

Modeling and Analysis of the Type I Power Saving Class in IEEE 802.16 with General Traffic (Invited Paper)

Ming-Fang Chen[†] and Hung-Yun Hsieh^{†§}

[†]Graduate Institute of Communication Engineering

[§]Department of Electrical Engineering

National Taiwan University

Taipei, Taiwan 106

hyhsieh@cc.ee.ntu.edu.tw

ABSTRACT

IEEE 802.16 defines three types of power saving classes for supporting the sleep mode operations on mobile stations using different types of traffic. While related work has developed analytical models to evaluate the performance of the power saving operations in IEEE 802.16 networks, most of these endeavors are limited to the Poisson traffic model and hence their capability is limited in capturing the characteristics of the sleep mode operation under realistic traffic distributions. In this paper, we propose an analytic model for capturing the behaviors of IEEE 802.16 sleep mode operations under the type I power saving class with arbitrary traffic distribution, including the Pareto and 4IPP traffic models. We compare the proposed model against existing work, and explain how such a general model allows a better investigation of the impact of different traffic models on optimal performance of IEEE 802.16 power saving operations.

Categories and Subject Descriptors

C.2.1 [Computer-Communication Networks]: Network Architecture and Design—*Wireless communication*

General Terms

Performance, Theory

Keywords

IEEE 802.16e, sleep mode, Pareto traffic model, 4IPP traffic model

1. INTRODUCTION

The IEEE 802.16 standard specifies a broadband wireless access technology that aims to provide high-speed, high-bandwidth and high-capacity multimedia services for resi-

dential as well as enterprise applications. While the standard was originally conceived to support fixed broadband access, it has now been extended to support mobile access in consideration of the promising application areas for mobile services [8].

One important extension of IEEE 802.16 to support the mobile station (MS) is the design of the power saving operations. Clearly, for scenarios where an MS does not need to be active in transmitting or receiving data all the time, it is critical to support energy-efficient power saving operations so the lifetime of the MS can be prolonged. Toward this objective, IEEE 802.16 defines two operation modes for each MS including the normal mode and the sleep mode. During the normal operation of an MS, it may conduct pre-negotiated periods of absence from the serving base station (BS) air interface and enter the sleep mode for power saving, where the MS may power down one or more physical operation components to minimize its usage of the power and other air interface resources.

In the IEEE 802.16 standard, three types of power saving mechanisms named Power Saving Classes (PSCs) are defined for supporting different applications used by an MS. Each PSC consists of one or more interleaving sleep and listening windows with different parameters for optimal application performance during the power saving operation. Briefly, the PSC of type I is used for non-real-time traffic with burst behavior such as web browsing, while the PSC of type II is designed for real-time traffic such as VoIP service and the PSC of type III is recommended for multicast and management connections.

While power saving mechanism has the benefit of minimizing power usage on MSs, it may potentially increase the access delay exposed to applications. A clear understanding of the impact of different power saving parameters on application performance thus is important for balancing the tradeoffs between energy consumption and access delay. Related work has focused on analyzing and profiling the performance of power saving mechanisms in IEEE 802.16 for different types of PSCs, and several analytical models have been proposed [3, 6, 9, 15, 16]. Most of these analytical models, however, are limited to using the Poisson traffic model along with making several simplified assumptions on IEEE 802.16 sleep mode operations. While the Poisson traffic model is an elegant analytical tool for describing bursty traffic, it nonetheless is limited in its capability of capturing realistic traffic behaviors that can be used for mobile

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WICON '08, November 17-19, 2008, Maui, Hawaii, USA

Copyright 2008 ICST 978-963-9799-36-3 ...\$5.00.

access in IEEE 802.16 networks [1]. Optimal parameters for the power saving operations derived under the Poisson traffic model thus can be suboptimal and even biased for operations under other more realistic traffic models.

In this paper, we propose an analytical model for modeling the power saving operations of IEEE 802.16 with general traffic. We focus on the power saving class of type I in consideration of the complicated set of parameters involved in controlling the power saving operations compared to the other two power saving classes. We use the packet response delay as the performance metric, and aim to investigate the impact of different type I parameters including the initial and final sleep window sizes. The proposed model provides a framework for investigating the relation between packet delay and different power saving parameters for arbitrary traffic models. In addition to using the Poisson traffic model, we also consider how the Pareto traffic model and the 4IPP traffic model proposed by the IEEE 802.16 working group for simulating the performance of the MAC/PHY operations in IEEE 802.16 [1] can be incorporated in this framework. Evaluation results compare the proposed model with existing work under the Poisson traffic model, and show how different traffic models affect the performance of IEEE 802.16 power saving operations.

The rest of this paper is organized as follows. Section 2 gives a brief description of the power saving class of type I, and then discusses related work on modeling the power saving operations of IEEE 802.16. Section 3 presents the proposed analytical model for general traffic, while Section 4 shows numerical results of the proposed model and compares the proposed model against related work. Finally, Section 5 concludes the paper.

2. BACKGROUND AND RELATED WORK

2.1 Power Saving Class of Type I

The sleep mode defined in the IEEE 802.16 specification is a state where an MS conducts pre-negotiated periods of absence from the serving BS air interface for minimizing usage of power and air interface resources. Three power saving classes (PSCs) that differ by their parameter sets, procedures of activation and deactivation, and policies of MS availability for data transmission are defined in IEEE 802.16 for supporting the sleep mode operation.

