

Energy Management Analysis and Enhancement in IEEE 802.16e WirelessMAN

Yan Zhang, *Member, IEEE*, Yang Xiao, *Senior Member, IEEE*, and Victor C. M. Leung, *Fellow, IEEE*

Abstract—One strategy for energy management is an indispensable component in the emerging IEEE 802.16e wireless metropolitan area networks (WirelessMAN) supporting mobility. An efficient energy saving mechanism is the basis for the guarantee of a long operating lifetime for a mobile subscriber station (MSS), because MSSs are normally powered by rechargeable batteries. In this paper, we will characterize and model the energy-saving scheme in the IEEE 802.16e WirelessMAN. A comprehensive analysis is performed to study the specified sleep mode with generalized traffic processes. The performance metrics are derived with respect to energy consumption and packet delay to evaluate the tradeoff in the energy management strategy. We then propose an enhanced scheme to offer favorable performance tradeoff by adaptively adjusting the sleep windows. The numerical result indicates that the new policy can significantly reduce power consumption. Simulation results are presented to validate the analytical model, which provides potential guidelines for the efficient management of limited energy.

Index Terms—Energy consumption, energy efficiency, IEEE 802.16e, medium access control (MAC), mobile worldwide interoperability for microwave access (WiMAX), packet delay, sleep mode, WiMAX, wireless metropolitan area networks (WirelessMAN).

I. INTRODUCTION

WITH THE support for mobile capability and seamless handoff in the latest specification, the IEEE 802.16 Working Group is attempting to develop high-speed, high-bandwidth, and high-capacity standards for both fixed and mobile broadband wireless access [1]–[6]. The mobile version of the IEEE 802.16e standard was approved in December 2005 and can achieve data rates of up to 15 Mb/s in 5 MHz channel bandwidth. The standard is particularly specified for the emerging wireless metropolitan area networks (WirelessMAN), which can provide high-speed multimedia services as an excellent alternative to replacing costly cable access networks such as fiber optic links, coaxial systems, and digital subscriber line links [2]. The IEEE 802.16 standard defines the physical (PHY) layer and Medium Access Control (MAC) layer. PHY

includes three different specifications: 1) WirelessMAN single-carrier (SC) modulation; 2) WirelessMAN 256-carrier orthogonal frequency division multiplexing (OFDM) modulation; and 3) WirelessMAN 2048-carrier OFDM modulation. The IEEE 802.16e MAC layer comprises three sublayers: 1) Service Specific Convergence Sublayer (SSCS); 2) MAC Common Part Sublayer (MAC CPS); and 3) Privacy Sublayer (PS). SSCS acts as the interface to higher layers and provides a mapping function from the diverse transport layer traffic to the flexible MAC. MAC CPS is the core component that includes access control, collision resolution, scheduling, and bandwidth requests. PS ensures secure connection establishment, access authentication, key exchange, and data privacy.

Given that mobile subscriber stations (MSSs) are usually powered by rechargeable batteries and given the promising mobility capability in IEEE 802.16e [3], it is important to develop an efficient mechanism to conserve energy in wireless systems and networks, to maximize the working lifetime of mobile stations, and to enhance network resilience [7]. To efficiently manage limited power, a sleep-mode operation is specified in the IEEE 802.16e MAC protocol [3]. An MSS goes into the sleep state if it has been idle for a period without data transmission/user activity. This period is normally standardized and serves as a timer in practical implementation. In the literature, the sleep strategy has been discussed in [8] and [10]. The work in [10] proposed an analytical model to investigate the energy consumption in IEEE 802.16e by considering the downlink traffic, i.e., from the base station (BS) to MSSs. The study in [8] investigated the energy consumption in IEEE 802.16e by considering both downlink and uplink (to be sent from an MSS to the BS) traffic. The study in [14] presented the performance of two different power-saving classes. It was observed that, in previous studies, specific Poisson assumptions were employed for the data packet in both downlink and uplink traffic for analytical simplicity. In other words, an exponential distribution function was assumed for both downlink and uplink packet interarrival times. However, it has been reported that the exponential distribution may be inappropriate for approximating the packet data traffic process [15], [16]. Motivated by this observation, we will perform an extensive analysis of the specified sleep mode with generalized traffic processes instead of memoryless Poisson process. Based on the analysis and simulation given as follows, it is shown that the scenario with generalized traffic processes requires significantly different analytical technique. It could substantially complicate the system dynamics and, more importantly, exhibit considerably different characteristics. This serves as the major difference between our study and the previous work. In addition, we present a new energy-saving policy. The result demonstrates that the

Manuscript received May 6, 2008; revised October 20, 2008 and November 11, 2008. First published February 2, 2009; current version published August 14, 2009. The review of this paper was coordinated by Prof. C. Lin.

Y. Zhang is with Simula Research Laboratory, 1325 Lysaker, Norway (e-mail: yanzhang@ieee.org).

Y. Xiao is with the Department of Computer Science, University of Alabama, Tuscaloosa, AL 35487-0286 USA (e-mail: yangxiao@ieee.org).

V. C. M. Leung is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: vleung@ece.ubc.ca).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TVT.2009.2014383

proposed scheme can significantly reduce energy consumption with acceptable delay.

To preserve universal approximation, general applicability, and ease of calculation, we employ the hyper-Erlang distribution as the uplink and downlink packet interarrival time. The hyper-Erlang distribution has been proven to arbitrarily closely approximate the distribution of any positive random variable and measured data [18]. This general approximation property has been applied to various scenarios. The work in [12] proposed the hyper-Erlang model for the cell residence time. Recently, in the General Packet Radio Service (GPRS) subsystem of the Universal Mobile Telecommunications System networks, the work in [13] modeled the serving GPRS support node (SGSN) residence time as a hyper-Erlang distribution to evaluate authentication signaling traffic. To our best knowledge, this paper is the first investigation that applies a universal approximation distribution function in energy management scheme analysis.

Our main contributions are summarized as follows:

- 1) Construct a system model, and develop an analytical technique under generalized traffic processes to study the performance tradeoff in the WirelessMAN energy saving mechanism with respect to the energy consumption and packet delay.
- 2) Evaluate performance tradeoffs in two distinct scenarios, including the cases when only downlink traffic is considered and when both downlink and uplink traffic is taken into account.
- 3) Propose a new energy saving policy that works by adaptively controlling the initial size of the sleep window.
- 4) Perform numerical examination to validate the analysis model via simulations and investigate the interaction between critical parameters and the performance metrics.

The rest of this paper is organized as follows. In Section II, we describe the standardized energy-saving mechanism in the IEEE 802.16e WirelessMAN. In Section III, the system model is developed to evaluate the tradeoff in the case of generalized downlink traffic process. In Section IV, the scenario that considers both downlink and uplink traffic processes is investigated. In Section V, an enhanced energy management scheme is proposed to significantly reduce energy consumption by adaptively adjusting the sleep window interval. Numerical results are given in Section VI, followed by concluding remarks in Section VII.

II. SLEEP MODE IN IEEE 802.16e

Fig. 1 shows the sleep-mode message sequence between the BS and MSS. The energy saving mechanism in the sleep mode is particularly designed for mobile terminals. Before entering the sleep mode, the MSS sends a request message MOB-SLP-REQ to the BS for permission to enter the sleep mode. Upon receiving the request, the BS replies with the response message MOB-SLP-RSP. This response message indicates the parameters initial-sleep window T_{\min} , final-sleep window T_{\max} , and listening window L . Upon receiving MOB-SLP-RSP, the MSS enters into the sleep mode. We now focus on the sleep mode mechanism. The duration of the first sleep interval T_1 is equal

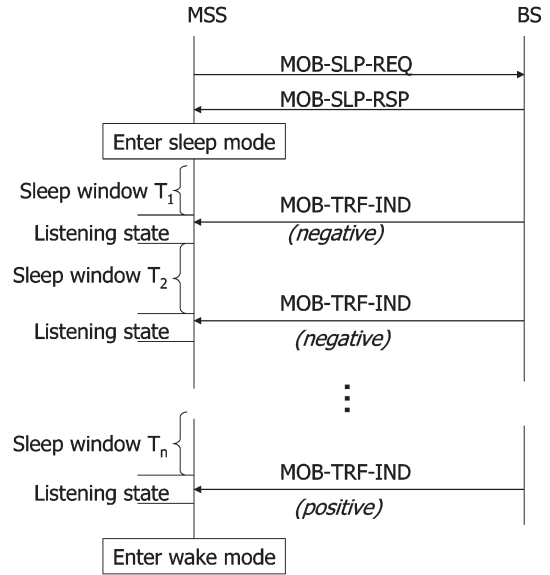


Fig. 1. Sleep-mode operation sequence between an MSS and the BS in IEEE 802.16e.

to the initial-sleep window T_{\min} . After the first sleep interval, the MSS transits into a listening state and listens to the traffic indication message MOB-TRF-IND, which is broadcast from the BS. This message indicates whether there has been traffic addressed to the MSS during its sleep interval T_1 . If MOB-TRF-IND indicates the message in the negative, the MSS continues in the sleep mode after the listening interval L . If MOB-TRF-IND positively indicates, the MSS will return to the wake mode. We term the sleep interval and its subsequent listening interval as a cycle.

If the MSS continues in the sleep mode, the next sleep window starts from the end of the previous listening window, and the new sleep interval will be twice the length of the preceding sleep interval. This process is repeated, as long as the sleep interval does not exceed the final-sleep window T_{\max} . When the MSS has reached T_{\max} , the length of the sleep interval will remain constant at T_{\max} . That is, the length of sleep window in the n th cycle is given by

$$T_n = \begin{cases} T_{\min}, & n = 1 \\ \min(2^{n-1}T_{\min}, T_{\max}), & n > 1. \end{cases} \quad (1)$$

For presentation purposes, we let $T_{\max} = T_{N_{\max}-1} = 2^{N_{\max}-1}T_1$. This equation also indicates that the maximum number of sleep windows before reaching the final-sleep window is N_{\max} . We define the set of all possible lengths of sleep window as $\mathcal{T} = \{T_1, T_2, \dots, T_{N_{\max}-1}\}$.

