a suitable service time distribution. Here, the service time is defined by the time period needed to transmit the HOL (Head Of Line) packet in the buffer successfully. Accordingly, to consider such effects, we assume that the service times of packets have a general service time distribution. As the arrival process to the buffer in BS, we adopt the Poisson process for the convenience in the analysis [4–11]. In addition, due to the sleep-mode operation, our queueing system has vacations.

Based on above assumptions, to analyze the sleep-mode operation mathematically, the buffer in the BS is modelled by the $M/GI/1/K$ queueing system with multiple vacations, exhaustive services and setup times [12, 13]. The contributions of this paper are as follows. First, we mathematically model the sleep-mode operation as exactly as possible and obtain the packet blocking probability, the power consumption per unit time and the average packet response delay. Second, as performance measures, we consider the average packet response delay and the power consumption per unit time per effective arrival, which is newly proposed to combine the average packet blocking probability and the power consumption per unit time simultaneously. Third, we provide a detailed procedure to get the optimal sleep-mode operation which satisfies the delay requirement and minimizes the power consumption per unit time per effective arrival.

The rest of this paper is organized as follows. In section 2, the sleep-mode operation in the IEEE 802.16e standard is described. In section 3, we use the $M/GI/1/K$ queueing system with multiple vacations, exhaustive services and setup times to analyze the sleep-mode operation. In section 4 and 5, we analyze the system behaviors during the vacation and busy periods, respectively. In section 6, we obtain the packet blocking probability, the power consumption per unit time and the average packet response delay. Based on our analysis, we propose a detailed procedure to get the optimal sleep-mode operation. In section 7, we give our conclusions.

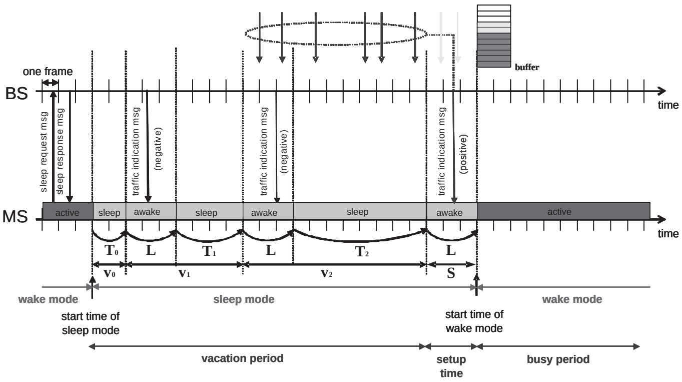

Fig. 1 The sleep-mode operation in IEEE802.16e

2. The Operation of Sleep-Mode in IEEE802.16e

We consider downlink transmission from a BS to an MS where the Power Saving Class of type I is used. In this case, the MS has two modes: sleep-mode and wake-mode. During a sleep-mode period, the MS powers down to reduce the battery consumption and there is no packet transmission between the BS and the MS. So, the buffer in the BS accommodates incoming packets addressed to the MS until the sleep-mode period ends. When the sleep-mode period ends, a wake-mode period starts immediately and there are packet transmissions between the BS and the MS during the wake-mode period. After the wake-mode period, a new sleep-mode period begins again and this alternating procedure will continue.

To enter a new sleep-mode period after a wake-mode period, the MS sends a Sleep Request message (MOB_SLP-REQ) to the BS and waits for a Sleep Response message (MOB_SLP-RSP) from the BS. The MOB_SLP-REQ contains the relevant parameters regarding the sleep-mode period such as *initial-sleep window, final-sleep window base, listening window, final-sleep window exponent and start frame number for the first sleep window.* When the BS receives the MOB_SLP-REQ, if there is no downlink traffic for the MS, then the BS sends a positive MOB_SLP-RSP message to the MS with the same parameters as in the MOB_SLP-REQ message that it has received. Otherwise, the BS sends a negative MOB_SLP-RSP message. After receiving the MOB_SLP-RSP, the MS can determine whether to begin a new sleep-mode period or not. If it has received a negative MOB_SLP-RSP (i.e., no approval message), it continues to be in wake-mode and waits for another packet transmissions. If it has received a positive MOB_SLP-RSP (i.e., approval message), it begins a new sleep-mode period at the frame specified as *start frame number for the first sleep window.* A sleep-mode period may consist of a single or multiple sleep intervals as shown in Fig. 1. The length of the first sleep interval is equal to the *initial-sleep* *window*, denoted by T_0. After the first sleep interval, the MS wakes for a fixed time period, called the listening interval, of length *listening window* L, to check the Traffic Indication message (MOB_TRF-IND). The message is broadcasted by the BS during the listening interval. It indicates the presence of the buffered traffic addressed to the MS in the BS. If the MS has received a negative MOB_TRF-IND, i.e., no buffered traffic for it, it goes back to the sleep-mode again to start the second sleep interval with the doubled length $T_1(=2T_0)$. Otherwise, the MS begins a new wake-mode period after the listening interval. The MS keeps the above procedure during the sleep-mode period until it receives a positive MOB_TRF-IND. In the IEEE 802.16e standard [3], the length T_i of the i-th sleep interval in a sleep-mode period, if any, is computed as follows:

$$\begin{aligned} T_0 &= \textit{initial-sleep window}, \\ T_i &= \min\{2T_{i-1},\ T_{max}\}, \qquad i \geq 1, \\ T_{max} &= \textit{final-sleep window base} \cdot 2^{\textit{final-sleep window exponent}}. \end{aligned}$$

3. Mathematical Modelling of The Sleep-Mode Operation

In this section, we consider the buffer in the BS which accommodates packets addressed to the MS, and analyze the operation of the sleep-mode in IEEE 802.16e explained in section 2.

We assume that the packet arrival process follows a Poisson process with rate λ and the service times of packets are independent and identically distributed (i.i.d) with common distribution function $F(x)$. We also assume that the buffer in the BS for the MS is of finite size $K-1$, and single server for packet transmission. Then, if there are K packets in the system including the packet being transmitted, newly arriving packets are blocked and discarded. For this reason, K is called the system size from now on. The packets in the buffer are transmitted based on the FCFS (First-Come-First-Served) service discipline.

Since the operation of IEEE 802.16e is based on frame times, a frame time is considered as a unit time and we assume that the time axis is divided into unit times in our analysis. To model our system, we first consider a sleep-mode period. The sleep-mode period starts with the first sleep interval of length T_0. This first sleep interval is called *the* 0*-th sub-vacation interval* in our model. If there is no packet during the 0-th sub-vacation interval, the MS receives a negative MOB_TRF-IND message during the following listening interval and starts immediately the second sleep interval after the listening interval. Since the switching to the wake-mode for the MS after the second sleep interval depends on whether there is at least one packet arrival during the listening and second sleep intervals, the sum of the listening and second sleep intervals is called *the first sub-vacation interval.* Similarly, the MS has the $(i+1)$-th sub-vacation interval if there is no packet during the i-th sub-vacation interval which consists of a listening interval and the $(i+1)$-th sleep interval.