Each PSC is composed of sleep and listening windows sequence and may be repeatedly activated and deactivated. PSC of type I is used for best effort (BE) and non-real-time variable rate (NRT-VR) type services, where its sleep window increases exponentially from an initial value up to the maximum value during the sleep mode. PSC of type II, on the other hand, is designed for unsolicited grant service (UGS) and real-time variable rate (RT-VR) service, where its sleep windows are constant and repeated alternately with listening windows according to the generation interval of the real-time traffic. PSC of type III is recommended for multicast connections and management connections, where it consists of one sleep window of fixed duration. While each PSC is designed for different types of traffic with a specific set of parameters, in this paper we focus on the type I PSC since it has the most complicated set of parameters among the three. In the following, we give a brief description of the operation of the PSC of type I.

To activate the PSC of type I, an MS sends a sleep request

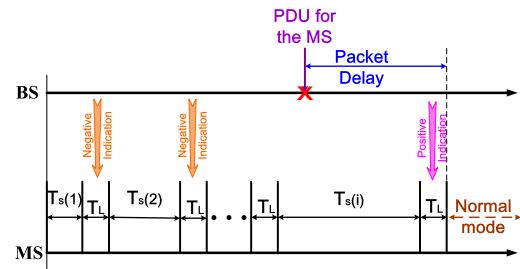


Figure 1: Packet delay in the PSC of type I

message (MOB_SLP-REQ) to the BS, which then responds with a sleep response message (MOB_SLP-RSP) containing negotiated parameters, including the start frame number for the first sleep window, initial sleep window (T_{min}), final sleep window base (B_{max}) and exponent (E_{max}), and the listening window (T_L). The power saving operation becomes active at the designated start frame with a sleep window of duration T_{min} . At the end of the initial sleep window, the MS transits to the listening state for a duration of T_L where it listens to the traffic indication message (MOB_TRF-IND) to determine whether there is pending traffic addressed to it. A traffic indication message is sent by the BS on the broadcast CID or sleep mode multicast CID during the listening window to alert the MS of the availability of downlink (DL) traffic when the MS is in the sleep state. If MOB_TRF-IND has a positive indication, the MS proceeds to deactivate the PSC for receiving the incoming traffic. Otherwise, if MOB_TRF-IND has a negative indication, the sleep mode operation will be continued. The duration of each next sleep window is doubled the size of the previous one, upper bounded by the final window value T_{max} determined by $T_{max} = B_{max} \cdot 2^{E_{max}}$.

Note that if any packet arrives during the sleep state of the MS, BS should wait to transmit the packet until the start time of the upcoming listening window for sending a positive indication. The packet response delay incurred thus depends on the duration of the sleep window when it arrives at the MS as shown in Figure 1. We show in Section 3 an analytical model for calculating the packet delay for MSs operated under the PSC of type I.

2.2 Related Work

Related work has investigated the analytical performance of the power saving operations in IEEE 802.16. In [14], the author proposes an analytical model for the PSC of type I in IEEE 802.16, considering downlink traffic that follows the Poisson distribution. The paper evaluates the average number of sleep cycles, sleeping time, energy consumption, and frame response delay. The model is extended by [3,6,9,15,16] with different considerations, but overall this line of analysis follows the per-packet analysis where the entire power saving procedure is divided into sub-procedures, and each sub-procedure considers only the operation for one packet. The arrival of a packet terminates the sub-procedure where the desired performance metrics can be determined.

Another line of analysis follows the Markov chain model for analyzing the power saving operations of IEEE 802.16 [5,

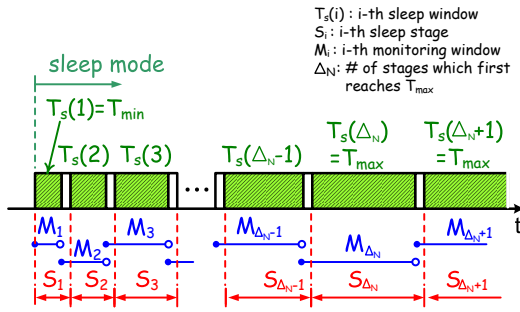


Figure 2: Sleep and monitoring windows

10–13, 17]. In these models, the power saving operation is modeled using Markov chain with different states, including the normal state and states for individual sleep cycles in the sleep mode. Steady-state probability for each of the state thus can be evaluated and the operation of the power saving mechanism can be modeled.

Existing analysis assumes the use of the Poisson process as the traffic model. While the Poisson process is simple, it does not captures the characteristics of realistic traffic distributions, and hence the analytical model based on the Poisson process is limited in its scope. We show in the following sections how we generalize the analytical model of the PSC of type I, and use it to investigate the performance of the power saving operation under different traffic models beyond the Poisson model.