Fig. 2 shows the wake mode and sleep mode of an MSS. It may be shown that the MSS alternates between the wake mode and the sleep mode throughout its lifetime. When there are downlink packets to a sleeping MSS, the downlink packets are buffered in the BS, and the MSS exits the sleep mode in the next listening interval. When there are uplink packets to a sleeping MSS, the MSS will immediately terminate the sleep mode. It is noteworthy that other procedures are independent of the energy saving scheme, which is particularly defined for the sleep mode. For instance, the energy consumption during data

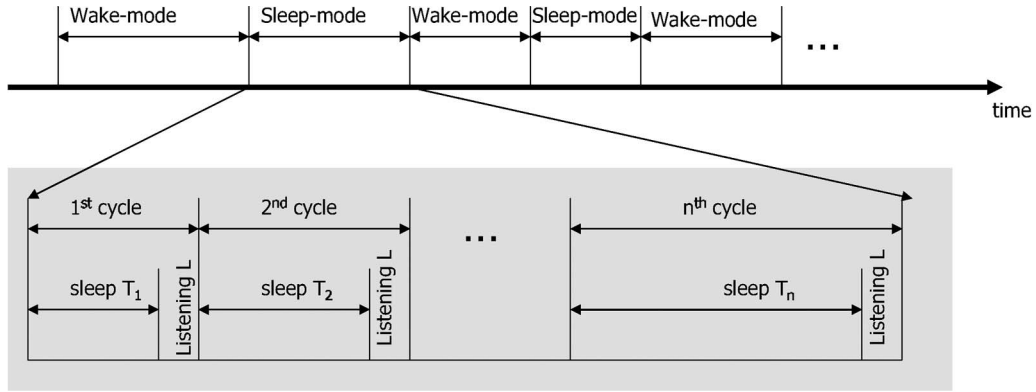


Fig. 2. Wake-mode and sleep-mode transition in the IEEE 802.16e.

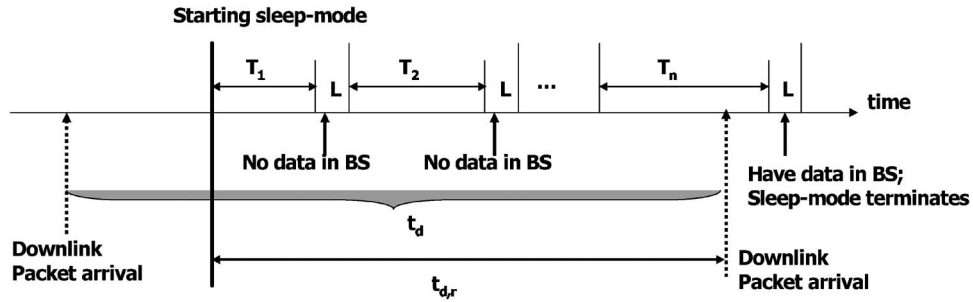


Fig. 3. System model when considering downlink traffic.

communications is largely determined by the amount of transmitted data and the specific policies during data transmission. Hence, the modeling, analysis, and simulation for the sleep mode in this paper are kept unchanged in case other procedures are considered.

III. ENERGY CONSUMPTION WITH DOWNLINK TRAFFIC

Fig. 3 shows the system model for analyzing the energy consumption and packet delay, considering only the downlink traffic. We denote the downlink packet interarrival time by t_d , the mean downlink packet interarrival time by $1/\lambda_d$, the probability density function (pdf) by $f_{t_d}(t)$, the cumulative distribution function (cdf) by $F_{t_d}(t)$, and the Laplace transform (LT) of pdf by $f_{t_d}^*(s)$. To preserve universal approximation, general applicability, and ease of calculation, we suppose that t_d follows the hyper-Erlang distribution with pdf [17], [18]

$$f_{t_d}(t) = \sum_{i=1}^{H_d} \alpha_{d,i} \frac{(m_{d,i} \lambda_{d,i})^{m_{d,i}} t^{m_{d,i}-1}}{(m_{d,i}-1)!} e^{-m_{d,i} \lambda_{d,i} t} \quad (2)$$

where $\sum_{i=1}^{H_d} \alpha_{d,i} = 1$, H_d and $m_{d,i}$ are positive integers, and $\alpha_{d,i}$ and $\lambda_{d,i}$ are positive real constants with $0 \leq \alpha_{d,i} \leq 1$, $\lambda_{d,i} \geq 0$. The parameters H_d , $\alpha_{d,i}$, and $m_{d,i}$ in (2) determine the shape and scale of the specific distribution. The average value of t_d is calculated as

$$\frac{1}{\lambda_d} = \sum_{i=1}^{H_d} \frac{\alpha_{d,i}}{\lambda_{d,i}}.$$

The LT of the hyper-Erlang distribution is given by

$$f_{t_d}^*(s) = \sum_{i=1}^{H_d} \alpha_{d,i} \left(\frac{m_{d,i} \lambda_{d,i}}{s + m_{d,i} \lambda_{d,i}} \right)^{m_{d,i}}.$$

Furthermore, for a real number α , integer m , and time variable t , we have the following formula:

$$\int_0^t \frac{\alpha x^{m-1}}{(m-1)!} e^{-\alpha x} dx = 1 - \sum_{j=0}^{m-1} \frac{(\alpha t)^j}{j!} e^{-\alpha t}.$$

Applying the aforementioned identity, we calculate the cdf of hyper-Erlang t_d as

$$F_{t_d}(t) = \int_0^t f_{t_d}(x) dx = 1 - \sum_{i=1}^{H_d} \alpha_{d,i} \sum_{j=0}^{m_{d,i}-1} \frac{(m_{d,i} \lambda_{d,i} t)^j}{j!} e^{-m_{d,i} \lambda_{d,i} t}. \quad (3)$$

We define the downlink packet residual interarrival time as the duration from an intermediate moment between two consecutive packets to the arrival time of the next packet. Let $t_{d,r}$ denote the downlink packet residual interarrival time with pdf $f_{t_{d,r}}(t)$, cdf $F_{t_{d,r}}(t)$, and LT of pdf $f_{t_{d,r}}^*(s)$. Referring to the Residual Life Theorem [17], we express the pdf of $t_{d,r}$ as

$$f_{t_{d,r}}(t) = \lambda_d \int_t^\infty f_{t_d}(x) dx = \lambda_d [1 - F_{t_d}(t)] \quad (4)$$

and the LT of pdf as

$$f_{t_{d,r}}^*(s) = \frac{\lambda_d [1 - f_{t_d}^*(s)]}{s}.$$

Substituting (2) into (4), we obtain the pdf of $t_{d,r}$ in the case of hyper-Erlang t_d . We have

$$f_{t_{d,r}}(t) = \lambda_d \sum_{i=1}^{H_d} \alpha_{d,i} \sum_{j=0}^{m_{d,i}-1} \frac{(m_{d,i}\lambda_{d,i})^j}{j!} e^{-m_{d,i}\lambda_{d,i}t}. \quad (5)$$

In addition, the cdf of $t_{d,r}$ is calculated as

$$F_{t_{d,r}}(t) = 1 - \lambda_d \sum_{i=1}^{H_d} \frac{\alpha_{d,i}}{m_{d,i}\lambda_{d,i}} \times \sum_{j=0}^{m_{d,i}-1} \sum_{k=0}^j \frac{(m_{d,i}\lambda_{d,i}t)^k}{k!} e^{-m_{d,i}\lambda_{d,i}t}. \quad (6)$$

A. Energy Consumption

Let N denote a random variable that represents the number of cycles that an MSS experiences when the sleep mode terminates. In Fig. 3, we suppose that the sleep mode is terminated during the n th cycle. This implies that 1) there is no downlink packet during the first, second, and up to the $(n-1)$ st cycles, and 2) there is a downlink packet during the n th sleep window. Under the condition that the MSS terminates the sleep mode during the n th cycle, we let E_n denote the energy consumption, provided that the MSS terminates the sleep mode in the n th cycle. Let \bar{E}_n denote the average value of E_n , where \bar{X} denotes the average value of the random variable X . Let \mathcal{E} represent the energy consumption during the sleep mode. Hence, the average consumed energy is expressed as

$$\bar{\mathcal{E}} = \sum_{n=1}^{\infty} \Pr(N = n) \bar{E}_n. \quad (7)$$

In the aforementioned equation, the value of N is highly dependent on sleeping-window parameters and the downlink packet traffic characteristics. Based on a theoretical analysis perspective, it has taken into account all possibilities for the variable N .

Let E_S and E_L denote the consumed energy units per unit time in a sleep interval and a listening interval, respectively. Then, during the n th cycle, the average energy consumption is the summation of consumed energy in the sleep window and in the listening window when the MSS terminates the sleep mode. We have

$$\bar{E}_n = \left[\sum_{j=1}^n T_j \right] E_S + (n-1)LE_L. \quad (8)$$

Now, we will develop the probability distribution for the random variable N . For presentation purposes, we let W_n

denote the summation of the first, second, and up to the n th sleep window and listening window, i.e.,

$$W_n = \sum_{j=1}^n (T_j + L) \quad n \geq 1$$

with the initial condition $W_0 = 0$.

The situation $N = 1$ implies that there are downlink packet arrivals during T_1 or, equivalently, the interval $t_{d,r}$ is shorter than the first sleep window. Hence, we express

$$\begin{aligned} \Pr(N = 1) &= \Pr(t_{d,r} \leq T_1) \\ &= \int_0^{T_1} f_{t_{d,r}}(t) dt \\ &= F_{t_{d,r}}(T_1) \\ &= 1 - \lambda_d \sum_{i=1}^{H_d} \frac{\alpha_{d,i}}{m_{d,i}\lambda_{d,i}} \\ &\quad \times \sum_{j=0}^{m_{d,i}-1} \sum_{k=0}^j \frac{(m_{d,i}\lambda_{d,i}T_1)^k}{k!} e^{-m_{d,i}\lambda_{d,i}T_1}. \end{aligned} \quad (9)$$

For the case $N = n (n \geq 2)$, no downlink packets arrive during the first, second, and up to the $(n-1)$ st cycle, but there are downlink packets during T_n . This scenario is equivalent to the interval $t_{d,r}$ being shorter than $W_{n-1} + T_n$ but longer than W_{n-1} . Hence, we have

$$\begin{aligned} \Pr(N = n) &= \Pr(W_{n-1} - L < t_{d,r} \leq W_n - L) \\ &= \int_{W_{n-1}-L}^{W_n-L} f_{t_{d,r}}(t) dt \\ &= F_{t_{d,r}}(W_n - L) - F_{t_{d,r}}(W_{n-1} - L) \\ &= \lambda_d \sum_{i=1}^{H_d} \frac{\alpha_{d,i}}{m_{d,i}\lambda_{d,i}} \sum_{j=0}^{m_{d,i}-1} \sum_{k=0}^j \frac{(m_{d,i}\lambda_{d,i})^k}{k!} e^{m_{d,i}\lambda_{d,i}L} \\ &\quad \times [(W_{n-1} - L)^k e^{-m_{d,i}\lambda_{d,i}W_{n-1}} \\ &\quad - (W_n - L)^k e^{-m_{d,i}\lambda_{d,i}W_n}]. \end{aligned} \quad (10)$$

Substituting (8), (9), and (10) into (7), we can obtain the energy consumption. The simple form of summation of hyper-Erlang leads to the summation expression of performance metrics and will reduce the computation complexity.

B. Packet Delay

A longer sleep mode enables the MSS to save more energy and allows a longer MSS serving time. However, the longer the sleep mode, the longer the time for the MSS to respond to the downlink packet arrivals. Fig. 3 shows that, for a downlink packet arrival to a sleeping MSS, the packet has to wait for the MSS's response until the MSS completes the ongoing sleep window, transits to a listening state, and finally enters the wake

mode. Hence, the MSS cannot instantaneously handle downlink packet traffic. Such introduced delay for a downlink packet can reflect the adverse contribution of the sleep mode. In this paper, if not specified, we use the term packet delay or simply delay to represent the MSS's delayed response time for a downlink packet.