Now, assume that the MS is at the end of the n-th sub-vacation interval and that there is at least one packet arrival during the n-th sub-vacation interval. In this case, the MS receives a positive MOB_TRF-IND message from the BS during the listening interval following the n-th sub-vacation interval. So, the MS ends the sleep-mode period consisting of $n+1$ sub-vacation intervals, called *a vacation period* in our model, and starts a new wake-mode period, called *a busy period* in our model, to receive packets from the BS. Note that there is always a listening interval between a vacation period and the following busy period, which is called *the setup time period* in our model. For details, refer to Fig. 1.

From the above assumptions and explanation, our system can be modelled by the discrete time $M/GI/1/K$ queueing system with multiple vacations, exhaustive services and setup times [12, 13]. Here, GI implies that service times are general and independent. Note that our system has the exhaustive service discipline because the MS begins the vacation period only when there is no packet for the BS to transmit. If we assume that the time to transmit the control messages such as MOB_SLP-REQ, MOB_SLP-RSP and MOB_TRF-IND is zero, the i-th sub-vacation interval, denoted by v_i, is given by

$$v_i = \begin{cases} T_0, & i = 0, \\ L + T_i, & i = 1, 2, ... \end{cases}$$

Recall that L is the length of a listening interval, T_0 is the initial sleep interval, and T_i is given by

$$T_i = \min\{2T_{i-1}, T_{max}\}, \qquad i = 1, 2, ...$$

Since T_i's are all fixed, all sub-vacation intervals, $v_i, i = 0, 1, 2, ...$, are of fixed lengths.

Let V, S and B be the vacation period, setup time period and busy period in our model, respectively. Then, we can define a service cycle C by $C = V+S+B$, i.e., a service cycle consists of a vacation period, a setup time period and a busy period.

In the subsequent sections, we analyze the system behaviors during the vacation period and busy period separately and obtain the performance measures such as the power consumption per unit time per effective arrival and the average packet response delay.

4. The Vacation Period Analysis

In this section, we analyze the length of a vacation period and the number of sub-vacation intervals in steady state. To do this, we first obtain the probability mass function of a vacation period. By the definition of a vacation period V given in section 3, the probability mass function of V is given by

$$V = \begin{cases} S_0 & \text{with prob. } 1 - e^{-\lambda S_0}, \\ S_n & \text{with prob. } e^{-\lambda S_{n-1}}(1 - e^{-\lambda v_n}),\ n \geq 1, \end{cases} \tag{1}$$

where $S_n = \sum_{i=0}^{n} v_i$ and $S_0 = v_0$.

For our analysis, we use M to denote the index of the sub-vacation interval such that $v_j = L + T_{max}$ for all $j \geq M$ and $v_j < L + T_{max}$ for all $0 \leq j < M$. That is, $M = \min\{j : v_j = L + T_{max}\}$. Note that the value of M depends on T_0 and T_{max}. For example, if $T_0 = 2$ and $T_{max} = 1024$, then $M = 9$. Then, from equation (1) the expectation $E[V]$ of a vacation period is given by

$$\begin{aligned} \mathbf{E}[V] &= \sum_{j=0}^{\infty} S_j \mathbf{Pr}[V = S_j] \\ &= S_0 + \sum_{j=1}^{M} e^{-\lambda S_{j-1}} v_j + \frac{v_M}{1 - e^{-\lambda v_M}} e^{-\lambda S_M}. \end{aligned}$$

The detail derivation is given in Appendix A.1.

Next, let N_I denote the maximum of sub-vacation interval indexes during a vacation period in steady state. By a similar argument as above, we also obtain the expectation $E[N_I]$ as follows:

$$\begin{aligned} \mathbf{E}[N_I] &= \sum_{j=1}^{\infty} j \mathbf{Pr}[V = S_{j-1}] \\ &= 1 + \sum_{j=0}^{M} e^{-\lambda S_j} + \frac{e^{-\lambda v_M}}{1 - e^{-\lambda v_M}} e^{-\lambda S_M}. \end{aligned}$$

The detail derivation is given in Appendix A.2.

5. The Busy Period Analysis

In this section, we analyze the length of a busy period in steady state. Since the length of a busy period is closely related with the number of backlogged packets at the beginning of the busy period, we start with the analysis of the number of backlogged packets at the end of a vacation period.

5.1 Distribution of the number of packets at the end of a vacation period

Let N_V be the number of backlogged packets at the end of a vacation period in steady state. By the definition, the distribution of N_V is obtained as follows:

$$\mathbf{Pr}[N_V = i] = \sum_{n=0}^{\infty} \mathbf{Pr}[N_V = i | N_I = n] \mathbf{Pr}[N_I = n].$$

Here, recall that N_I denotes the maximum of sub-vacation interval indexes during a vacation period in steady state. Observing that there should be at least one packet at the end of a vacation period, we have the probability mass function of N_V as follows:

Theorem 1: The probability mass function of the number N_V of backlogged packets at the end of a vacation period is given by

$$\begin{aligned} \mathbf{Pr}[N_V = i] &= \sum_{n=0}^{M} \frac{(\lambda v_n)^i e^{-\lambda S_n}}{i!} + \frac{\frac{(\lambda v_M)^i}{i!} e^{-\lambda v_M}}{1 - e^{-\lambda v_M}} e^{-\lambda S_M}, \\ & \qquad (i = 1, 2, ..., K-1) \\ \mathbf{Pr}[N_V = K] &= 1 - \sum_{i=1}^{K-1} \mathbf{Pr}[N_V = i]. \end{aligned}$$

The proof of Theorem 1 is given in Appendix A.3.

5.2 Distribution of the number of packets at the beginning of a busy period

Let N_S be the number of backlogged packets in the system at the beginning of a busy period in steady state. Since there is always a setup time period between a vacation period and a busy period, N_S satisfies $N_S = N_V + A_S$ where A_S denotes the number of packets newly arriving during the setup time period. Then, by Theorem 1 the distribution of N_S is derived as follows.

Theorem 2: The probability mass function of the number N_S of backlogged packets at the beginning of a busy period is given by

$$\begin{aligned} \mathbf{Pr}[N_S = j] &= \sum_{i=1}^{j} \frac{(\lambda S)^{j-i} e^{-\lambda S}}{(j-i)!} \times \\ & \left\{ \sum_{n=0}^{M} \frac{(\lambda v_n)^i e^{-\lambda S_n}}{i!} + \frac{\frac{(\lambda v_M)^i e^{-\lambda S_M}}{i!} e^{-\lambda v_M}}{1 - e^{-\lambda v_M}} \right\}, \\ & \qquad (j = 1, 2, ..., K-1) \\ \mathbf{Pr}[N_S = K] &= 1 - \sum_{j=1}^{K-1} \mathbf{Pr}[N_S = j]. \end{aligned}$$

The proof of Theorem 2 is given in Appendix A.4.