3. ANALYTICAL MODEL

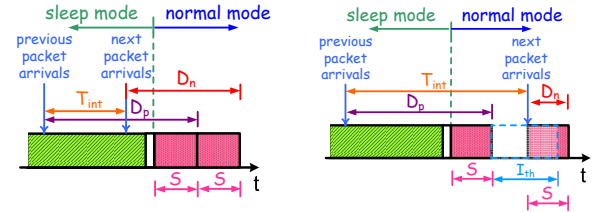
To model the operation of the PSC of type I in IEEE 802.16, we start from the case when an MS is in the normal mode. After being idle in the normal mode for a threshold of time I_{th} , the MS decides to enter the sleep mode following the procedure described in Section 2.1. For the MS, the sleep mode can be divided into several sleep stages as shown in Figure 2. The i^{th} sleep stage S_i consists of a sleep window $T_S(i)$ followed by a listening window T_L , where

$$T_S(i) = \begin{cases} 2^{i-1}T_{min}, & \text{if } 2^{i-1}T_{min} < T_{max} \\ T_{max}, & \text{if } 2^{i-1}T_{min} \geq T_{max} \end{cases} \quad (1)$$

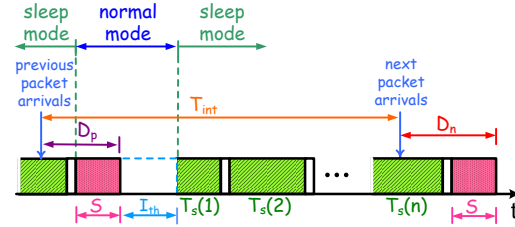
Let Δ_N^{th} be the stage when the sleep window size is equal to T_{max} . We have

$$\Delta_N = \lceil \log_2 \frac{T_{max}}{T_{min}} + 1 \rceil. \quad (2)$$

Let random variable T_{int} be the inter-arrival time between successive packet arrivals. Note that if packets arrive during the i^{th} sleep window, the BS will schedule a positive traffic indication to the MS in the following listening window. If packets arrive during the listening window, however, we assume that the BS will not immediately respond with a positive indication in the current listening window (consider, for example, the case that packets arrive after the negative indication has been delivered to the MS). Instead, the BS will wake up the MS until the next listening window. The monitoring window for sending traffic indication in the following listening window thus can be defined as shown in



(a) Next packet comes within D_p (b) Next packet comes after D_p



(c) Next packet comes during the n^{st} monitoring window

Figure 3: Delay analysis for the PSC of type I

Figure 2, where each monitoring window is T_L ahead of the sleep stage, and the first monitoring window includes only the initial sleep window. After the MS wakes up, the packet undergoes a fixed service time S . The delay D_n incurred on the packet between its arrival and service completion can then be determined. For simplicity, we only consider the downlink traffic in the following analysis.

To find the packet delay, let N be the number of sleep stages in the sleep mode. The delay D_n experienced by packet n depends on the delay D_p of the previous packet, where three cases can be identified as follows: (a) packet n arrives before the previous packet is completely served, (b) packet n arrives after the previous packet is served, but before the MS enters the sleep mode again, and (c) packet n arrives when the MS is back in the normal mode. Figure 3 illustrates the three cases with relevant parameters.

For case (a), packet n arrives before the previous packet is completely served as shown in Figure 3(a). Denoting this event as $e_{0'}$, the probability that this event happens can be expressed as

$$Pr\{e_{0'}\} = Pr\{T_{int} < D_p\} = F_T(D_p),$$

where $F_T(t)$ is the cumulative density function (CDF) of the inter-arrival time T_{int} . The delay D_n experienced by the packet is equal to

$$D_n|e_{0'} = D_p + S - T_{int}. \quad (3)$$

For case (b), packet n arrives after the previous packet is served, but before the MS enters the sleep mode again as shown in Figure 3(b). Denoting this event as $e_{0''}$, the

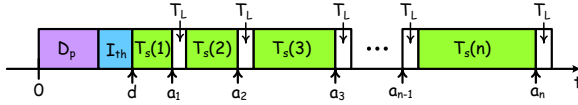


Figure 4: Definition of a_i

probability that this event happens can be expressed as

$$\begin{aligned} Pr\{e_{0''}\} &= Pr\{D_p \leq T_{int} < D_p + I_{th}\} \\ &= F_T(D_p + I_{th}) - F_T(D_p), \end{aligned}$$

and the delay is equal to

$$D_n|_{e_{0''}} = S. \quad (4)$$

For case (c), packet n arrives after the previous has been served and the MS has entered the sleep mode again (the idle threshold I_{th} is passed) as shown in Figure 3(c). Denoting e_i as the event that packet n arrives during the i^{th} monitoring window, we have for $i = 1$

$$\begin{aligned} Pr\{e_1\} &= Pr\{D_p + I_{th} \leq T_{int} < D_p + I_{th} + T_S(1)\} \\ &= F_T(D_p + I_{th} + T_S(1)) - F_T(D_p + I_{th}). \end{aligned}$$

The delayed incurred by packet n is

$$D_n|_{e_1} = D_p + I_{th} + T_S(1) + T_L + S - T_{int},$$

and the number of sleep stages $N = 1$. By induction, for any e_i , $i \geq 2$, we have

$$\begin{aligned} Pr\{e_i\} &= F_T(D_p + I_{th} + (i-1)T_L + \sum_{j=1}^i T_S(j)) \\ &\quad - F_T(D_p + I_{th} + (i-2)T_L + \sum_{j=1}^{i-1} T_S(j)). \end{aligned}$$

The delay incurred by packet n is

$$D_n|_{e_i} = D_p + I_{th} + \sum_{j=1}^i T_S(j) + iT_L + S - T_{int}, \quad (5)$$

and the number of sleep stages $N = i$.