Let \mathcal{D} denote the delay of a downlink packet and let $\bar{\mathcal{D}}$ denote the average packet delay. If a downlink packet arrives at the MSS's first sleep window, it has to wait in the BS until the first listening window. Alternatively, when the residual downlink interarrival time $t_{d,r}$ is shorter than T_1 , the delay is equal to the difference between $t_{d,r}$ and T_1 . In such a situation, the delay \mathcal{D} is expressed as

$$\mathcal{D} = T_1 - t_{d,r}; \quad \text{if } t_{d,r} \leq T_1. \quad (11)$$

Define an alternative random variable v_1 for $t_{d,r}$ as

$$v_1 = \begin{cases} t_{d,r}, & t_{d,r} \leq T_1 \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

Then, the average value of v_1 is given by

$$\begin{aligned} \bar{v}_1 &= \frac{\int_0^{T_1} x f_{t_{d,r}}(x) dx}{\int_0^{T_1} f_{t_{d,r}}(x) dx} \\ &= \frac{A}{F_{t_{d,r}}(T_1)} \end{aligned} \quad (13)$$

where

$$\begin{aligned} A &= \int_0^{T_1} x f_{t_{d,r}}(x) dx \\ &= \int_0^{T_1} x \lambda_d \sum_{i=1}^{H_d} \alpha_{d,i} \sum_{j=0}^{m_{d,i}-1} \frac{(m_{d,i} \lambda_{d,i} x)^j}{j!} e^{-m_{d,i} \lambda_{d,i} x} dx \\ &= j \lambda_d \sum_{i=1}^{H_d} \alpha_{d,i} \sum_{j=0}^{m_{d,i}-1} \frac{(m_{d,i} \lambda_{d,i})^j}{j!} \int_0^{T_1} x^{j+1} e^{-m_{d,i} \lambda_{d,i} x} dx. \end{aligned} \quad (14)$$

To illustrate (14), we define

$$\Theta_n(a, b; \beta) = \int_a^b x^n e^{-\beta x} dx.$$

The recursive algorithm for calculating $\Theta_n(a, b; \beta)$ is derived and presented in Appendix A. Then, the result for (14) is expressed as

$$A = \lambda_d \sum_{i=1}^{H_d} \alpha_{d,i} \sum_{j=0}^{m_{d,i}-1} \frac{(m_{d,i} \lambda_{d,i})^j}{j!} \Theta_{j+1}(0, T_1; m_{d,i} \lambda_{d,i}). \quad (15)$$

If a downlink packet arrives at the MSS during the n th sleep window, the packet has to wait in the BS until the next listening

window. Equally, when the residual downlink interarrival time $t_{d,r}$ is shorter than $W_n - L$ but longer than $W_{n-1} - L$, the delay is equal to the difference between $t_{d,r}$ and $W_n - L$. In such a situation, the delay \mathcal{D} is expressed as

$$\mathcal{D} = W_n - L - t_{d,r} \text{ if } W_{n-1} - L < t_{d,r} \leq W_n - L. \quad (16)$$

For this situation, we define an alternative random variable v_n for the variable $t_{d,r}$ in the range as

$$v_n = \begin{cases} t_{d,r}, & W_{n-1} - L < t_{d,r} \leq W_n - L \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

Then, the average value of v_n is given by

$$\begin{aligned} \bar{v}_n &= \frac{\int_{W_{n-1}-L}^{W_n-L} x f_{t_{d,r}}(x) dx}{\int_{W_{n-1}-L}^{W_n-L} f_{t_{d,r}}(x) dx} \\ &= \frac{B_n}{F_{t_{d,r}}(W_n - L) - F_{t_{d,r}}(W_{n-1} - L)} \end{aligned} \quad (18)$$

where B_n is given in Appendix B.

Considering all possible N , (11), and (16), we express the packet delay as

$$\mathcal{D} = (T_1 - v_1) \Pr(N = 1) + \sum_{n=2}^{\infty} (W_n - L - v_n) \Pr(N = n).$$

Taking the expectation operation on the two sides of the aforementioned equation, we obtain the average delay as

$$\begin{aligned} \bar{\mathcal{D}} &= (T_1 - \bar{v}_1) \Pr(N = 1) + \sum_{n=2}^{\infty} (W_n - L - \bar{v}_n) \Pr(N = n) \\ &= [T_1 \Pr(N = 1) - A] \\ &\quad + \sum_{n=2}^{\infty} [(W_n - L) \Pr(N = n) - B_n] \end{aligned} \quad (19)$$

where A and B_n are given in (15) and (49), respectively.

IV. ENERGY CONSUMPTION WITH DOWNLINK AND UPLINK TRAFFIC

We now investigate the tradeoff with respect to energy consumption and packet delay, considering both downlink and uplink traffic processes. The motivation is mainly due to the immediate termination of the MSS sleep mode when there is uplink traffic or there are external operations to be performed. As a consequence, the time at which the sleep mode is terminated varies according to the type of traffic. It is, thus, necessary to differentiate the type of traffic when determining the duration of sleep intervals, in addition to considering energy consumption. As a result, to comprehensively evaluate the mechanism for energy management, we will also study the sleep mode performance, considering both downlink and uplink traffic. In this paper, we will not present the case when only uplink traffic is considered. The main reason is due to the immediate termination of a sleep mode upon an uplink packet arrival. Thus, the energy consumption in this situation is straightforwardly the summation of the energy in a number of sleeping windows before an uplink packet arrival.

We denote the uplink packet interarrival time by t_u , the mean interarrival time by $1/\lambda_u$, the pdf by $f_{t_u}(t)$, the cdf by $F_{t_u}(t)$, and the LT of the pdf by $f_{t_u}^*(s)$. Let $t_{u,r}$ denote the uplink packet residual interarrival time with pdf $f_{t_{u,r}}(t)$, cdf $F_{t_{u,r}}(t)$, and LT of the pdf $f_{t_{u,r}}^*(s)$. Similarly, referring to the Residual Life Theorem [17], we express the $t_{u,r}$ pdf as

$$f_{t_{u,r}}(t) = \lambda_u \int_t^{\infty} f_{t_u}(x) dx$$

and the LT of pdf as

$$f_{t_{u,r}}^*(s) = \frac{\lambda_u [1 - f_{t_u}^*(s)]}{s}.$$

For the hyper-Erlang distributed t_u with pdf [17], [18], we have

$$f_{t_u}(t) = \sum_{i=1}^{H_u} \alpha_{u,i} \frac{(m_{u,i} \lambda_{u,i})^{m_{u,i}} t^{m_{u,i}-1}}{(m_{u,i}-1)!} e^{-m_{u,i} \lambda_{u,i} t} \quad (20)$$

where $\sum_{i=1}^{H_u} \alpha_{u,i} = 1$, H_u and $m_{u,i}$ are positive integers, and $\alpha_{u,i}$ and $\lambda_{u,i}$ are positive real constants, with $0 \leq \alpha_{u,i} \leq 1$, $\lambda_{u,i} \geq 0$. The parameters H_u , $\alpha_{u,i}$, and $m_{u,i}$ in (2) determine the shape and scale of the specific distribution. The average value of t_u is calculated as

$$\frac{1}{\lambda_u} = \sum_{i=1}^{H_u} \frac{\alpha_{u,i}}{\lambda_{u,i}}.$$

Similarly, we obtain the pdf of $t_{u,r}$ in the case of hyper-Erlang t_u as

$$f_{t_{u,r}}(t) = \lambda_u \sum_{i=1}^{H_u} \alpha_{u,i} \sum_{j=0}^{m_{u,i}-1} \frac{(m_{u,i} \lambda_{u,i} t)^j}{j!} e^{-m_{u,i} \lambda_{u,i} t}$$

and the cdf of $t_{u,r}$ is calculated as

$$F_{t_{u,r}}(t) = 1 - \lambda_u \sum_{i=1}^{H_u} \frac{\alpha_{u,i}}{m_{u,i} \lambda_{u,i}} \times \sum_{j=0}^{m_{u,i}-1} \sum_{k=0}^j \frac{(m_{u,i} \lambda_{u,i} t)^k}{k!} e^{-m_{u,i} \lambda_{u,i} t}. \quad (21)$$

Let λ denote the summation of total uplink and downlink packet arrival rates, i.e., $\lambda = \lambda_u + \lambda_d$. For hyper-Erlang distributed t_u and t_d , we have

$$\lambda = \left(\sum_{i=1}^{H_u} \frac{\alpha_{u,i}}{\lambda_{u,i}} + \sum_{i=1}^{H_d} \frac{\alpha_{d,i}}{\lambda_{d,i}} \right)^{-1}.$$

A. Energy Consumption

Under the condition that the MSS terminates the sleep mode during the n th cycle, we distinguish three cases, as shown in Fig. 4. Accordingly, we let $\overline{E}_n^{(k)}$ denote the consumed energy

during the case (k) ($k \in \{a, b, c\}$), provided that the MSS terminates the sleep mode in the n th cycle. In addition, let $\phi_n^{(k)}$ be the probability that case (k) ($k \in \{a, b, c\}$) will occur under the condition that the MSS terminates the sleep mode in the n th cycle. Taking these three possibilities into account, the average consumed energy $\overline{\mathcal{E}}$ during a sleep mode is, hence, expressed as

$$\overline{\mathcal{E}} = \sum_{n=1}^{\infty} \left[\overline{E}_n^{(a)} \phi_n^{(a)} + \overline{E}_n^{(b)} \phi_n^{(b)} + \overline{E}_n^{(c)} \phi_n^{(c)} \right]. \quad (22)$$

1) *Case 1—Sleep mode is terminated during the n th sleep window due to the arrival of an uplink packet:* In this case, as shown in Fig. 4(a), there is an uplink packet that results in the sleep mode being terminated during the n th sleep window. This situation indicates that there are no downlink packets during the j^{th} ($j = 1, 2, \dots, n-1$) sleep interval. In addition, an uplink packet will arrive during the n th sleep window, which implies that $t_{u,r}$ should be greater than W_{n-1} and concurrently less than $(W_{n-1} + T_n)$. In such a case, we define an alternative random variable y_n that satisfies

$$y_n = \begin{cases} t_{u,r}, & W_{n-1} < t_{u,r} < W_{n-1} + T_n \\ 0, & \text{otherwise.} \end{cases} \quad (23)$$

Then, the average value of y_n is given by

$$\begin{aligned} \overline{y}_n &= \frac{\int_{W_{n-1}}^{W_{n-1}+T_n} x f_{t_{u,r}}(x) dx}{\int_{W_{n-1}}^{W_{n-1}+T_n} f_{t_{u,r}}(x) dx} \\ &= \frac{C_n^{(a)}}{R_n^{(a)}} \end{aligned} \quad (24)$$

where $C_n^{(a)}$ and $R_n^{(a)}$ are given in Appendix C.