5.3 The length of a busy period

In this subsection, we analyze the length of a busy period based on the results obtained in subsections 5.1 and 5.2. Recall that packets in the buffer are transmitted based on the FCFS discipline. However, the length of a busy period does not depend on the order in which packets in the buffer are transmitted [12, 14–16]. That is, the length of a busy period of the system with the FCFS discipline is identical to that of the system with the LCFS (Last-Come-First-Served) discipline. So, for convenience in the analysis, we consider a new system which is identical to our system except that the LCFS discipline is used. Note that the method of considering a system with the LCFS discipline is widely used for the busy period analysis, e.g., [14].

Now we assume that there are j backlogged packets in the buffer at the beginning of a busy period. Then, due to the LCFS service discipline the busy period generated by the j backlogged packets can be divided into j sub-busy periods, each of which is generated by the service of a backlogged packet as follows. When a busy period starts, our system immediately starts the service of the j-th backlogged packet and continue its service for all subsequent packets that newly arrive at the system until it can start the service of the $(j-1)$-th backlogged packet. That is, the service of the j-th backlogged packet generates the first sub-busy period which ends with the start of the service of the $(j-1)$-th backlogged packet. In addition, since there are only $K-j$ empty waiting rooms in the buffer for this case, the length of the first sub-busy period in our system is identical to the length of a busy period of the ordinary $M/GI/1/K-j+1$ queueing system. Similarly, the i-th sub-busy period is generated by the service of the $(j-i+1)$-th backlogged packet. Since there are $K-j+i-1$ empty waiting rooms in the buffer for this case, the length of the i-th sub-busy period is identical to the length of a busy period of the ordinary $M/GI/1/K-j+i$ queueing system.

Let B_m and $B_m^*(s)$ be the length of a busy period and its LST (Laplace-Stieltjes Transform) in the ordinary $M/GI/1/m+1$ queueing system. Then, we have the following Theorem [16].

Theorem 3: The LST of the length of a busy period for the ordinary $M/GI/1/m+1$ queueing system with m waiting rooms is

$$B_m{}^*(s) = \frac{u_0(s)}{1 - \sum_{k=1}^{m-1} u_k(s) \prod_{j=m-k+1}^{m-1} B_j{}^*(s) - \sum_{k=m}^{\infty} u_k(s) \prod_{j=1}^{m-1} B_j{}^*(s)}.$$

The expectation $\mathbf{E}[B_m]$ is

$$\mathbf{E}[B_m] = \frac{1}{q_0} \left\{ \mathbf{E}[X] + \sum_{j=1}^{m-1} \mathbf{E}[B_j] \mathbf{Q}_{m-j} \right\},$$

where $u_k(s) = \int_0^\infty \frac{(\lambda x)^k}{k!} e^{-(\lambda+s)x} dF(x)$, $F(x)$ is the distribution function of the service time, $q_k = u_k(0)$ and $\mathbf{Q}_j = 1 - \sum_{k=0}^{j} q_k$. (Note that when the lower limit of a product (a summation, resp.) is greater that the upper limit, the product (the summation, resp.) is taken to be 1 (0, resp.).)

Then, from our observation above the length of a busy period $B(j)$ generated by j backlogged packets in the buffer is given by

$$B(j) \overset{d}{=} B_{K-j} + B_{K-j+1} + \cdots + B_{K-1}. \tag{2}$$

Therefore, using equation (2), Theorem 2 and Theorem 3, we can derive the expectation of a busy period B in our system as follows.

Theorem 4: The expectation of a busy period in our system is given by

$$\mathbf{E}[B] = \sum_{j=1}^{K-1} \sum_{i=K-j}^{K-1} \mathbf{E}[B_i] \mathbf{Pr}\left[N_S = j\right] + \left(\mathbf{E}[X] + \sum_{i=1}^{K-1} \mathbf{E}[B_i] \right) \mathbf{Pr}[N_S = K],$$

where X is the service time of a packet.

The proof of Theorem 4 is given in Appendix A.5.

6. Performance Analysis

In this section, we propose two performance measures, called the *the power consumption per unit time per effective arrival* and *the average packet response delay*. Based on these two performance measures, for given arrival rate λ, the first sleep interval T_0, the system size K and the mean service time of a packet $\mathbf{E}[X]$, we obtain the optimal T_{max} value. Then, by using the results of T_{max} for each T_0, we will determine the optimal set of (T_0, T_{max}) for a given delay requirement to minimize the power consumption per unit time per effective arrival.

The parameters related with the power consumption are as follows. Let E_S, E_W and E_L be the power consumption units per unit time in sleep-mode, wake-mode and a listening interval, respectively. In addition, since the MS consumes the power to switch the mode, let $E_{on-switch}$ and $E_{off-switch}$ be the power consumption units for the switch-on action and the switch-off action, respectively. The switch-on (switch-off, resp.) action means that the MS changes its state from sleep (wake or listen, resp.) to listen (sleep, resp.).

6.1 The power consumption per unit time per effective arrival

In this subsection, we obtain the power consumption per unit time per effective arrival in our system. To do this, we first derive the power consumption per unit time of an MS. Then, we get the packet blocking probability. Let p_V, p_S and p_B be the amounts of total power consumption during a vacation period, a setup time period and a busy period, respectively. Then, by the definitions the expectations $\mathbf{E}[p_V], \mathbf{E}[p_S]$ and $\mathbf{E}[p_B]$ are given as follows:

$$\begin{aligned}\mathbf{E}[p_V] &= E_S S_0 + \sum_{j=1}^{M} e^{-\lambda S_{j-1}}(E_S T_j + E_L L) \\ &\quad + \frac{E_S T_M + E_L L}{1 - e^{-\lambda v_M}} e^{-\lambda S_M}, \\ \mathbf{E}[p_S] &= E_L E[S] = E_L L, \\ \mathbf{E}[p_B] &= E_W E[B].\end{aligned}$$

Note that $\mathbf{E}[p_V]$ can be derived from $\mathbf{E}[V]$ in section 4. Then, the power consumption per unit time of the MS, which is denoted by PC and defined by

$$PC = \frac{\mathbf{E}[\text{total power consumption in a cycle}]}{\mathbf{E}[\text{total length of a cycle}]},$$

is given by

$$PC = \frac{\mathbf{E}[p_V] + \mathbf{E}[p_S] + \mathbf{E}[p_B] + \mathbf{E}[N_I](E_{on-switch} + E_{off-switch})}{\mathbf{E}[C]} \tag{3}$$

Let P_B and ρ be the packet blocking probability and the probability that the server is busy at an arbitrary time in steady state, respectively. Since our system has a single server and is of finite size K, it satisfies that

$$\rho = \lambda(1 - P_B)E[X] \tag{4}$$

where $E[X]$ denotes the mean service time of a packet. On the other hand, by its definition ρ can be also obtained as follows:

$$\rho = \frac{\mathbf{E}[B]}{\mathbf{E}[C]}. \tag{5}$$

Then, by combining equations (4) and (5), we finally obtain P_B as follows:

$$P_B = 1 - \frac{\mathbf{E}[B]}{\lambda \mathbf{E}[X]\mathbf{E}[C]}. \tag{6}$$

To consider two performance measures P_B and PC simultaneously, we propose a new combined performance measure PC_e, called the power consumption per unit time per effective arrival and defined by

$$PC_e = \frac{PC}{\lambda(1 - P_B)}.$$

Note that PC_e is the average actual power consumption to transmit a packet.