Now define

$$\begin{aligned} a_0 &= d = D_p + I_{th} \\ a_1 &= a_0 + T_S(1) \\ a_i &= a_{i-1} + T_L + T_S(i) \\ &= D_p + I_{th} + (i-1)T_L + \sum_{j=1}^i T_S(j), \quad \forall i \geq 2. \end{aligned} \quad (6)$$

as shown in Figure 4. The mean number of sleep stages can be expressed as

$$\begin{aligned} E[N] &= \sum_{i=0}^{\infty} iPr\{N = i\} \\ &= \sum_{i=1}^{\infty} [1 - F_T(a_{i-1})], \end{aligned} \quad (7)$$

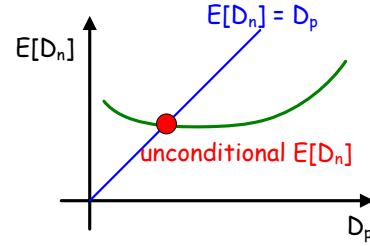


Figure 5: Evaluation of unconditional packet delay $E[D]$

and the mean packet delay can be expressed as

$$\begin{aligned} E[D_n] &= E[D_n|_{e_{0''}}]Pr\{e_{0''}\} + E[D_n|_{e_{0''}}]Pr\{e_{0''}\} \\ &\quad + \sum_{i=1}^{\infty} \{E[D_n|_{e_i}]Pr\{e_i\}\} \\ &= S + D_p + I_{th} - \mu_T - \int_{D_p}^d F_T(t)dt \\ &\quad + \sum_{i=1}^{\infty} [T_S(i) + T_L][1 - F_T(a_{i-1})], \end{aligned} \quad (8)$$

where $\mu_T = E[T_{int}]$ is the mean inter-arrival time.

Therefore, given the delay experienced by one packet D_p , the expected packet delay of the following packet D_n can be obtained using (8) as long as the distribution of packet inter-arrival time $F_T(t)$ is known. Note that (8) is in fact the conditional packet delay $E[D_n] = E[D|D_p]$. To find the unconditional packet delay $E[D]$, we need to take the expected value of $E[D_n]$. Since the expected value of the previous packet delay D_p is equal to the expected value of the next packet delay D_n , we can solve

$$E[E[D_n]] = E[D_p]$$

from (8) to obtain the solution D_p^* as the unconditional packet delay $E[D]$.

While it is difficult to obtain the closed-form expression for $E[D]$ with arbitrary traffic distribution $F_T(t)$, one can still use (8) for investigating the relation between mean packet delay and different power saving parameters as follows. As shown in Figure 5, given the $E[D_n]$ vs. D_p curve based on (8), one can plot the straight line $E[D_n] = D_p$. The cross point of the two curves is the unconditional mean packet delay $E[D]$. We validate the result through simulation in the following section.

4. EVALUATION RESULTS

We have presented in Section 3 an analytical model for finding the mean packet delay in the type I power saving class under general traffic. We show in this section some results for different types of traffic, including the Poisson, Pareto, and 4IPP traffic models. For lack of space, we focus only on the mean packet delay as the performance metric.

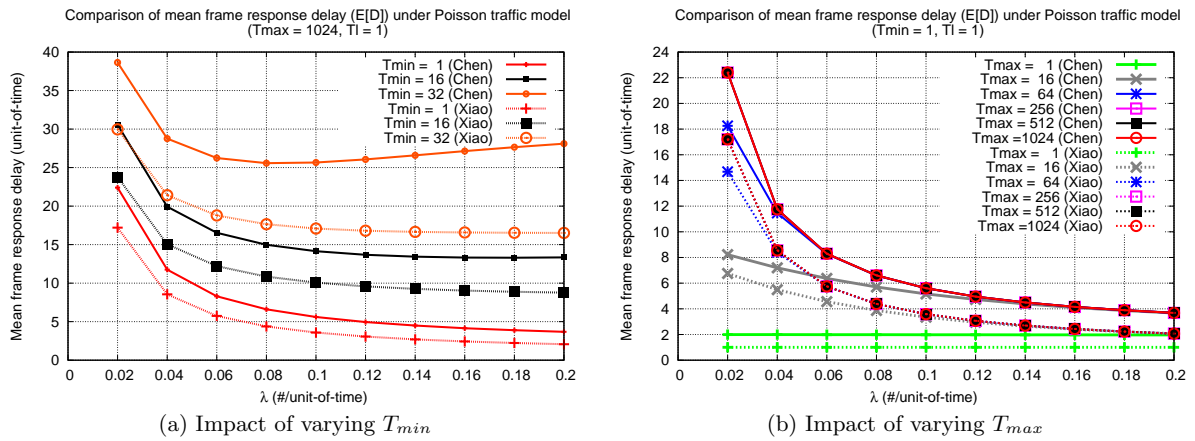


Figure 6: Mean packet delay under the Poisson traffic model

4.1 Poisson Traffic

The Poisson traffic model is used popularly in the literature for analyzing the 802.16 power saving performance. In this section, we compare the proposed model under the Poisson traffic model against the model proposed by Xiao in [14].