In this situation, the consumed energy is the summation of energy consumption in the sleep window and in the listening window during the first, second, and up to the $(n-1)$ st cycles, plus the energy consumption in the sleep duration in the n th cycle right before sleep termination. Consequently, the energy for Case 1 is expressed as

$$\begin{aligned} E_n^{(a)} &= \sum_{j=1}^{n-1} T_j E_S + (n-1) L E_L \\ &\quad + \left[y_n - \sum_{j=1}^{n-1} T_j - (n-1)L \right] E_S \\ &= [y_n - (n-1)L] E_S + (n-1) L E_L. \end{aligned} \quad (25)$$

Taking the expectation operation, the average energy consumption is given by

$$\overline{E}_n^{(a)} = [\overline{y}_n - (n-1)L] E_S + (n-1) L E_L \quad (26)$$

where \overline{y}_n is given in (24).

Now, we calculate the probability for Case 1. If there is an uplink packet arrival during the first cycle, the sleep mode is

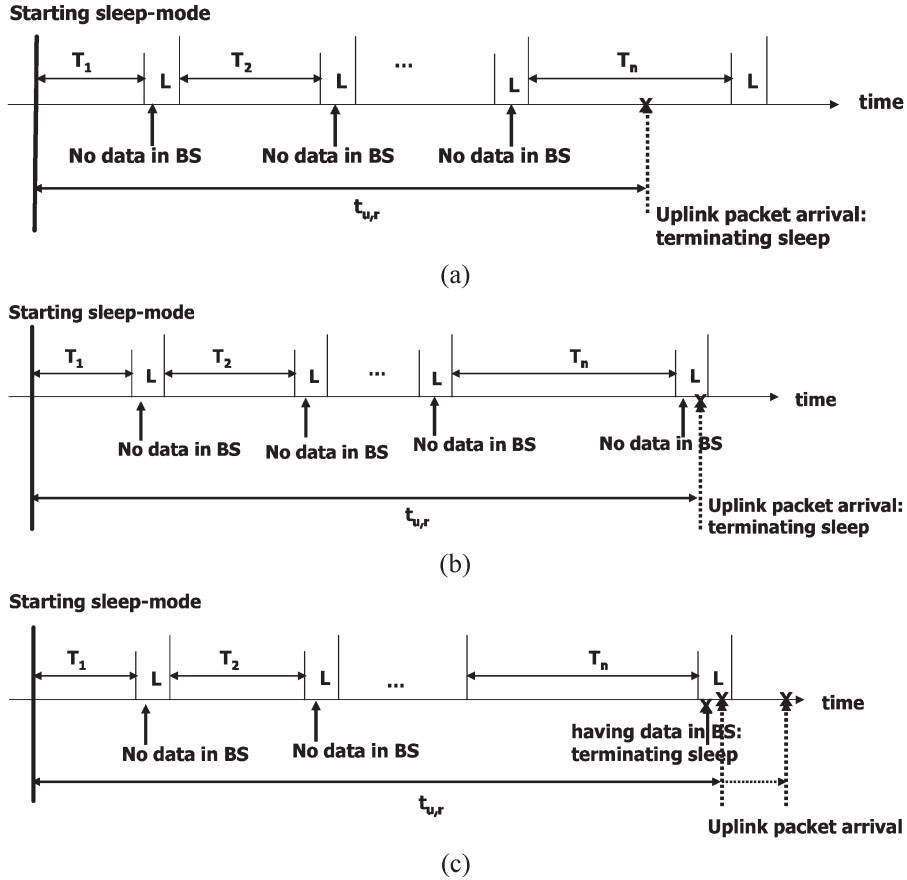


Fig. 4. Three different and exclusive cases for termination of the sleep mode. (a) Sleep mode is terminated during the n th sleep window due to the arrival of an uplink packet. (b) Sleep mode is terminated during the n th listening window due to the arrival of an uplink packet. (c) Sleep mode is terminated during the n th listening window due to the arrival of a downlink packet.

instantly terminated. This scenario indicates that t_{ur} is shorter than the first sleep window length. Hence, we have

$$\begin{aligned}
 \phi_1^{(a)} &= \Pr(t_{u,r} \leq T_1) = \int_0^{T_1} f_{t_{u,r}}(x) dx \\
 &= F_{t_{u,r}}(T_1) \\
 &= 1 - \lambda_u \sum_{i=1}^{H_u} \frac{\alpha_{u,i}}{m_{u,i} \lambda_{u,i}} \\
 &\quad \times \sum_{j=0}^{m_{u,i}-1} \sum_{k=0}^j \frac{(m_{u,i} \lambda_{u,i} T_1)^k}{k!} e^{-m_{u,i} \lambda_{u,i} T_1}. \quad (27)
 \end{aligned}$$

For the situation $N = n (n \geq 2)$, no downlink packets arrive during the first, second, and up to the $(n-1)$ st cycle. Moreover, there are uplink packets arrival in the n th sleep interval. Hence, the probability that Case 1 will occur is given by

$$\begin{aligned}
 \phi_n^{(a)} &= \Pr(W_{n-1} < t_{u,r} \leq W_{n-1} + T_n; t_{d,r} > W_{n-1} - L) \\
 &= \Pr(W_{n-1} < t_{u,r} \leq W_{n-1} + T_n) \Pr(t_{d,r} > W_{n-1} - L) \\
 &= \int_{W_{n-1}}^{W_{n-1}+T_n} f_{t_{u,r}}(x) dx \int_{W_{n-1}-L}^{\infty} f_{t_{d,r}}(y) dy \\
 &= [F_{t_{u,r}}(W_{n-1} + T_n) - F_{t_{u,r}}(W_{n-1})] \\
 &\quad \times [1 - F_{t_{d,r}}(W_{n-1} - L)]. \quad (28)
 \end{aligned}$$

2) *Case 2—Sleep mode is terminated during the n th listening window due to the arrival of an uplink packet*: In this case, as shown in Fig. 4(b), there is an uplink packet that results in the sleep mode being terminated during the n th listening interval. This indicates that there are no packets during the j^{th} ($j = 1, 2, \dots, n$) sleep interval. In addition, $t_{u,r}$ should be greater than $W_{n-1} + T_n (= W_n - L)$ and also less than W_n . In such a case, we define an alternative random variable z_n that satisfies

$$z_n = \begin{cases} t_{u,r}, & W_n - L < t_{u,r} < W_n \\ 0, & \text{otherwise.} \end{cases} \quad (29)$$

Then, the average value of z_n is given by

$$\begin{aligned}
 \bar{z}_n &= \frac{\int_{W_n-L}^{W_n} x f_{t_{u,r}}(x) dx}{\int_{W_n-L}^{W_n} f_{t_{u,r}}(x) dx} \\
 &= \frac{C_n^{(b)}}{R_n^{(b)}}. \quad (30)
 \end{aligned}$$

Following a similar reasoning that leads to the expressions for $C_n^{(a)}$ and $R_n^{(a)}$, we obtain the results for items $C_n^{(b)}$ and

$R_n^{(b)}$, i.e.,

$$C_n^{(b)} = \lambda_u \sum_{i=1}^{H_u} \alpha_{u,i} \sum_{j=0}^{m_{u,i}-1} \frac{(m_{u,i} \lambda_{u,i})^j}{j!} \times \Theta_{j+1}(W_n - L, W_n; m_{u,i} \lambda_{u,i}) \quad (31)$$

$$\begin{aligned} R_n^{(b)} &= F_{t_{u,r}}(W_n) - F_{t_{u,r}}(W_n - L) \\ &= \lambda_u \sum_{i=1}^{H_u} \frac{\alpha_{u,i}}{m_{u,i} \lambda_{u,i}} \sum_{j=0}^{m_{u,i}-1} \sum_{k=0}^j \frac{(m_{u,i} \lambda_{u,i})^k}{k!} \\ &\quad \times e^{-m_{u,i} \lambda_{u,i} W_n} [(W_n - L)^k e^{m_{u,i} \lambda_{u,i} L} - (W_n)^k]. \end{aligned} \quad (32)$$

The energy for Case 2 is expressed as

$$\begin{aligned} E_n^{(b)} &= \sum_{j=1}^n T_j E_S + (n-1) L E_L \\ &\quad + \left[z_n - \sum_{j=1}^n T_j - (n-1)L \right] E_L \end{aligned} \quad (33)$$

$$= \sum_{j=1}^n T_j E_S + \left[z_n - \sum_{j=1}^n T_j \right] E_L \quad (34)$$

where, in the right hand side of (33), the first item accounts for the energy consumption in the sleep interval during the first, second, and up to the n th cycles, the second item represents for the energy consumption in the listening interval during the first, second, and up to the $(n-1)$ st cycles, and the third item refers to the energy consumption during the listening interval in the n th cycle before sleep termination. Based on the aforementioned equation, the average energy consumption is given by

$$\begin{aligned} \overline{E_n^{(b)}} &= \sum_{j=1}^n T_j E_S + \left[\overline{z_n} - \sum_{j=1}^n T_j \right] E_L \\ &= \sum_{j=1}^n T_j E_S + \left[\frac{C_n^{(b)}}{R_n^{(b)}} - \sum_{j=1}^n T_j \right] E_L. \end{aligned} \quad (35)$$

Now, we calculate the probability that Case 2 will occur. For $N = 1$, no downlink packets arrive during the first sleep interval, and there are uplink packets arrival in the first listening interval. We have

$$\begin{aligned} \phi_1^{(b)} &= \Pr(T_1 < t_{u,r} \leq T_1 + L; t_{d,r} > T_1) \\ &= \int_{T_1}^{T_1+L} f_{t_{u,r}}(x) dx \int_{T_1}^{\infty} f_{t_{d,r}}(x) dx \\ &= [F_{t_{u,r}}(T_1 + L) - F_{t_{u,r}}(T_1)] [1 - F_{t_{d,r}}(T_1)]. \end{aligned} \quad (36)$$

For $N = n \geq 2$, the probability that Case 2 will occur is given by

$$\begin{aligned} \phi_n^{(b)} &= \Pr(W_n - L < t_{u,r} < W_n; t_{d,r} > W_n - L) \\ &= \Pr(W_n - L < t_{u,r} < W_n) \Pr(t_{d,r} > W_n - L) \\ &= \int_{W_n-L}^{W_n} f_{t_{u,r}}(x) dx \int_{W_n-L}^{\infty} f_{t_{d,r}}(y) dy \\ &= [F_{t_{u,r}}(W_n) - F_{t_{u,r}}(W_n - L)] \\ &\quad \times [1 - F_{t_{d,r}}(W_n - L)]. \end{aligned} \quad (37)$$

3) *Case 3—Sleep mode is terminated during the n th listening window due to the arrival of a downlink packet:* In this case, as shown in Fig. 4(c), there are downlink packets during the n th sleeping interval, which are temporarily buffered in the BS. Moreover, during this sleep interval T_n , there is no uplink frame. Upon completion of the n th sleep window, the MSS enters the listening state, will receive the broadcast message MOB-TRF-IND with positive indication, and will then terminate the sleep mode. Thus, the energy is given by

$$E_n^{(c)} = \overline{E_n^{(c)}} = \sum_{j=1}^n T_j E_S + (n-1) L E_L \quad (38)$$

where, on the right hand side of (38), the first item accounts for the energy consumption in the sleep interval during the first, second, and up to the n th cycles, and the second item represents for the energy consumption in the listening interval during the first, second, and up to the $(n-1)$ st cycles.