6.2 The average packet response delay

In this subsection, we propose the average packet response delay defined by the sum of queueing delay in the buffer and transmission delay from the BS to the MS. To do this, first, we obtain the distribution of the number of backlogged packets in the system (called the queue length) immediately after a service completion by applying the embedded Markov chain method. Second, we derive the queue length distribution at an arbitrary time. We consider a set of embedded Markov points which are those points in time when packets leave the system after service completion. Let π_j^d be the steady state probability that j packets are left in the system immediately after service completion $(0 \le j \le K - 1)$. Let L_n be the number of packets left behind in the system immediately after the n-th Markov point $(n = 1, 2, ...)$. Then the π_j^d is represented as follows.

Theorem 5: The steady state probability π_j^d that j packets are left in the system immediately after a service completion is given by

$$\begin{aligned}\pi_j^d &= \lim_{n\to\infty} \mathbf{Pr}[L_n = j], \qquad 0 \le j \le K-1 \\ &= \begin{cases} \dfrac{1}{\sum_{j=0}^{K-1} \pi_j'}, & j = 0 \\ \pi_0^d \pi_j', & 1 \le j \le K-1, \end{cases}\end{aligned}$$

where

$$\begin{aligned}\pi_0' &= 1, \\ \pi_{j+1}' &= \frac{1}{a_0}\left\{\pi_j' - \sum_{i=1}^{j} \pi_i' a_{j-i+1} - \sum_{k=1}^{j+1} a_{j-k+1}\mathbf{Pr}[N_S = k]\right\}, \\ &\qquad\qquad \text{for } 0 \le j \le K-2, \\ a_k &\triangleq \int_0^\infty \frac{(\lambda x)^k}{k!} e^{-\lambda x} dF(x), \ k = 0, 1, ...,\end{aligned}$$

and $\mathbf{Pr}[N_S = k]$ are given in Theorem 2.

The proof of Theorem 5 is given in Appendix A.6.

Let Q_k be the probability that there are k packets in the system including the packet being transmitted at an arbitrary time $(k = 0, 1, ..., K)$. To derive Q_k by using π_j^d, we let π_j^a be the probability that an arriving packet finds j packets in the system, $(j = 0, 1, ..., K)$. From the PASTA (Poisson Arrivals See Time Average) property,

$$\pi_j^a = Q_j, \quad 0 \le j \le K. \tag{7}$$

Since the state changes only by unit steps, by Burke's theorem [12],

$$\pi_j^a = (1 - Q_K)\pi_j^d, \qquad 0 \le j \le K-1. \tag{8}$$

Therefore, combining equations (7) and (8), we obtain

$$Q_j = (1 - Q_K)\pi_j^d, \qquad 0 \le j \le K-1, \tag{9}$$

Note that Q_K is the blocking probability. So, by equation (6) Q_K is, in fact, given as

$$Q_K = 1 - \frac{\mathbf{E}[B]}{\lambda \mathbf{E}[X]\mathbf{E}[C]}.$$

Then using equation (9) and Theorem 5 we can obtain $\{Q_j | \, 0 \le j \le K\}$. Let L_e and D be the number of packets in the system at an arbitrary time and the response delay of an arbitrary packet in the system, respectively. Then the expectation $\mathbf{E}[L_e]$ is given by

$$\mathbf{E}[L_e] = \sum_{k=0}^{K} kQ_k \tag{10}$$

Then, from Little's formula, the average packet response delay $E[D]$ is derived as follows:

$$\mathbf{E}[D] = \frac{1}{\lambda(1-Q_K)}\mathbf{E}[L_e]. \tag{11}$$

6.3 The Procedure to get the optimal sleep-mode operation

For simulation studies, we develop a MATLAB code to simulate the system. The simulation condition is as follows. Since the sleep-mode operation is performed between one BS and one MS and we consider downlink transmission, one BS and one MS are considered in our simulation. We assume an ideal wireless channel model and generate 3×10^5 frames for each simulation. We also assume that the time to transmit the control messages is zero. Here, the control messages are MOB_SLP-REQ, MOB_SLP-RSP, and MOB_TRF-IND. Since the purpose of the standardized sleep-mode operation in IEEE 802.16e is to save the power consumption for low-rate traffic environment, we use the expected interarrival time $I_A(=1/\lambda) = 8, 16, 32, 64$ frames. For all examples in this subsection, otherwise mentioned, we assume that the service time of a packet has the geometric distribution with mean $\mathbf{E}[X] = 2$, and the system size K is 10. Note that the choice of $\mathbf{E}[X] = 2$ and $K = 10$ is an example. The system size K is not a critical parameter in our model due to the fact that the packet blocking probability is very low in low-rate traffic environment. We can choose any other values of $\mathbf{E}[X]$ and K in our analysis. We follow the guidelines of IEEE 802.16e standard for the initial sleep window T_0, the final sleep window T_{max}, and the listening interval L. The length L of a listening interval is fixed. We assume that the length of L is equal to 1. Since there is no general information of the actual parameters for the power consumption units, we use $E_S : E_L : E_W = 1 : 10 : 10$ and $E_{on-switch} : E_{off-switch} = 30 : 20$, as given in [5,10,17–19].

In this subsection, we first propose a procedure to get the optimal value of T_{max} for the delay-guaranteed services. That is, for given mean service time $\mathbf{E}[X]$, system size K and the initial sleep interval T_0 we show how to get the optimal value of T_{max} based on two performance measures – the power consumption per unit time per effective arrival in subsection 6.1 and the average packet response delay in subsection 6.2. Then, by investigating the results, we show how to determine the optimal values of T_0 and T_{max} for a given delay requirement to minimize the power consumption per unit time per effective arrival.

In Fig. 2, we change the value of T_{max} and the packet arrival rate λ, and plot the resulting value of $\mathbf{E}[D]$ and PC_e for given mean service time $\mathbf{E}[X]$, system size K and the initial sleep interval T_0. In the figure, 'Sim', 'Num' and 'I_A' denotes the results obtained by simulation, the results obtained by our analysis and the expected interarrival time $(= 1/\lambda)$ of the packets (in frame), respectively. As seen from the figure, our analytic results are well matched with the simulation results, which partially verifies the validity of our analysis. From the figure, we also see that the value of T_{max} affects $\mathbf{E}[D]$ and PC_e for each fixed value of I_A, and the effect of T_{max} on $\mathbf{E}[D]$ and PC_e becomes more significant as I_A increases. The reason for this is that, as I_A decreases, it occurs more frequently that the sleep-mode period is terminated before the length of a vacation period reaches the maximum possible value (related with T_{max}).