To start, we cite the results derived in [14], where the mean packet delay is

$$E[D_n]_{Xiao} = \frac{1}{2} \sum_{j=1}^{\infty} [T_S(j) + T_L] e^{-\lambda \sum_{i=1}^{j-1} [T_S(i) + T_L]} - \frac{1}{2} \sum_{j=1}^{\infty} [T_S(j) + T_L] e^{-\lambda \sum_{i=1}^j [T_S(i) + T_L]} \quad (9)$$

for the Poisson traffic model with rate λ .

In the proposed model under the Poisson traffic model, the CDF of the inter-arrival time is the exponential distribution:

$$F_T(t) = 1 - e^{-\lambda t}, \quad t \geq 0. \quad (10)$$

By substituting (10) into (8), the conditional mean packet delay can be rewritten as

$$E[D_n]_{Chen} = S + D_p - \frac{1}{\lambda} + \frac{1}{\lambda} (e^{-\lambda D_p} - e^{-\lambda d}) + \sum_{i=1}^{\infty} [T_S(i) + T_L] e^{-\lambda a_i - 1}, \quad (11)$$

where the suffix is used to denote the proposed model.

To compare the two models, note first that the model in [14] considers the delay of the first packet that initiates the sleep mode, and hence $D_p = 0$ in addition to setting $I_{th} = 0$ and $S = 0$. In the proposed model, however, we consider continuous runs of the sleep mode operation, and hence the delay experienced by one packet will depend on the previous packet delay, the idle threshold, and the packet service time. The second difference between the two models is that in [14] the first monitoring window includes a *virtual* listening window that corresponds to the time for receiving the MOB_SLP-RSP message from the BS, whereas in the proposed model the first monitoring window includes only the initial sleep window as shown in Figure 2. Finally, since [14]

is proposed for Poisson traffic, it leverages the memoryless property of Poisson traffic for simplifying the model. For example, for packets arriving during the i^{th} monitoring window, the model in [14] uses a delay of $\frac{1}{2} [T_S(i) + T_L]$, assuming that there is only one packet arrival in each sleep stage. In the proposed model, on the other hand, the delay is expressed using (5). The difference in modeling the sleep mode operation results in the delay obtained in [14] to be smaller than the actual delay.

Figure 6 shows the mean packet delay obtained from the model in [14] and the proposed model using (9) and (11) respectively. Figure 6(a) compares the mean packet delay for different power saving parameters T_{min} , while Figure 6(b) compares the mean delay for different values of T_{max} . We find that the delay obtained in [14] is indeed smaller than that obtained in the proposed model as we explained before. Also, note that the difference of the two models decreases with increasing traffic load when $E[N]$ is larger than 1 (e.g. for $T_{min} = 1$ in Figure 6(a)). The reason is that heavy load makes the MS wake up sooner and results in smaller number of sleep stages. The sleep window thus does not grow into a large value, thus rendering a smaller difference between the two models. On the other hand, if packets come during the first sleep window (for which $E[N] = 1$), $E[T_{int} | T_{int} \in [0, a_1]]$ decreases with increasing λ , and hence the difference increases accordingly. Intuitively, as $\lambda \rightarrow \infty$, the packet delay will approach $T_{min} + T_L$. It can be seen from Figure 6(a) and Figure 6(b) that the proposed model observes the desired trend while the model in [14] does not.

Note that for any given values of T_{min} and T_{max} , the mean packet delay $E[D_n]$ decreases with increasing arrival rate λ when $E[N] \gg 1$, and it increases with λ when $E[N] \rightarrow 1$. Intuitively, a larger sleep window results in a larger packet delay. When arrival rate λ varies from 0.02 to 0.2 in Figure 6(a), for example, for curves such as $T_{min} = 1$ or $T_{min} = 16$ (where $E[N]$ is much larger than 1), the delay decreases because the sleep window decreases with larger λ . Otherwise, for curves such as $T_{min} = 32$ and when $\lambda \geq 0.08$ (where $E[N]$ approaches 1), the delay increases with increasing λ because the packets arrive very soon in the first sleep window.

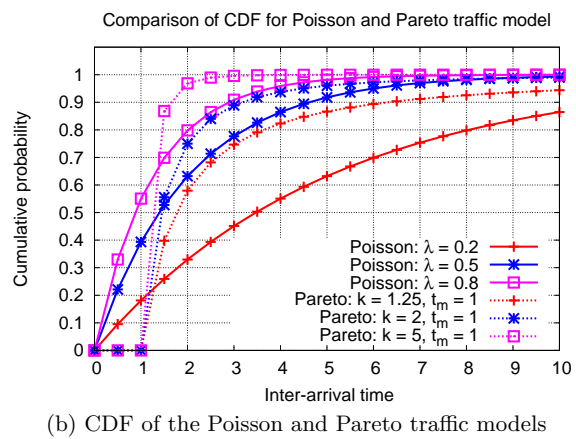
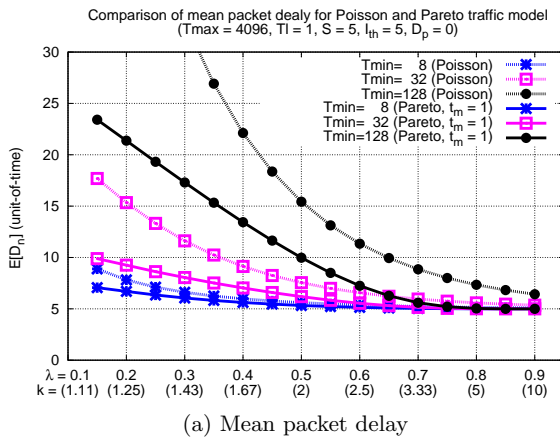


Figure 7: Comparison of the Poisson and Pareto traffic models

4.2 Pareto Traffic

In this section, we consider the operation of the type I power saving class under the Pareto traffic model. We first show the analytical results and then compare the delay performance for Poisson and Pareto traffic models.