For $N = n = 1$, the probability that this situation will occur is given by

$$\begin{aligned} \phi_1^{(c)} &= \Pr(t_{d,r} \leq T_1; t_{u,r} > T_1 + L) \\ &= \Pr(t_{d,r} \leq T_1) \Pr(t_{u,r} > T_1 + L) \\ &= \int_0^{T_1} f_{t_{d,r}}(x) dx \int_{T_1+L}^{\infty} f_{t_{u,r}}(y) dy \\ &= F_{t_{d,r}}(T_1) [1 - F_{t_{u,r}}(T_1)]. \end{aligned} \quad (39)$$

For $N = n \geq 2$, the probability that this situation will occur is given by

$$\begin{aligned} \phi_n^{(c)} &= \Pr(W_{n-1} - L < t_{d,r} \leq W_n - L; t_{u,r} > W_n) \\ &= \Pr(W_{n-1} - L < t_{d,r} \leq W_n - L) \Pr(t_{u,r} > W_n) \\ &= \int_{W_{n-1}-L}^{W_n-L} f_{t_{d,r}}(x) dx \int_{W_n-L}^{\infty} f_{t_{u,r}}(y) dy \\ &= [F_{t_{d,r}}(W_n - L) - F_{t_{d,r}}(W_{n-1} - L)] \\ &\quad \times [1 - F_{t_{u,r}}(W_n)]. \end{aligned} \quad (40)$$

Substituting (26), (28), (35), (37), (38), and (40) into (22), we can calculate the power consumption while an MSS is

in the sleep mode. Note that the formula for energy consumption is expressed as a summation operator, which is due to the hyper-Erlang downlink and uplink packet interarrival time. The summation calculation substantially reduces the computation complexity and makes for easy programming using mathematics tools.

B. Downlink Packet Delay

If an uplink packet arrives, the sleep mode will instantly be terminated. Hence, the waiting time of an uplink packet is zero. In this section, we focus on the derivation of the downlink packet delay. Following the partition of the indicated three cases and referring to Fig. 4, we let $\overline{D_n^{(k)}}$ denote the downlink packet delay during case (k) ($k \in \{a, b, c\}$), provided that the MSS terminates the sleep mode in the n th cycle. Then, the average downlink packet delay is given by

$$\overline{D} = \sum_{n=1}^{\infty} \left[\overline{D_n^{(a)}} \phi_n^{(a)} + \overline{D_n^{(b)}} \phi_n^{(b)} + \overline{D_n^{(c)}} \phi_n^{(c)} \right]. \quad (41)$$

1) *Case 1—Sleep mode is terminated during the n th sleep window due to the arrival of an uplink packet:* In this situation, as shown in Fig. 4(a), the expressions for the downlink packet delay differ according to the value of random variable N . For $N = n = 1$, an uplink packet arrives during the first sleep window T_1 . Hence, the packet delay is equal to y_1 , i.e.,

$$D_1^{(a)} = y_1$$

where y_1 is given by (23), with $n = 1$. This equation results in the average delay, i.e.,

$$\overline{D_1^{(a)}} = \overline{y_1} = \frac{C_1^{(a)}}{R_1^{(a)}} \quad (42)$$

where $C_1^{(a)}$ and $R_1^{(a)}$ are, respectively, given in (50) and (51) by setting $n = 1$.

For $N = n \geq 2$, an uplink packet arrives during the n th sleep window. Based on the defined alternative random variable y_n in (23), the downlink packet waits for duration $y_n - W_{n-1}$ until the uplink packet arrives. Hence, the downlink packet delay is given by

$$D_n^{(a)} = y_n - W_{n-1}; n \geq 2. \quad (43)$$

This leads to the average delay, i.e.,

$$\begin{aligned} \overline{D_n^{(a)}} &= \overline{y_n} - W_{n-1} \\ &= \frac{C_n^{(a)}}{R_n^{(a)}} - W_{n-1}; n \geq 2 \end{aligned} \quad (44)$$

where $C_n^{(a)}$ and $R_n^{(a)}$ are given in (50) and (51), respectively.

2) *Case 2—Sleep mode is terminated during the n th listening window due to the arrival of an uplink packet:* For the situation in Fig. 4(b), an uplink packet arrives during the n th listening window and hence terminates the sleep mode. Based

on the defined random variable z_n , the downlink packet delay is given by

$$\overline{D_n^{(b)}} = \overline{z_n} - (W_n - L); n \geq 1. \quad (45)$$

3) *Case 3—Sleep mode is terminated during the n th listening window due to the arrival of a downlink packet:* For the situation in Fig. 4(c), the expressions for the downlink packet delay differ according to the value of random variable N . For $N = n = 1$, the delay is given by

$$\overline{D_1^{(c)}} = T_1 - \overline{v_1}. \quad (46)$$

For $N = n \geq 2$, downlink packets arrive during the n th sleep window. These packets wait in the BS until the next immediate listening window. Based on the defined alternative random variable v_n , the downlink packet waits for duration $(W_n - L - v_n)$. Hence, the downlink packet delay is given by

$$\overline{D_n^{(c)}} = W_n - L - \overline{v_n}; n \geq 2. \quad (47)$$

Substituting (42), (44), (28), (45), (37), (46), (47), and (40) into (41), we can calculate the average downlink packet delay. Similarly, the delay formula is expressed as a summation operator, which is due to the hyper-Erlang downlink and uplink packet interarrival time.

V. ENHANCED ENERGY-MANAGEMENT MECHANISM

As stated in Section II, an MSS alternates between the sleep mode and the wake mode. We term a sleep and the immediate next sleep mode as *neighboring sleep modes*. Fig. 5 shows the indices of all sleep modes during an MSS's lifetime. During the k th sleep mode, an MSS experiences a number of sleep windows before sleep mode termination. We define the duration of all sleep windows as the set $\{T_1^{(k)}, T_2^{(k)}, \dots, T_n^{(k)}\}$ for the k th ($k = 1, 2, \dots$) sleep mode, where $T_i^{(k)}$ denotes the i th sleep window during the k th sleep mode state. $T_1^{(k)}$ represents the first sleep window, whereas $T_n^{(k)}$ refers to the last sleep window during the k th sleep mode state.

In the enhanced scheme, the initial-sleep window during the k th sleep mode is set as $T_1^{(k)}$ instead of the constant T_{\min} , as specified in the standard. The reasoning behind this setting is given as follows. In the standard, the length T_{\min} is fixed without considering the traffic pattern. The MSS's termination of the sleep mode during the n th sleep window implies that this sleep window is not suitable for the sleep length for this specific sleep mode. From a statistical point of view, in the long run, the average duration of the sleep window right before the last sleep window (i.e., $T_{n-1}^{(k-1)}/2$) is suitable. Furthermore, using the moving average in a number of previous consecutive neighboring sleep modes, the initial-sleep window $T_1^{(k)}$ should be more appropriate. We assign the initial-sleep window $T_1^{(k)}$, which is equal to the moving average of q previous average sleep windows that come next to the last sleep window. $T_{n-1}^{(k-i)}/2$ is

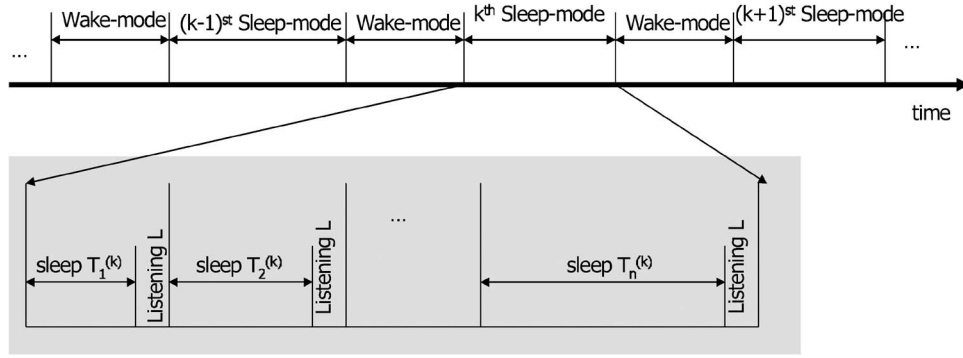


Fig. 5. Enhanced energy-management mechanism in IEEE 802.16e.

the average length of the sleep window right before the final sleep window in the $(k - i)$ th sleep mode; thus, we have

$$T_1^{(k)} = \frac{\frac{T_{n-1}^{(k-1)}}{2} + \frac{T_{n-1}^{(k-2)}}{2} + \dots + \frac{T_{n-1}^{(k-q)}}{2}}{q} = \frac{\sum_{i=1}^q T_{n-1}^{(k-i)}}{2q} \quad (48)$$

where q is a positive integer. Note that, in this equation, $T_{n-1}^{(k-i)}/2$ is set as $T_n^{(k-i)}/2$ if $n = 1$. When $T_1^{(k)}$ is smaller than T_{\min} , $T_1^{(k)}$ is set as T_{\min} . When $T_1^{(k)}$ is greater than T_{\max} , $T_1^{(k)}$ is set as T_{\max} . Moreover, after the calculation in (48), $T_1^{(k)}$ may not exactly be one of the values in the set $\mathcal{T} = \{T_1, T_2, \dots, T_{N_{\max}-1}\}$. In such a case, we choose

$$T_1^{(k)} = \begin{cases} T_i, & |T_1^{(k)} - T_i| < |T_1^{(k)} - T_{i+1}| \\ T_{i+1}, & |T_1^{(k)} - T_i| \geq |T_1^{(k)} - T_{i+1}| \end{cases}$$

where $i = 1, 2, \dots, N_{\max} - 1$.

In summary, the enhanced energy-management mechanism is presented as follows. An MSS defines a variable to store the updated initial-sleep window length. Whenever the MSS enters the sleep mode, it chooses the first sleep window $T_1^{(k)}$ based on (48). After the first sleep interval, the MSS enters a listening state and listens to the traffic indication message MOB-TRF-IND broadcast from the BS. If MOB-TRF-IND indicates in the negative, the MSS continues in the sleep mode. The next sleep window will be twice the size of the preceding sleep interval if $2 \times T_1^{(k)}$ is not more than T_{\max} . Otherwise, the next sleep window remains fixed at T_{\max} . This process is repeated until the sleep mode is terminated. It is clear that only the first length of the sleep window length differs from the scheme in IEEE 802.16e, which implies that our proposed scheme is compatible with the standardized energy-management scheme in IEEE 802.16e WirelessMAN.