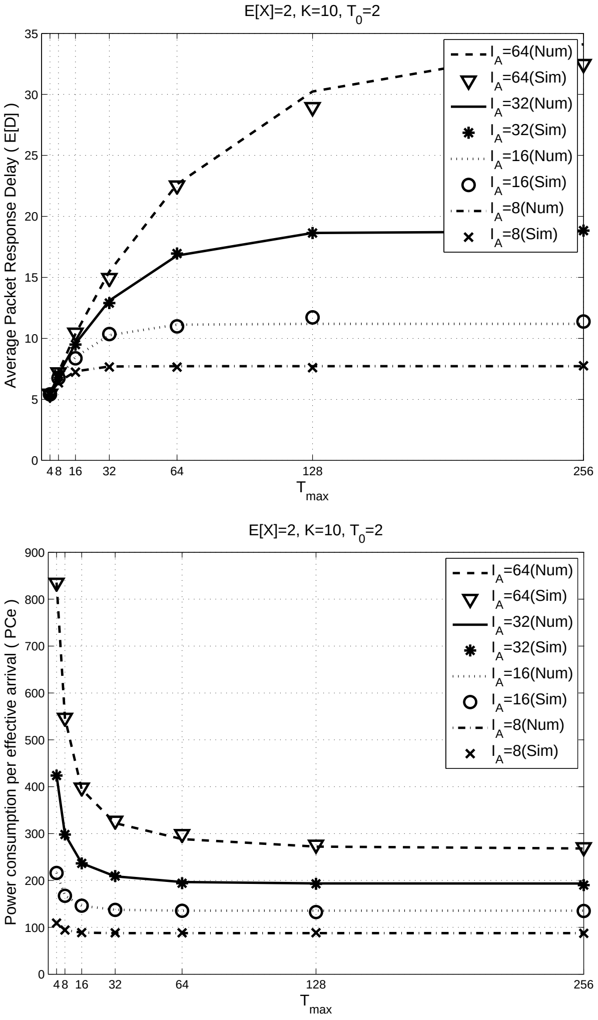

Fig. 2 $\mathbf{E}[D]$ and PC_e for different I_A and T_{max}

Another observation from Fig. 2 is that, in most cases except $I_A = 64$ frame, when T_{max} is greater than 128, T_{max} does not affect $\mathbf{E}[D]$ and PC_e any more. This result is in accordance with previously known results in [11,21]. That is, when the value I_A is small (the high-rate traffic environment in this case), the parameter T_{max} does not affect the system performance significantly. However, since the sleep-mode operation is

Table 1 Optimal value of T_{max} given delay requirement (T_0=2)

I_A delay requirement	64	32	16	8
10 frames	16	16	16	1024
15 frames	16	32	1024	1024
20 frames	32	1024	1024	1024
25 frames	64	1024	1024	1024
30 frames	128	1024	1024	1024

considered to save the power consumption for low-rate traffic environment (under which the effect of T_{max} is significant), finding a suitable value of T_{max} is important for low-rate traffic environment.

To get the optimal value of T_{max}, first note that the average packet response delay $\mathbf{E}[D]$ increases and the power consumption per unit time per effective arrival PC_e decreases as we change the value of T_{max}. This implies that we can not improve $\mathbf{E}[D]$ and PC_e simultaneously. Hence, we assume that a certain level of delay requirement is given. Then we consider the following twofold procedure. In the first step, for given mean service time $\mathbf{E}[X]$, system size K, initial sleep interval T_0 and interarrival time I_A, we select the values of T_{max} with which the value of $\mathbf{E}[D]$ is lower than the delay requirement. In the second step, among the selected values of T_{max}, we then choose the optimal value of T_{max} which minimizes the value of PC_e. In all procedures in this subsection, we assume the value of T_{max} is of the form $2^M T_0$ for simplicity.

Using Fig. 2 and the twofold procedure, we can choose the optimal value of T_{max} as follows when the other parameters are given as $\mathbf{E}[X] = 2$, $K = 10$ and $T_0 = 2$. For instance, assume that the delay requirement is 15 frames and $I_A = 32$ frames. Then, from the delay requirement we can first select $T_{max} = 4, 8, 16, 32$ because they result in the average packet response delay lower than the delay requirement. Among these selected values of T_{max}, we check the corresponding values of PC_e from Fig. 2 and finally choose $T_{max} = 32$ as the optimal value which minimizes PC_e. For other values of delay requirement and interarrival times, we can perform the same procedure to obtain the optimal value of T_{max} and the results are summarized in Table 1. Note that the maximum length of T_{max} is fixed and equal to 1024 [3], so we do not consider values greater than 1024 for T_{max} to obtain Table 1.

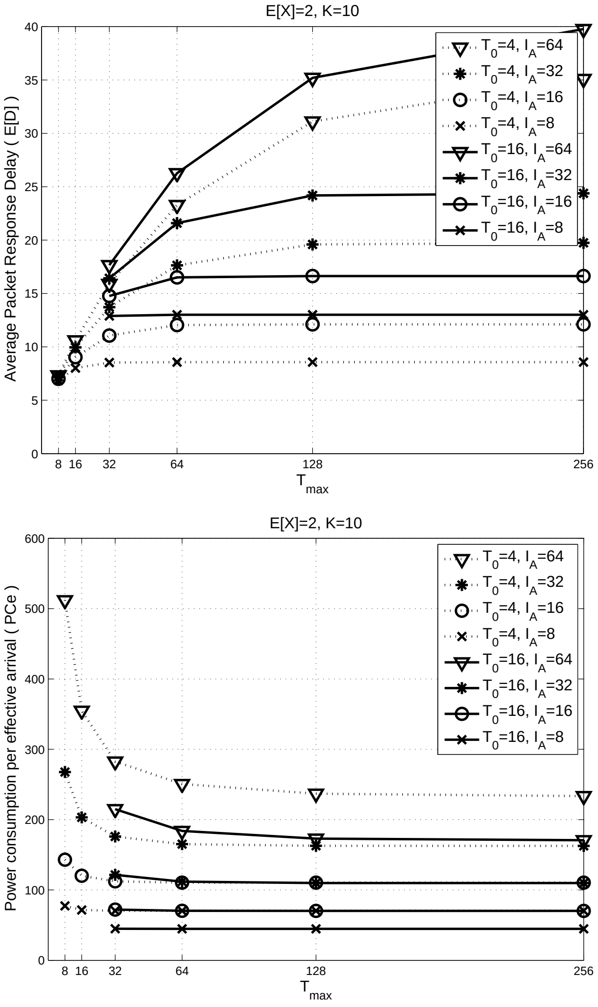

Fig. 3 The effect of T_0

In Fig. 3, we change the value of T_0 and plot the resulting value of $\mathbf{E}[D]$ and PC_e. From the figure, we see that, as T_0 increases, $\mathbf{E}[D]$ increases and PC_e decreases. The reason for this is as follows. As T_0 increases, the sleep intervals become longer and accordingly the number of packets stored in the buffer during the sleep intervals increases. This results in the increase in $\mathbf{E}[D]$. On the contrary, as T_0 increases, the MS has longer sleep-mode periods and switches its mode less frequently. Noting that the power consumption units $E_{on-switch}$ and $E_{off-switch}$ are greater than the power consumption units E_S, E_W and E_L in sleep-mode, wake-mode and listening intervals, the less the MS switches its mode, the less it consumes its power. Thus, PC_e decreases. The optimal values of T_{max} for various values of T_0 are summarized in Table 2 and 3. In Table 3, the notation * means that the optimal value of T_{max} does not exist. This can happen since the value of T_0 is relatively large compared with the delay requirement. That is, if T_0 is large, then the length of sleep-mode period is large. So, the average packet response delay is large and accordingly the delay requirement is violated for any value of T_{max} in this case.