The distribution of the Pareto traffic model can be written as

$$F_T(t) = 1 - \left(\frac{t_m}{t}\right)^k, \quad k > 0, \quad t_m > 0, \quad t \geq t_m, \quad (12)$$

where t_m and k are scale and shape parameters respectively. The mean inter-arrival time is

$$\mu_T = \frac{kt_m}{k-1}, \quad k > 1, \quad t_m > 0. \quad (13)$$

By substituting (12) and (13) into (8), we have

$$E[D_n] = S + D_p + I_{th} - \frac{kt_m}{k-1} - W + \sum_{i=1}^{\infty} [T_S(i) + T_L] \left(\frac{t_m}{a_{i-1}}\right)^k, \quad (14)$$

where d and a_i are defined as (6), and W is given depending on the values of t_m as shown in the following:

case 1: $t_m < D_p$

$$W = I_{th} + \frac{t_m^k}{k-1} \left(\frac{1}{d^{k-1}} - \frac{1}{D_p^{k-1}} \right),$$

case 2: $D_p \leq t_m < d$

$$W = D_p + I_{th} - t_m + \frac{t_m^k}{k-1} \left(\frac{1}{d^{k-1}} - \frac{1}{D_p^{k-1}} \right),$$

case 3: $d \leq t_m$

$$W = 0.$$

To investigate the impact of traffic distribution on the performance of the power saving operation, we compare the mean packet delay under the Poisson and Pareto traffic model when the mean inter-arrival time of the two mod-

els are the same. Note that the mean inter-arrival times of the Poisson and Pareto traffic models are $\frac{1}{\lambda}$ and $\frac{kt_m}{k-1}$ respectively. Considering the case with $t_m = 1$, we have the following relation between λ and k for the two models to have the same mean inter-arrival time:

$$k = \frac{1}{1 - \lambda t_m} = \frac{1}{1 - \lambda}.$$

It can be observed from Figure 7(a) that the curve with the Poisson traffic model has larger packet delay than the curve the Pareto traffic model even though the mean inter-arrival time of the two models is the same. To understand why, we show the CDF of the two models for three different traffic loads in Figure 7(b). It can be observed that the cumulative probability for the inter-arrival time being smaller than $t_m = 1$ is larger in the Poisson traffic model than in the Pareto traffic model. On the other hand, the Pareto traffic model practically has a larger cumulative probability for the inter-arrival time being larger than $t_m = 1$, especially when λ is small. Therefore, the Pareto traffic model has more packets that have smaller inter-arrival times. The result is that the sleep mode has a shorter duration under the Pareto traffic model, and hence the Pareto traffic model has smaller packet delay than the Poisson traffic model. The result echoes our claim that traffic distribution has an important impact on the performance of the power saving operation.

4.3 4IPP Traffic

PSC of type I is recommended for non-real-time traffic, and the 4IPP traffic model is proposed in [1] to describe HTTP/FTP traffic for use in the simulation of IEEE 802.16. In this section, we first briefly describe the 4IPP traffic model, and then present the delay performance of the PSC of type I under the 4IPP model.

4.3.1 4IPP traffic model

The 4IPP model is a superposition of four Interrupted Poisson Processes (IPPs) in which each IPP spans a distinct time frame to generate the self-similar traffic found in Internet traffic [1, 7]. The IPP is a special case of the Markov-Modulated Poisson Process (MMPP) [4] with two

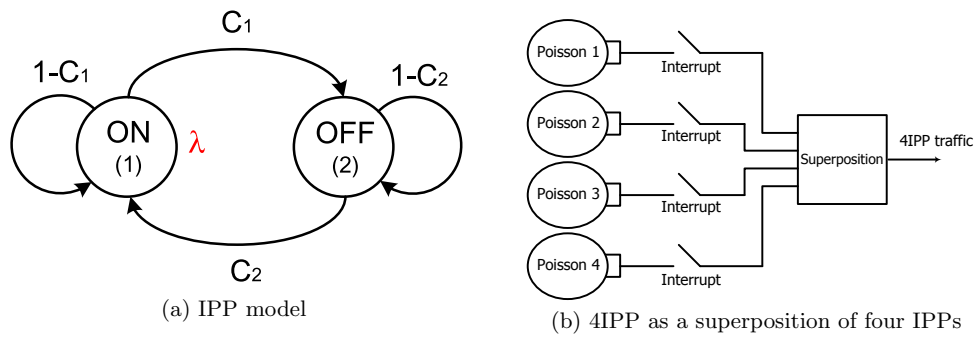


Figure 8: 4IPP traffic model

states as shown in Figure 8(a). During the ON state (state 1), the IPP generates packets following the Poisson process with rate λ , while during the OFF state (state 2), the IPP does not generate any packet. C_1 and C_2 specify the transition probability rate between ON and OFF states, where the mean sojourn time in state 1 (ON time) is $1/C_1$, and that in state 2 (OFF time) is $1/C_2$. Clearly, the Poisson process is a special case of IPP with parameters $C_1 = 0$ and $C_2 = 1$.