Let $\mathcal{E}^{(k)}$ denote the consumed energy in the k th sleep mode during an MSS's lifetime. Then, the average consumed energy during the sleep mode in the enhanced scheme is expressed as

$$\bar{\mathcal{E}} = \lim_{k \rightarrow \infty} \frac{\mathcal{E}^{(1)} + \mathcal{E}^{(2)} + \dots + \mathcal{E}^{(k)}}{k}.$$

Let $\mathcal{D}^{(k)}$ denote the delay in the k th sleep mode during an MSS's lifetime. Then, the average delay during a sleep mode

period in the enhanced scheme is given by

$$\bar{\mathcal{D}} = \lim_{k \rightarrow \infty} \frac{\mathcal{D}^{(1)} + \mathcal{D}^{(2)} + \dots + \mathcal{D}^{(k)}}{k}.$$

VI. NUMERICAL RESULTS

In this section, our major objectives are to validate the analytical model via the extensive simulation results and to show the improved performance of our proposed energy-saving scheme. Illustrative numerical examples are also presented to demonstrate the performance tradeoff. The results will additionally show the interaction between the performance metrics and critical settings. To verify the analysis, we simulate the MSS behavior in a 2-D plane using Monte Carlo simulation [19] to obtain the simulation results. A subprogram is written to simulate an MSS behavior. In particular, for each MSS, a random variable is generated to determine the moment of the next packet arrival. Based on this setup, we can calculate the required sleeping period and, hence, the consumed energy in this period. Each simulation result is the average of 100 000 MSS energy consumption. If not specified, we choose the following default parameters: 1) $L = 1$; 2) $T_{\min} = 1$; and 3) $T_{\max} = 1024$. Here, L , T_{\min} , and T_{\max} are measured in units of MAC frames. The consumed energy units are set as $E_L = 0.045$ W and $E_S = 1.5$ W [9]. In what follows, without loss of generality, we employ Erlang distributed downlink and uplink packet interarrival times. Reasons for this employment include the general formulation for the exponential distribution and the straightforward extension into a hyper-Erlang distribution, which has been reported to preserve the universal approximation capability. Then, the parameters in the hyper-Erlang distributed t_d in (2) become $H_d = 1$, $\alpha_{d,1} = 1$, $\lambda_{d,1} = \lambda_d$, and the stage $m_{d,1}$. For presentation purposes, we drop the index of the stage parameter $m_{d,1}$ and use m_d instead. In this case, m_d -stage Erlang distributed t_d has the pdf $f_{t_d}(t) = [(m_d \lambda_d)^{m_d} t^{m_d-1} / (m_d - 1)!] e^{-m_d \lambda_d t}$. Similarly, m_u -stage Erlang distributed t_u has the pdf $f_{t_u}(t) = [(m_u \lambda_u)^{m_u} t^{m_u-1} / (m_u - 1)!] e^{-m_u \lambda_u t}$.

A. Analytical Model Validation: Downlink Traffic Process

In this section, a number of illustrative numerical examples will be presented to demonstrate the validity of both the analytical model and the simulation program.

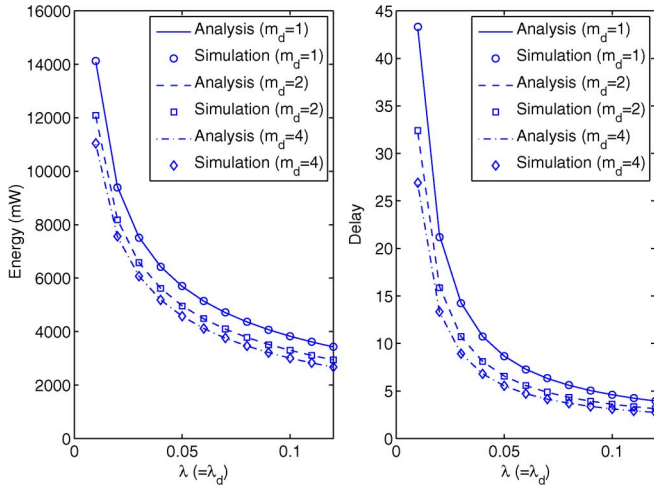


Fig. 6. Energy consumption and delay during the sleep mode in terms of downlink packet arrival rate λ_d . ($T_{min} = 1$ and $T_{max} = 1024$).

Fig. 6 shows the energy consumption and packet delay during sleep modes in terms of downlink packet arrival rate λ_d . The parameters are set as $T_{min} = 1$ and $T_{max} = 1024$. The simulation results are presented (indicated by symbol) to validate the analytical model. It is clear that the results of simulation and analysis match with each other very well. With more frequent downlink packet arrivals, i.e., larger λ_d , the sleep mode is terminated with higher probability, leading to shorter sleep mode duration and, consequently, less energy consumption and shorter downlink packet delay during the sleep mode. This setup will eventually result in more energy consumption during wake modes in the MSS's lifetime. As previously stated, one of the main motivations of this paper is the traffic process generalization. As a consequence, it is necessary to investigate the traffic distributions effect, i.e., whether energy consumption and packet delay reflect a significant variation with different traffic processes. The different stage m_d represents different traffic processes, where $m_d = 1$ is equivalent to the Poisson arrival process. The comparison shows that a smaller m_d leads to greater energy consumption during the sleep mode and longer packet delay. This scenario can be explained as follows. A smaller m_d indicates greater variance in the traffic-arrival process. In such a case, many more MSSs with longer t_d will be observed. This phenomenon will significantly contribute to the longer sleeping period and, consequently, more energy consumption \mathcal{E} and larger packet delay \bar{D} during the sleep mode.

Fig. 7 shows the energy consumption and packet delay during sleep modes in terms of downlink packet arrival rate λ_d with $T_{min} = 16$ and $T_{max} = 1024$. Fig. 8 shows the energy consumption and packet delay during sleep modes in terms of downlink packet arrival rate λ_d . The parameters are set as $T_{min} = 1$ and $T_{max} = 64$. Again, the comparison indicates that the analysis is consistent with the simulation. A similar effect for m_d can be observed. In addition, these two figures also indicate that the energy consumption and delay during the sleep mode decrease with higher packet arrival rates.

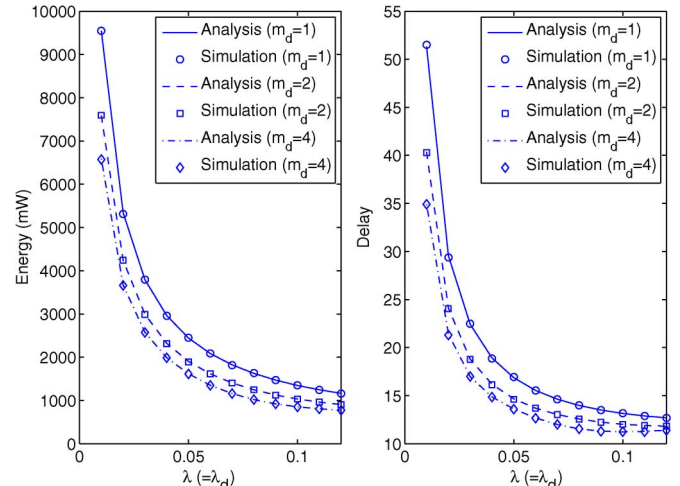


Fig. 7. Energy consumption and delay during the sleep mode in terms of downlink packet arrival rate λ_d . ($T_{min} = 16$ and $T_{max} = 1024$).

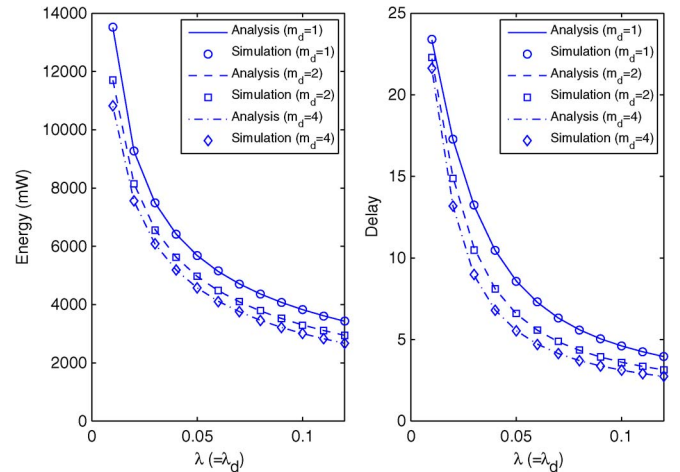


Fig. 8. Energy consumption and delay during the sleep mode in terms of downlink packet arrival rate λ_d . ($T_{min} = 1$ and $T_{max} = 64$).

B. Enhanced Scheme Versus Standard Scheme in IEEE 802.16e: Downlink Traffic Process

Fig. 9 shows the reduced energy consumption and introduced extra delay in our proposed mechanism with different downlink traffic processes. This example demonstrates the condition when only a downlink traffic process is considered. We let $\mathcal{E}_{802.16e}$ and $\mathcal{E}_{Enhanced}$ denote the energy consumption during the sleep mode in the standard scheme in IEEE 802.16e and in our proposed scheme, respectively. The saved energy consumption is then expressed as $(\mathcal{E}_{802.16e} - \mathcal{E}_{Enhanced})/\mathcal{E}_{802.16e}$. Similarly, we let $\mathcal{D}_{802.16e}$ and $\mathcal{D}_{Enhanced}$ denote the packet delay during the sleep mode in the standard scheme in IEEE 802.16e and in our proposed scheme, respectively. The extra delay introduced by the proposed scheme is then expressed as $(\mathcal{D}_{Enhanced} - \mathcal{D}_{802.16e})/\mathcal{D}_{802.16e}$. For ease of comparison, we have employed an identical scale on the y-axis for the two subfigures. The comparison indicates that, for a fixed arrival rate, power consumption can significantly be reduced. For instance, for fixed $\lambda_d = 0.02$, the saved energy can be 37%. We also notice that the introduced extra delay is around 25% under the same condition, which is much smaller than

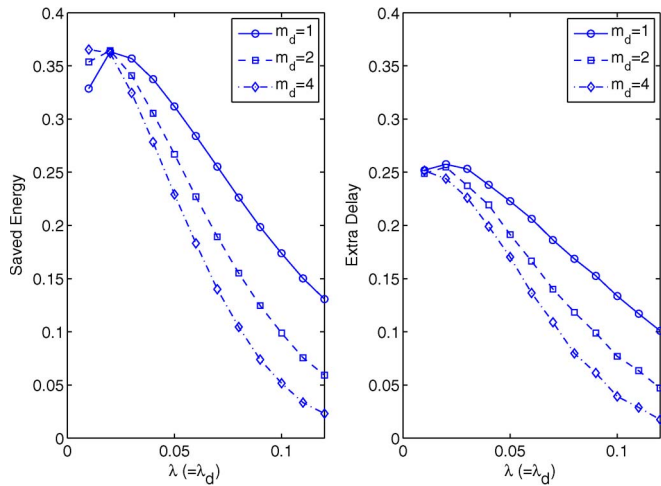


Fig. 9. Saved energy consumption and extra delay in terms of downlink packet arrival rate λ_d .