Finally, using the above procedure for each given T_0 and other parameters, we can determine the opti-

Table 2 Optimal value of T_{max} given delay requirement (T_0=4)

I_A delay requirement	64	32	16	8
10 frames	8	16	16	1024
15 frames	16	32	1024	1024
20 frames	32	1024	1024	1024
25 frames	64	1024	1024	1024
30 frames	64	1024	1024	1024

Table 3 Optimal value of T_{max} given delay requirement (T_0=16)

I_A delay requirement	64	32	16	8
10 frames	*	*	*	*
15 frames	*	*	32	1024
20 frames	32	32	1024	1024
25 frames	32	1024	1024	1024
30 frames	64	1024	1024	1024

Table 4 Optimal values of T_0 and T_{max} ($I_A = 64$ and delay requirement=15 frames)

(T_0, T_{max})	PC_e
(2, 16)	394.3040
(3, 24)	321.8611
(4, 16)	354.7521
(5, 20)	312.7799
(6, 24)	283.7988
(7, 14)	349.4597
(8, 16)	320.6569
(9, 18)	297.8579
(10, 20)	279.3499
(11, 22)	264.0157
(12, 24)	251.0954
(13, *)	-
(14, *)	-
(15, *)	-
(16, *)	-

mal values of T_0 and T_{max} that minimize the power consumption per unit time per effective arrival PC_e. For instance, assume that the delay requirement is 15 frames and the interarrival time I_A is 64 frames. In this case, we compute the optimal value of T_{max} and the corresponding value of PC_e for each value of T_0, and the results are given in Table 4. Then, obviously the optimal values of (T_0, T_{max}) are (12, 24) because they minimize PC_e. From this result, we see that a relatively large value of T_0 would minimize the power consumption per unit time per effective arrival for low-rate traffic environment. For another delay requirements and traffic conditions, we use the same procedure to determine the optimal values of T_0 and T_{max}.

7. Conclusions

In this paper, we model and analyze the sleep-mode operation in IEEE 802.16e WMAN using the $M/GI/1/K$ queueing system with multiple vacations, exhaustive services and setup times. We analyze the wake-mode period as well as the sleep-mode period together. We also consider the power consumption for switching the mode. Based on the analysis, we consider two performance measures – the power consumption per unit time per effective arrival and the average packet response delay. By considering these two performance measures together, we obtain the optimal sleep-mode operation that satisfies a given delay requirement and minimizes the power consumption per unit time per effective arrival.

References

[1] IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE Standard 802.16, 2004.

[2] C. Eklund, R. B. Marks, K.L. Stanwood, and S. Wang, "IEEE standard 802.16: A technical overview of the wirelessMAN air interface for broadband wireless access", IEEE Communications Magazine, vol.40, no.6, pp.98-107, June 2002.

[3] Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, IEEE Standard 802.16e, 2006.

[4] Dinh Thi Thuy Nga, Min-Gon Kim, and Minho Kang, "Delay-guaranteed Energy Saving Algorithm for the Delay-sensitive Applications in IEEE 802.16e Systems", IEEE Trans. Consumer Electron., vol.53, no.4, pp.1339-1347, Nov. 2007.

[5] Yang Xiao, "Energy saving mechanism in the IEEE 802.16e wireless MAN", IEEE Communications Letters, vol.9, no.7, pp.595-597, July 2005.

[6] Jun-Bae Seo, Seung-Que Lee, Nam-Hoon Park, Hyong-Woo Lee, and Choong-Ho Cho, "Performance analysis of sleep mode operation in IEEE802.16e", Vehicular Technology Conference, vol.2, no.26-29, pp.1169-1173, Sept. 2004.

[7] Jaehyuk Jang, Kwanghun Han, and Sunghyun Choi, “Adaptive Power Saving Strategies for IEEE 802.16e Mobile Broadband Wireless Access”, Asia-Pacific Conference on Communications, pp.1-5, Aug. 2006.

[8] Min-Gon KIM, JungYul CHOI, Bokrae JUNG, and Minho KANG, “Adaptive Power Management Mechanism Considering Remaining Energy in IEEE 802.16e”, IEICE Trans. Commun., vol.E90-B, no.9, pp.2621-2624, Sep. 2007.

[9] Min-Gon Kim, Minho Kang, and Jung Yul Choi, “Performance Evaluation of the Sleep Mode Operation in the IEEE 802.16e MAC”, International Conference on Advanced Communication Technology, vol.1, pp.602-605, Feb. 2007.

[10] Fangmin Xu, Wei Zhong, and Zheng Zhou, “A Novel Adaptive Energy Saving Mode in IEEE 802.16E System”, Military Communications conference, pp.1-6, Oct. 2006.

[11] Guojun Dong, Chengjun Zheng, Hongxia Zhang, and Jufern Dai, “Power saving class I sleep mode in IEEE 802.16e system”, Advanced Communication Technology, vol.3, pp.1487-1491, Feb. 2007.

[12] Hideaki Takagi, Queueing Analysis: A Foundation of Performance Evaluation, Volume 1:Vacation and priority Systems, Amsterdam: North Holland, 1991.

[13] Tony T. Lee, “M/G/1/N queue with vacation time and exhaustive service discipline”, Operation Research, vol.32, no.4, pp.774-784, 1984.

[14] Hideaki Takagi, Queueing Analysis: A Foundation of Performance Evaluation, Volume 2:Finite systems, Amsterdam: North Holland, 1993.

[15] Leonard Kleinrock, Queueing systems, volume1: Theory. New York: John Wileys & Sons, 1975.

[16] Louis W. Miller, “A note on the busy period of an M/G/1 finite queue”, Operation Research, vol.23, no.6, pp.1179-1182, Nov. 1975.

[17] Yang Xiao, Haizhon Li, Yi Pan, Kui Wu, and Jie Li, “On optimizing energy consumption for mobile handsets”, IEEE Transactions on Vehicular Technology, vol.53, no.6, pp.1927-1941, Nov. 2004.

[18] Ying-Wen Bai, and Ching-Ho Lai, “A bitmap scaling and rotation design for SH1 low power CPU”, Proceedings of the 2nd ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems MSWiM, pp.101-106, 1999.

[19] Ching-Long Su, Chi-Ying Tsui, and Alvin M Despain, “Saving power in the control path of embedded processors”, IEEE Design and Test of Computers, vol.11, no.4, pp.24-31, 1994.