The 4IPP model consists of four IPP models with parameters λ_i , C_{i1} , and C_{i2} as shown in Figure 8(b). For an IPP model we define the generator matrix \mathbf{Q} to be the transition matrix of the modulating Markov chain and rate matrix $\mathbf{\Lambda}$ to be the matrix whose diagonal elements contain the arrival intensities. Therefore, for the i^{th} IPP model we have

$$\mathbf{Q}_i = \begin{bmatrix} -C_{i1} & C_{i1} \\ C_{i2} & -C_{i2} \end{bmatrix}, \quad \mathbf{\Lambda}_i = \begin{bmatrix} \lambda_i & 0 \\ 0 & 0 \end{bmatrix}, \quad i = 1, 2, 3, 4.$$

Since the superposition of MMPPs is again an MMPP model, 4IPP can also be characterized by a generator matrix \mathbf{Q} and a rate matrix $\mathbf{\Lambda}$ from individual generator matrices \mathbf{Q}_i and rate matrices $\mathbf{\Lambda}_i$. The matrices \mathbf{Q} and $\mathbf{\Lambda}$ for the 4IPP model can be written as

$$\mathbf{Q} = \mathbf{Q}_1 \oplus \mathbf{Q}_2 \oplus \mathbf{Q}_3 \oplus \mathbf{Q}_4,$$

$$\mathbf{\Lambda} = \mathbf{\Lambda}_1 \oplus \mathbf{\Lambda}_2 \oplus \mathbf{\Lambda}_3 \oplus \mathbf{\Lambda}_4,$$

where \oplus represents the Kronecker-sum.

To generate 4IPP traffic, different set of parameters (λ_i , C_{i1} , and C_{i2}) can be chosen depending on the desired characteristics of the traffic. In [1] the authors provide several parameter sets for describing HTTP/FTP traffic in 802.16 simulation. Given any set of parameters, the CDF and hence the mean inter-arrival time of the 4IPP model can be obtained [2]. Figure 9 shows the CDF curves for 4IPP models with different average traffic rates as given in [1]. In the following simulation, we focus on the 4IPP model with an average traffic rate of 2kbps and mean inter-arrival time of 768.0125 ms.

4.3.2 Unconditional mean packet delay

While we have considered the mean packet delay conditioned on the previous packet delay D_p in previous sections for comparison with related work [14], the proposed model can in fact be used to find the unconditional packet delay

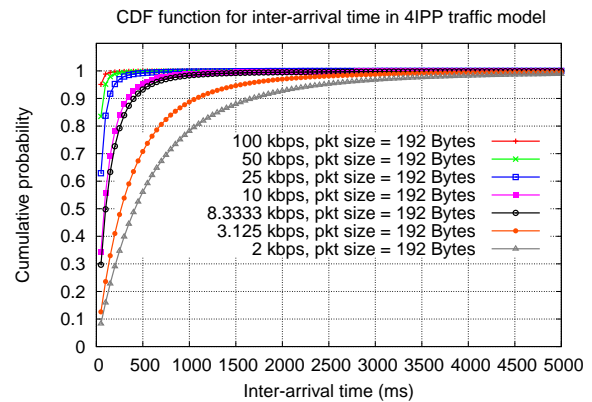


Figure 9: CDF of the 4IPP traffic model

as discussed in Section 3. In this section, we validate the correctness of the proposed model through simulation.

To proceed, the mean packet delay $E[D_n]$ for different values of D_p varying from 0 to 300 ms is obtained as shown in Figure 10. Note that as D_p increases, $E[D_n]$ first decreases and then increases. The reason is that the packet delay is $D_p + S - T_{int}$ in the case of Figure 3(a). The probability of this case is $F_T(D_p)$. If D_p increases, the packet delay $D_p + S - T_{int}$ also increases and the probability of occurrence increases. Therefore, the packet delay in this case gradually dominates the mean packet delay $E[D_n]$ when D_p increases. Eventually the overall mean packet delay increases as D_p increases.

As we discussed in Section 3, in a continuous run of the power saving operations under a train of packet arrivals, the packet delay D_n in the previous run is the previous packet delay D_p of the current run. Thus, the expected value of D_p should be equal to the expected value of D_n . By plotting an auxiliary line of $E[D_n] = D_p$ in Figure 10, we can find the cross points on the auxiliary line and different $E[D_n]$ curves. The D_p values of these cross points are then the unconditional mean packet delay $E[D]$ for different values of T_{min} and T_{max} .