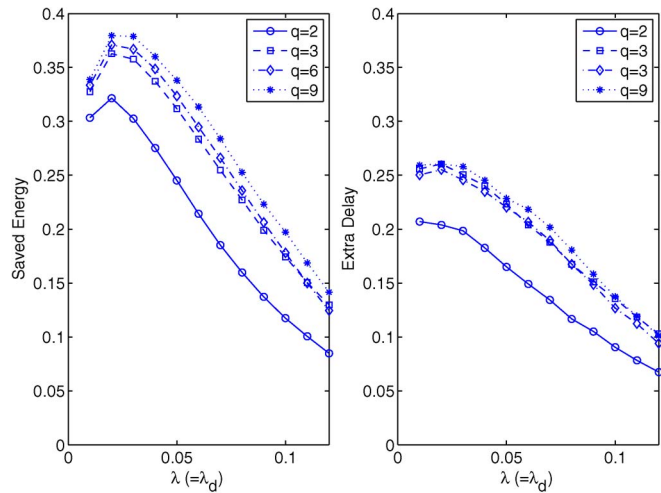


Fig. 10. Saved energy consumption and extra delay with different q .

the percentage with respect to the saved energy and indicates the advantage of our proposed strategy. The saved energy and extra delay decrease as the traffic arrival rate increases. Under other identical conditions, the percentage of saved energy is always higher than that of extra delay. Furthermore, the sleeping mechanism is standardized for best effort and non-real-time services, which is insensitive to a small increase in packet delay. As a consequence, with respect to the tradeoff between energy consumption and delay, our proposed scheme outperforms the standard scheme in IEEE 802.16e.

C. Effect of q on the Enhanced Scheme

Fig. 10 shows the effect of q on the saved energy and extra delay. It may be shown that various q could achieve different energy consumption and delay in the enhanced scheme. For a fixed arrival rate, $q = 2$ can substantially reduce power consumption with acceptable delay tradeoff. $q = 3$ can further reduce power consumption. After $q \geq 3$, there is an insignificant discrepancy with even larger q in terms of energy consumption and packet delay. Hence, the value of q is unnecessarily very large. The parameter q is essentially the moving average size.

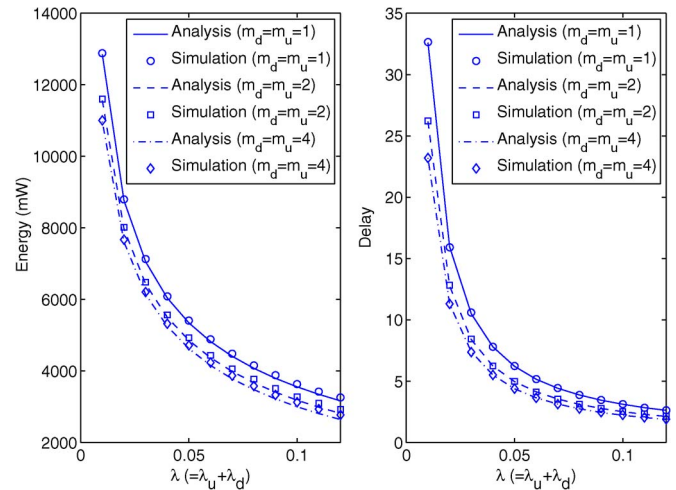


Fig. 11. Energy consumption and downlink packet delay during the sleep mode in terms of total arrival rate λ . ($T_{\min} = 1$ and $T_{\max} = 1024$).

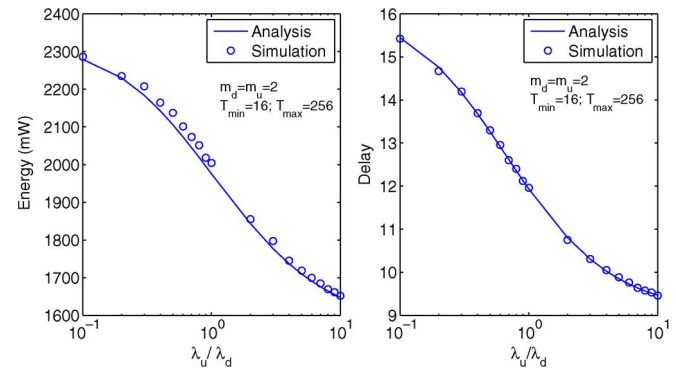


Fig. 12. Energy consumption and downlink packet delay during the sleep mode in terms of downlink and uplink arrival rate ratio λ_u/λ_d .

A larger q shows that more values should be saved, a longer sliding window is required for moving average operation, and concurrently, higher implementation complexity is involved. As a consequence, we are inclined to choose a small value of q with significantly reduced energy and acceptable delay. For instance, q can be chosen as 3 in this example.

D. Analytical Model Validation: Downlink- and Uplink Traffic Processes

Fig. 11 shows the energy consumption and packet delay during the sleep mode in terms of the total packet arrival rate λ ($\lambda_d = \lambda_u = \lambda/2$). It may be shown that the analysis and simulation results agree with each other. In addition, different downlink and uplink traffic processes are investigated. $m_d = 1$ or $m_u = 1$ indicates the Poisson downlink or uplink traffic process. Similar to Section VI-A, a similar effect of m_d or m_u can be observed and explained following similar reasoning. Note that we also checked the analysis and simulation results under other T_{\min} and T_{\max} situations and found agreement.

E. Effect of Downlink and Uplink Arrival Rate Ratio λ_u/λ_d

Fig. 12 shows the energy and delay during the sleep mode in terms of downlink and uplink arrival rate ratio λ_u/λ_d with fixed

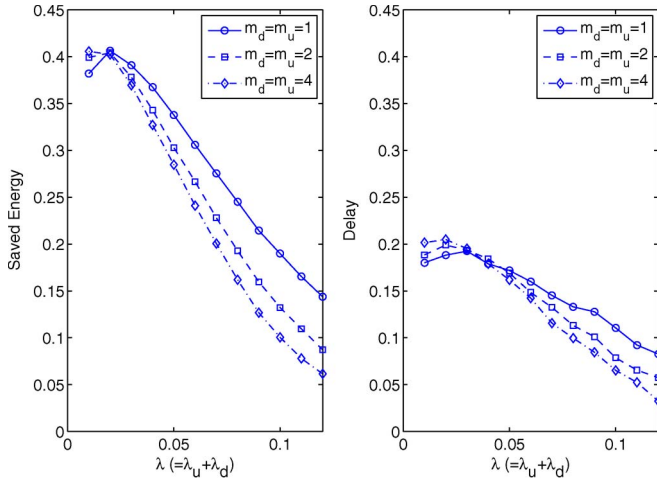


Fig. 13. Saved energy consumption and extra delay in terms of the total packet arrival rate $\lambda(\lambda_d = \lambda_u = \lambda/2)$.

$\lambda = 0.04$. The comparison shows that the consumed energy varies with the different traffic directions, because with more uplink frames, the sleep mode is instantaneously terminated by uplink frames with higher possibility, leading to shorter sleep mode duration and, consequently, less energy consumption during the sleep mode.

F. Enhanced Scheme Versus Standardized Scheme in IEEE 802.16e: Downlink and Uplink Traffic Processes

Fig. 13 shows the saved energy consumption and introduced extra delay in our proposed mechanism with different downlink and uplink traffic processes. This example demonstrates the situation when both downlink and uplink traffic processes are considered. The comparison indicates that, for a fixed arrival rate, power consumption can significantly be reduced. For instance, for fixed $\lambda_d = 0.02$, the saved energy can be 40%. We also notice that the corresponding introduced extra delay is around 20%, which is half of the percentage of the saved energy. Under other similar conditions, the percentage of the saved energy is always much higher than that of the extra delay. As a consequence, our proposed scheme outperforms the standardized scheme in IEEE 802.16e with respect to the tradeoff between energy consumption and delay, regardless of whether the uplink traffic process is considered.

VII. CONCLUSION AND FUTURE WORK

We have proposed an analytical model to evaluate energy management in the IEEE 802.16e WirelessMAN standard. Both downlink and uplink traffic processes have been relaxed to develop formulas that have general applications. The analysis has been validated by the simulations. Both the methodology and results provide potential guidance to efficiently manage energy. Furthermore, an enhanced energy management mechanism has been proposed by adaptively choosing the sleep window size. The scheme can significantly reduce energy consumption, and the introduced packet delay is much lower than the percentage of the saved energy. Under similar situations in the proposed strategy, the percentage of the saved energy

is always higher than that of the introduced extra delay. One interesting future topic is the tradeoff optimization by choosing appropriate sleeping-window parameters. To accomplish this task, we need to define a new system-level performance metric to simultaneously integrate the effects of energy consumption and delay. The comprehensive and detailed investigation on this research effort will be left to future work.

APPENDIX I

RECURSIVE ALGORITHM FOR CALCULATING $\Theta_n(a, b; \beta)$

$\Theta_n(a, b; \beta)$ is defined as

$$\Theta_n(a, b; \beta) = \int_a^b x^n e^{-\beta x} dx$$

where a, b , and β are real numbers, and n is a nonnegative integer. We have

$$\begin{aligned} \Theta_n(a, b; \beta) &= \int_a^b x^n e^{-\beta x} dx = \frac{1}{n+1} \int_a^b e^{-\beta x} dx^{n+1} \\ &= \frac{1}{n+1} \left[x^{n+1} e^{-\beta x} \Big|_a^b + \beta \int_a^b x^{n+1} e^{-\beta x} dx \right] \\ &= \frac{1}{n+1} \{ [b^{n+1} e^{-\beta b} - a^{n+1} e^{-\beta a}] + \beta \Theta_{n+1}(a, b; \beta) \}. \end{aligned}$$

Reformatting the aforementioned equation, we develop the recursive algorithm as follows:

$$\Theta_n(a, b; \beta) = \frac{1}{\beta} [n\Theta_{n-1}(a, b; \beta) - (b^n e^{-\beta b} - a^n e^{-\beta a})]$$

with initial condition

$$\Theta_0(a, b; \beta) = \int_a^b e^{-\beta x} dx = \frac{1}{\beta} (e^{-\beta a} - e^{-\beta b}).$$

APPENDIX II
DEVELOPING B_n

We have

$$\begin{aligned} B_n &= \int_{W_{n-1}-L}^{W_n-L} x f_{t_{d,r}}(x) dx \\ &= \int_{W_{n-1}-L}^{W_n-L} x \lambda_d \sum_{i=1}^{H_d} \alpha_{d,i} \sum_{j=0}^{m_{d,i}-1} \frac{(m_{d,i} \lambda_{d,i} x)^j}{j!} e^{-m_{d,i} \lambda_{d,i} x} dx \\ &= \lambda_d \sum_{i=1}^{H_d} \alpha_{d,i} \sum_{j=0}^{m_{d,i}-1} \frac{(m_{d,i} \lambda_{d,i})^j}{j!} \\ &\quad \times \Theta_{j+1}(W_{n-1} - L, W_n - L; m_{d,i} \lambda_{d,i}). \end{aligned} \tag{49}$$