[20] J. Banks, J. S. Carson II, B. L. Nelson, and D. M. Nicol Discrete-Event System Simulation, 3rd Edition, Prentice Hall, 2001.

[21] Yunju Park, and Gang Uk Hwang, “Performance Modelling and Analysis of the Sleep-Mode in IEEE802.16e WMAN”, IEEE 65th Vehicular Technology Conference, pp.2801-2806, April, 2007.

[22] Andreas Frey, and Yoshitaka Takahashi, “Explicit Solutions for the M/GI/1/N Finite Capacity Queues With and Without Vacation Time”, Proc. 15th International Teletraffic Congress, pp.507-516, June 1997.

[23] Hideaki Takagi, “Analysis of a Finite Capacity M/G/1 Queue with Resume Level”, Performance Evaluation, vol.5, no.3, pp.197-203, 1985.

Appendices

A.1. Derivation of $\mathbf{E}[V]$

$$
\begin{aligned}
\mathbf{E}[V] &= \sum_{j=0}^{\infty} S_j \mathbf{Pr}[V = S_j] \\
&= S_0(1 - e^{-\lambda S_0}) + \sum_{j=1}^{\infty} S_j e^{-\lambda S_{j-1}}(1 - e^{-\lambda v_j}) \\
&= S_0(1 - e^{-\lambda S_0}) + \sum_{j=1}^{\infty} (S_j e^{-\lambda S_{j-1}} - S_j e^{-\lambda S_j}) \\
&= S_0 + \sum_{j=1}^{\infty} e^{-\lambda S_{j-1}}(S_j - S_{j-1}) \\
&= S_0 + \sum_{j=1}^{M} e^{-\lambda S_{j-1}}(S_j - S_{j-1}) \\
&\quad + \sum_{j=M+1}^{\infty} e^{-\lambda S_{j-1}}(S_j - S_{j-1}) \\
&= S_0 + \sum_{j=1}^{M} e^{-\lambda S_{j-1}} v_j + \sum_{j=M+1}^{\infty} e^{-\lambda S_{j-1}} v_M \\
&= S_0 + \sum_{j=1}^{M} e^{-\lambda S_{j-1}} v_j \\
&\quad + e^{-\lambda S_M} v_M \left\{1 + e^{-\lambda v_M} + (e^{-\lambda v_M})^2 + \cdots\right\} \\
&= S_0 + \sum_{j=1}^{M} e^{-\lambda S_{j-1}} v_j + \frac{v_M}{1 - e^{-\lambda v_M}} e^{-\lambda S_M}.
\end{aligned}
$$

A.2. Derivation of $\mathbf{E}[N_I]$

$$\begin{aligned}
\mathbf{E}[N_I] &= \sum_{j=1}^{\infty} j\mathbf{Pr}[V = S_{j-1}] \\
&= (1 - e^{-\lambda S_0}) + \sum_{j=2}^{\infty} j e^{-\lambda S_{j-2}}(1 - e^{-\lambda v_{j-1}}) \\
&= (1 - e^{-\lambda S_0}) + \sum_{j=2}^{\infty}(j e^{-\lambda S_{j-2}} - j e^{-\lambda S_{j-1}}) \\
&= 1 + \sum_{j=0}^{\infty} e^{-\lambda S_j} \\
&= 1 + \sum_{j=0}^{M} e^{-\lambda S_j} + \sum_{j=M+1}^{\infty} e^{-\lambda S_j} \\
&= 1 + \sum_{j=0}^{M} e^{-\lambda S_j} \\
&\quad + e^{-\lambda S_M}\left\{e^{-\lambda v_M} + (e^{-\lambda v_M})^2 + \cdots\right\} \\
&= 1 + \sum_{j=0}^{M} e^{-\lambda S_j} + \frac{e^{-\lambda v_M}}{1 - e^{-\lambda v_M}} e^{-\lambda S_M}
\end{aligned}$$

A.3. Proof of Theorem 1

For $1 \le i \le K - 1$,

$$\begin{aligned}
\mathbf{Pr}[N_V = i] &= \sum_{n=0}^{\infty} \mathbf{Pr}[N_V = i | N_I = n]\mathbf{Pr}[N_I = n] \\
&= \mathbf{Pr}[N_V = i | N_I = 0]\mathbf{Pr}[N_I = 0] \\
&\quad + \sum_{n=1}^{M} \mathbf{Pr}[N_V = i | N_I = n]\mathbf{Pr}[N_I = n] \\
&\quad + \sum_{n=M+1}^{\infty} \mathbf{Pr}[N_V = i | N_I = n]\mathbf{Pr}[N_I = n] \\
&= \frac{\frac{(\lambda v_0)^i}{i!} e^{-\lambda v_0}}{1 - e^{-\lambda v_0}}(1 - e^{-\lambda v_0}) \\
&\quad + \sum_{n=1}^{M} \frac{\frac{(\lambda v_n)^i}{i!} e^{-\lambda v_n}}{1 - e^{-\lambda v_n}} e^{-\lambda S_{n-1}}(1 - e^{-\lambda v_n}) \\
&\quad + \frac{\frac{(\lambda v_M)^i}{i!} e^{-\lambda v_M}}{1 - e^{-\lambda v_M}} \sum_{k=0}^{\infty} e^{-\lambda(S_M + k v_M)}(1 - e^{-\lambda v_M}) \\
&= \sum_{n=0}^{M} \frac{(\lambda v_n)^i e^{-\lambda S_n}}{i!} + \frac{\frac{(\lambda v_M)^i}{i!} e^{-\lambda v_M}}{1 - e^{-\lambda v_M}} e^{-\lambda S_M}.
\end{aligned}$$

We have $\mathbf{Pr}[N_V = K] = 1 - \sum_{i=1}^{K-1} \mathbf{Pr}[N_V = i]$.

A.4. Proof of Theorem 2

For $1 \le j \le K - 1$,

$$\begin{aligned}
\mathbf{Pr}[N_S = j] &= \sum_{i=1}^{K-2} \mathbf{Pr}[N_S = j | N_V = i]\mathbf{Pr}[N_V = i] \\
&= \sum_{i=1}^{j} \mathbf{Pr}[A_S = j - i]\mathbf{Pr}[N_V = i]. \\
&= \sum_{i=1}^{j} \frac{(\lambda S)^{j-i} e^{-\lambda S}}{(j-i)!} \times \\
&\quad \left\{\sum_{n=0}^{M} \frac{(\lambda v_n)^i e^{-\lambda S_n}}{i!} + \frac{\frac{(\lambda v_M)^i}{i!} e^{-\lambda v_M}}{1 - e^{-\lambda v_M}} e^{-\lambda S_M}\right\}.
\end{aligned}$$

Similarly as in the proof of Theorem 1, we get $\mathbf{Pr}[N_S = K]$.