APPENDIX III
DEVELOPING $C_n^{(a)}$ AND $R_n^{(a)}$

We have

$$\begin{aligned}
 C_n^{(a)} &= \int_{W_{n-1}}^{W_{n-1}+T_n} x f_{t_{u,r}}(x) dx \\
 &= \int_{W_{n-1}}^{W_{n-1}+T_n} x \lambda_u \sum_{i=1}^{H_u} \alpha_{u,i} \\
 &\quad \times \sum_{j=0}^{m_{u,i}-1} \frac{(m_{u,i} \lambda_{u,i} x)^j}{j!} e^{-m_{u,i} \lambda_{u,i} x} dx \\
 &= \lambda_u \sum_{i=1}^{H_u} \alpha_{u,i} \sum_{j=0}^{m_{u,i}-1} \frac{(m_{u,i} \lambda_{u,i})^j}{j!} \\
 &\quad \times \Theta_{j+1}(W_{n-1}, W_{n-1} + T_n; m_{u,i} \lambda_{u,i}) \quad (50) \\
 R_n^{(a)} &= \int_{W_{n-1}}^{W_{n-1}+T_n} f_{t_{u,r}}(x) dx \\
 &= F_{t_{u,r}}(W_{n-1} + T_n) - F_{t_{u,r}}(W_{n-1}) \\
 &= \lambda_u \sum_{i=1}^{H_u} \frac{\alpha_{u,i}}{m_{u,i} \lambda_{u,i}} \sum_{j=0}^{m_{u,i}-1} \sum_{k=0}^j \frac{(m_{u,i} \lambda_{u,i})^k}{k!} \\
 &\quad \times e^{-m_{u,i} \lambda_{u,i} W_{n-1}} \\
 &\quad \times [(W_{n-1})^k - (W_{n-1} + T_n)^k e^{-m_{u,i} \lambda_{u,i} T_n}]. \quad (51)
 \end{aligned}$$

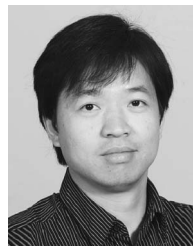
ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their careful reading and insightful comments that have significantly improved the quality of this paper.

REFERENCES

[1] Wireless MAN Working Group. [Online]. Available: <http://wirelessman.org/>
 [2] *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Std. 802.16-2004, Jun. 2004.
 [3] *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1*, IEEE Std. 802.16e-2005 and IEEE Std. 802.16-2004/Cor 1-2005 (Amendment and Corrigendum to IEEE Std. 802.16-2004), 2006.
 [4] I. Koffman and V. Roman, "Broadband wireless access solutions based on OFDM access in IEEE 802.16," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 96–103, Apr. 2004.
 [5] C. Cicconetti, L. Lenzini, E. Mingozzi, and C. Eklund, "Quality-of-service support in IEEE 802.16 networks," *IEEE Netw.*, vol. 20, no. 2, pp. 50–55, Mar./Apr. 2006.
 [6] A. Ghosh, D. R. Wolter, J. G. Andrews, and R. Chen, "Broadband wireless access with WiMAX/802.16: Current performance benchmarks and future potential," *IEEE Commun. Mag.*, vol. 43, no. 2, pp. 129–136, Feb. 2005.

[7] M. Zorzi, "Energy management in personal communications and mobile computing," *IEEE Pers. Commun.*, vol. 5, no. 3, p. 10, Jun. 1998.
 [8] Y. Zhang and M. Fujise, "Energy management in the IEEE 802.16e MAC," *IEEE Commun. Lett.*, vol. 10, no. 4, pp. 311–313, Apr. 2006.
 [9] E. S. Jung and N. H. Vaidya, "An energy efficient MAC protocol for wireless LANs," in *Proc. IEEE INFOCOM*, Jun. 2002, vol. 3, pp. 1756–1764.
 [10] Y. Xiao, "Energy-saving mechanism in the IEEE 802.16e wireless MAN," *IEEE Commun. Lett.*, vol. 9, no. 7, pp. 595–597, Jul. 2005.
 [11] M. A. Marsan, G. Ginella, R. Maglione, and M. Meo, "Performance analysis of hierarchical cellular networks with generally distributed call holding times and dwell times," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 248–257, Jan. 2004.
 [12] Y. G. Fang, "Hyper-Erlang distribution model and its application in wireless mobile networks," *ACM Wireless Netw.*, vol. 7, no. 3, pp. 211–219, May 2001.
 [13] Y. B. Lin and Y. K. Chen, "Reducing authentication signaling traffic in third-generation mobile network," *IEEE Trans. Wireless Commun.*, vol. 2, no. 3, pp. 493–501, May 2003.
 [14] L. Kong and D. H. K. Tsang, "Performance study of power saving classes of type I and II in IEEE 802.16e," in *Proc. 31st IEEE Conf. LCN*, Tampa, FL, Nov. 2006, pp. 20–27.
 [15] E. Casilari, F. J. Gonzalez, and F. Sandoval, "Modeling of HTTP traffic," *IEEE Commun. Lett.*, vol. 5, no. 6, pp. 272–274, Jun. 2001.
 [16] M. E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: Evidence and possible causes," *IEEE/ACM Trans. Netw.*, vol. 5, no. 6, pp. 835–846, Dec. 1997.
 [17] L. Kleinrock, *Queueing Systems*. New York: Wiley, 1975.
 [18] F. P. Kelly, *Reversibility and Stochastic Networks*. New York: Wiley, 1979.
 [19] G. S. Fishman, *Monte Carlo: Concepts, Algorithms and Applications*. New York: Springer-Verlag, 1997.



Yan Zhang (M'05) received the Ph.D. degree in electrical and electronics engineering from Nanyang Technological University, Singapore.
 Since August 2006, he has been with Simula Research Laboratory, Lysaker, Norway. He is a Guest Editor, a Regional Editor, an Associate Editor, or a member of the editorial board of several international journals. He is currently the Book Series Editor of *Wireless Networks and Mobile Communications* (Auerbach, CRC, Taylor & Francis). His research interests include resource, mobility, spectrum, data, energy, and security management in wireless networks and mobile computing.

Dr. Zhang is a member of Technical Program Committees of numerous international conferences, including the IEEE International Conference on Communications (ICC), the IEEE Global Telecommunications Conference (GlobeCom), the IEEE Wireless Communications and Networking Conference (WCNC), the IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), the IEEE Vehicular Technology Conference (VTC), and the IEEE Consumer Communications and Networking Conference (CCNC). He is also the Chair of the Organizing Committee for several international conferences. He is a Publicity Cochair for the Sixth IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS 2009), a Program Cochair for the Fifth International Conference on Broadband Communications, Networks, and Systems (Broadnets 2009), Program Co-Chair for the 2009 International Wireless Communication and Mobile Computing Conference (IWCMC), a Symposium Cochair for the 2008 and 2009 International Conference on Communications and Networking in China (ChinaCom), an Industrial Cochair for the Ninth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2008), a Program Cochair for the Fifth International Conference on Ubiquitous Intelligence and Computing (UIC 2008), a Program Vice Cochair for the Ninth IEEE International Symposium on Multimedia (ISM 2007), and the Publications Chair for the Fourth IEEE International Symposium on Wireless Communication Systems (ISWCS 2007). He received the Best Paper Award at the 21st IEEE International Conference on Advanced Information Networking and Applications (AINA 2007).



Yang Xiao (SM'04) received the M.S. and Ph.D. degrees in computer science and engineering from Wright State University, Dayton, OH.

He was a Medium Access Control (MAC) architect for the IEEE 802.11 standard enhancement before joining the Department of Computer Science, University of Memphis, Memphis, TN, in 2002. He is currently with the Department of Computer Science, University of Alabama, Tuscaloosa. He was a Referee/Reviewer for many funding agencies, as well as a Panelist for the U.S. National Science

Foundation and a member of the Telecommunications Expert Committee of the Canada Foundation for Innovation. He is currently the Editor-in-Chief of the *International Journal of Security and Networks*, the *International Journal of Sensor Networks*, and the *International Journal of Telemedicine and Applications*. He has published more than 300 papers in major journals, refereed conference proceedings, and book chapters related to these research areas. His research interests include security, telemedicine, sensor networks, and wireless networks.

Dr. Xiao is a member of the American Telemedicine Association. From 2001 to 2004, he was a Voting Member of the IEEE 802.11 Working Group. He has published more than 60 papers in various IEEE journals and magazines. He serves on the Technical Program Committee of more than 100 conferences, such as the IEEE International Conference on Computer Communications (INFOCOM), the IEEE International Conference on Distributed Computing Systems (ICDCS), the ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), the IEEE International Conference on Communications (ICC), the IEEE Global Telecommunications Conference (Globecom), and the IEEE Wireless Communications and Networking Conference (WCNC). He is an Associate Editor for several journals, including the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.



Victor C. M. Leung (S'75-M'89-SM'97-F'03) received the B.A.Sc. degree (with honors) in electrical engineering, for which he was awarded the APEBC Gold Medal as the head of the graduating class of the Faculty of Applied Science, and the Ph.D. degree in electrical engineering, under a Natural Sciences and Engineering Research Council Postgraduate Scholarship, from the University of British Columbia (UBC), Vancouver, BC, Canada, in 1977 and 1981, respectively.

From 1981 to 1987, he was a Senior Member of Technical Staff with MPR Teltech Ltd., specializing in the planning, design, and analysis of satellite communication systems. In 1988, he was a Lecturer with the Department of Electronics, Chinese University of Hong Kong, Shatin, Hong Kong. In 1989, he returned to UBC as a Faculty Member, where he is currently a Professor, the TELUS Mobility Research Chair in Advanced Telecommunications Engineering with the Department of Electrical and Computer Engineering, and a member of the Institute for Computing, Information, and Cognitive Systems. His research interests include wireless networks and mobile systems. He has published more than 380 journal and conference papers.

Dr. Leung is a Fellow of the Engineering Institute of Canada, a Fellow of the Canadian Academy of Engineering, and a Voting Member of the Association for Computing Machinery. He serves on the editorial board of several journals, including the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE TRANSACTIONS ON COMPUTERS. He has served on the Technical Program Committee (TPC) of numerous conferences. He was the TPC Vice Chair of the 2005 IEEE Wireless Communications and Networking Conference (WCNC), a General Cochair of the Eighth ACM/IEEE International Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems Symposium (MSWiM 2005), the General Chair of the ACM/IEEE Fourth International Conference on Heterogeneous Networking for Quality, Reliability, Security, and Robustness (QShine 2007), and the TPC Chair of the wireless networks and cognitive radio track of the 68th IEEE Vehicular Technology Conference (VTC-Fall 2008).