A.5. Proof of Theorem 4

$$\begin{aligned}
\mathbf{E}[B] &= \sum_{j=1}^{K} \mathbf{E}[B | N_S = j]\mathbf{Pr}[N_S = j] \\
&= \sum_{j=1}^{K-1} \mathbf{E}[B | N_S = j]\mathbf{Pr}[N_S = j] \\
&\quad + \mathbf{E}[B | N_S = K]\mathbf{Pr}[N_S = K] \\
&= \sum_{j=1}^{K-1} \mathbf{E}[B(j)]\mathbf{Pr}[N_S = j] \\
&\quad + (\mathbf{E}[X] + \mathbf{E}[B(K-1)])\,\mathbf{Pr}[N_S = K] \\
&= \sum_{j=1}^{K-1} \sum_{i=K-j}^{K-1} \mathbf{E}[B_i]\mathbf{Pr}[N_S = j] \\
&\quad + \left(\mathbf{E}[X] + \sum_{i=1}^{K-1} \mathbf{E}[B_i]\right)\mathbf{Pr}[N_S = K]
\end{aligned}$$

A.6. Proof of Theorem 5

Let π_j^d be the steady state probability that j packets are left in the system immediately after service completion $(0 \le j \le K - 1)$. Let L_n be the number of packets left behind in the system immediately after the n-th Markov point $(n = 1, 2, ...)$. Then the π_j^d is represented as follows:

$$\pi_j^d = \lim_{n \to \infty} \mathbf{Pr}[L_n = j], \qquad 0 \le j \le K - 1.$$

Let p_{ij} and a_k be the one step transition probability in the Markov chain and the probability that k packets arrive during a service time, respectively. Then p_{ij} and

a_k are represented as follows:

$$p_{i,j} \triangleq \mathbf{Pr}[L_{n+1} = j | L_n = i]$$
$$a_k \triangleq \int_0^\infty \frac{(\lambda x)^k}{k!} e^{-\lambda x} dF(x) \qquad k = 0, 1, ...$$

Then the one step transition probability p_{ij} is derived as follows:

$$P_{0,j} = \sum_{k=1}^{j+1} a_{j-k+1} \mathbf{Pr}[N_S = k], \qquad 0 \le j \le K-2, \tag{12}$$

$$P_{0,K-1} = \sum_{k=1}^{K} \sum_{l=K-k}^{\infty} a_l \mathbf{Pr}[N_S = k], \qquad j = K-1, \tag{13}$$

$$P_{i,j} = a_{j-i+1}, \qquad 1 \le i \le K-1,\ i-1 \le j \le K-2, \tag{14}$$

$$P_{i,K-1} = \sum_{l=K-i}^{\infty} a_l, \qquad 1 \le i \le K-1,\ j = K-1, \tag{15}$$
$$P_{ij} = 0, \qquad \text{otherwise.}$$

Note here that N_S is the number of backlogged packets at the beginning of a busy period. Also, the balance equations for the steady state probabilities are given by

$$\begin{aligned}\pi_j^d &= \sum_{i=0}^{K-1} \pi_i^d p_{i,j}, \qquad 0 \le j \le K-1,\\ &= \begin{cases} \displaystyle\sum_{i=0}^{j+1} \pi_i^d p_{i,j}, & 0 \le j \le K-2,\\ \displaystyle\sum_{i=0}^{K-1} \pi_i^d p_{i,j}, & j = K-1, \end{cases}\end{aligned} \tag{16}$$

$$\sum_{j=0}^{K-1} \pi_j^d = 1. \tag{17}$$

Then, for $0 \le j \le K-2$, by substituting (12) and (14) into (16), we have

$$\pi_j^d = \pi_0^d \sum_{k=1}^{j+1} a_{j-k+1} \mathbf{Pr}[N_S = k] + \sum_{i=1}^{j+1} \pi_i^d a_{j-i+1}. \tag{18}$$

Similarly, for $j = K-1$, by substituting (13) and (15) into (16), we have

$$\begin{aligned}\pi_{K-1}^d &= \pi_0^d \sum_{k=1}^{K} \sum_{l=K-k}^{\infty} a_l \mathbf{Pr}[N_S = k]\\ &\quad + \sum_{i=1}^{K-1} \pi_i^d \sum_{l=K-i}^{\infty} a_l\end{aligned} \tag{19}$$

Note that equation (19) is redundant, and that equations (17) and (18) provide K independent equations with K unknowns $\{\pi_j^d |\ 0 \le j \le K-1\}$. An efficient algorithm for computing $\{\pi_j^d |\ 0 \le j \le K-1\}$ can be given in terms of

$$\pi'_j \triangleq \frac{\pi_j^d}{\pi_0^d} \qquad 0 \le j \le K-1. \tag{20}$$

This π'_j is called an *upper Hessenberg matrix* [12,22,23]. It is easy to see from equation (18) that $\{\pi'_j |\ 0 \le j \le K-1\}$ can be recursively computed as follows. For $0 \le j \le K-2$

$$\begin{aligned}\pi'_0 &= 1\\ \pi'_j &= \frac{\pi_j^d}{\pi_0^d}\\ &= \frac{\pi_0^d \sum_{k=1}^{j+1} a_{j-k+1} \mathbf{Pr}[N_S = k] + \sum_{i=1}^{j+1} \pi_i^d a_{j-i+1}}{\pi_0^d}\\ &= \sum_{k=1}^{j+1} a_{j-k+1} \mathbf{Pr}[N_S = k] + \sum_{i=1}^{j+1} \pi'_i a_{j-i+1}\\ &= \sum_{k=1}^{j+1} a_{j-k+1} \mathbf{Pr}[N_S = k] + \sum_{i=1}^{j} \pi'_i a_{j-i+1} + \pi'_{j+1} a_0.\end{aligned}$$

Thus, for $0 \le j \le K-2$,

$$\pi'_{j+1} = \frac{1}{a_0} \left\{ \pi'_j - \sum_{i=1}^{j} \pi'_i a_{j-i+1} - \sum_{k=1}^{j+1} a_{j-k+1} \mathbf{Pr}[N_S = k] \right\}. \tag{21}$$

From equations (17) and (20), we have

$$\pi_0^d = \frac{1}{\sum_{j=0}^{K-1} \pi'_j}. \tag{22}$$

Hence, using equation (20), (21) and (22), we can get $\{\pi_j^d \mid 0 \le j \le K-1\}$.

Yunju Park received the B.S. degree in Mathematics from Kyungpook National University in 2002 and the M.S. degree in Mathematics from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2004. She is currently working toward the Ph.D. degree in the Department of Mathematical Sciences at KAIST. Her research interests include IEEE 802.16e and the power saving and performance modelling of wireless networks.

Gang Uk Hwang received the B.Sc., M.Sc., and Ph. D. degrees in Mathematics (Applied Probability) from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1991, 1993 and 1997, respectively. From February 1997 to March 2000, he was with Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea. From March 2000 to February 2002, he was a visiting scholar at the Department of Computer Sciences and Electrical Engineering in University of Missouri - Kansas City. Since March 2002, he has been with the Department of Mathematical Sciences and Telecommunication Engineering Program at KAIST, where he is an Associate Professor. His research interests include teletraffic theory, performance evaluation of communication systems, quality of service provisioning for wired/wireless networks and cross-layer design for wireless networks.