We validate through simulation the correctness of using the D_p values of the cross points in Figure 10 as the un-

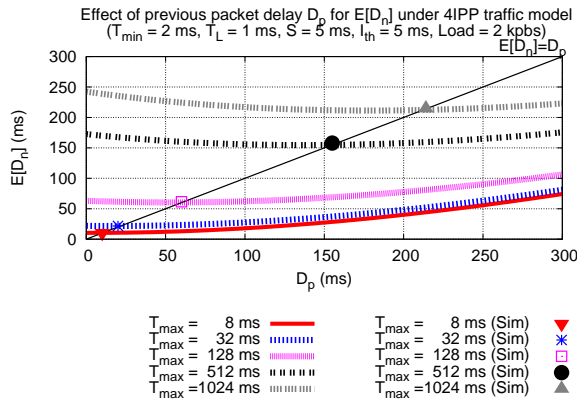


Figure 10: Finding unconditional packet delay

conditional mean packet delays. Unlike for the case of per-packet simulations where the previous packet delay D_p is given for each operation, in the continuous run packets are generated and processed one by one based on the operation of the PSC of type I. Therefore, the packet delay D_n of packet i will become the previous packet delay D_p of packet $i+1$. Given a power saving parameter pair (T_{min}, T_{max}) , the continuous run of the power saving operation will eventually yield the unconditional mean packet delay as labeled (Sim) in Figure 10. Exact values of the packet delay obtained through (a) crossing the $E[D_n]$ and $E[D_n] = D_p$ curves and (b) performing the continuous-run simulations are listed in Table 1. It can be observed that the two results agree with each other, thus validating the model that the unconditional mean packet delay can be obtained through the conditional mean packet delay following a straightforward approach.

4.3.3 Tradeoff between packet delay and energy consumption

In this section, we investigate the performance of the type I power saving class under the 4IPP model when T_{min} and T_{max} vary. For simplicity we consider the traffic with 2kbps traffic load, and set $D_p = 0$, listening window size $T_L = 1$, idle threshold $I_{th} = 5$, and service time $S = 20$.

Figure 11(a) and Figure 11(b) show the mean packet delay and number of sleep stages as T_{max} varies from 1 to 4096 while T_{min} is kept at a fixed value for each curve. The curves labeled **Num** denote the numerical results, and the dots labeled **Sim** denote the simulation results. Note that T_{max} should be larger than or equal to T_{min} for proper power saving operation, and hence all curves start from $T_{max} = T_{min}$ for different values of T_{min} . We observe from Figure 11 that there is a tradeoff between $E[D_n]$ and $E[N]$. Since N represents how many number of sleep stages the MS can sleep between successive arrivals, $E[N]$ is related to the energy consumption of the MS when the MS is in the sleep mode. Therefore, Figure 11 shows the tradeoff between packet delay and energy consumption: for any given value of T_{min} , as T_{max} increases, packet delay increases while energy consumption goes down. Similarly, for any given value of T_{max} , as T_{min} increases, packet delay increases while energy consumption goes down.

We note that in Figure 11, as T_{max} becomes very large,

Table 1: Verification of $E[D]$

(T_{min}, T_{max})	Numerical results	Simulation results
(2, 8)	10.4 ms	10.4 ms
(2, 32)	21.7 ms	21.7 ms
(2, 128)	60.4 ms	61.4 ms
(2, 512)	154.7 ms	157.8 ms
(2, 1024)	212.8 ms	214.6 ms

$E[D_n]$ and $E[N]$ saturate. The saturated points for different T_{min} curves can be obtained as follows. Let $x \geq 0$ be an arbitrarily small value, and let the term of $i = \Delta_N$ in (8) be smaller than x . We have

$$(2^{\Delta_N - 1} T_{min} + T_L)$$

$$* [1 - F_T(d + (\Delta_N - 2)T_L + (2^{\Delta_N - 1} - 1)T_{min})] \leq x \cong 0,$$

or equivalently,

$$F_T(d + \Delta_N T_L - 2T_L + 2^{\Delta_N - 1} T_{min} - T_{min}) \geq 1 - \frac{x}{2^{\Delta_N - 1} T_{min} + T_L}. \quad (15)$$

If Δ_N^* is the smallest Δ_N that satisfies (15), then $T_{max}^* = 2^{\Delta_N^* - 1} T_{min}$ is the saturated point for $E[D_n]$. That is, increasing the final window size T_{max} beyond T_{max}^* practically does not increase packet delay $E[D_n]$.

Similarly, the saturation point for $E[N]$ can be found through manipulating the $i = \Delta_N$ term in (7). Let

$$1 - F_T(d + (\Delta_N - 2)T_L + (2^{\Delta_N - 1} - 1)T_{min}) \leq y \cong 0,$$

or equivalently,

$$F_T(d + \Delta_N T_L - 2T_L + 2^{\Delta_N - 1} T_{min} - T_{min}) \geq 1 - y \quad (16)$$

for any arbitrarily small value of $y \geq 0$. If Δ_N^\dagger is the smallest Δ_N that satisfies (16), then $T_{max}^\dagger = 2^{\Delta_N^\dagger - 1} T_{min}$ is the saturated point for $E[N]$. That is, increasing the final window size T_{max} beyond T_{max}^\dagger practically does not increase packet delay $E[N]$. Due to lack of space, we do not delve into details of this issue.

5. CONCLUSIONS

In this paper, we propose an analytic model for capturing the behaviors of IEEE 802.16 sleep mode operations under the type I power saving class with arbitrary traffic distribution. The proposed model can be used for understanding the performance of the power saving operations in IEEE 802.16 under general traffic, including Poisson, Pareto and 4IPP traffic models. The proposed model analyzes the power saving operation under arrivals of multiple packets as opposed to considering only one packet in related work. In addition, the proposed model incorporates parameters such as packet service time and idle threshold not modeled in related work. We verify the model using both numerical results and simulation results, and compare the performance of the proposed model against existing work. Finally, we explain how such a general model allows a better investigation of the impact of different traffic models on the performance of IEEE 802.16 power saving operations.

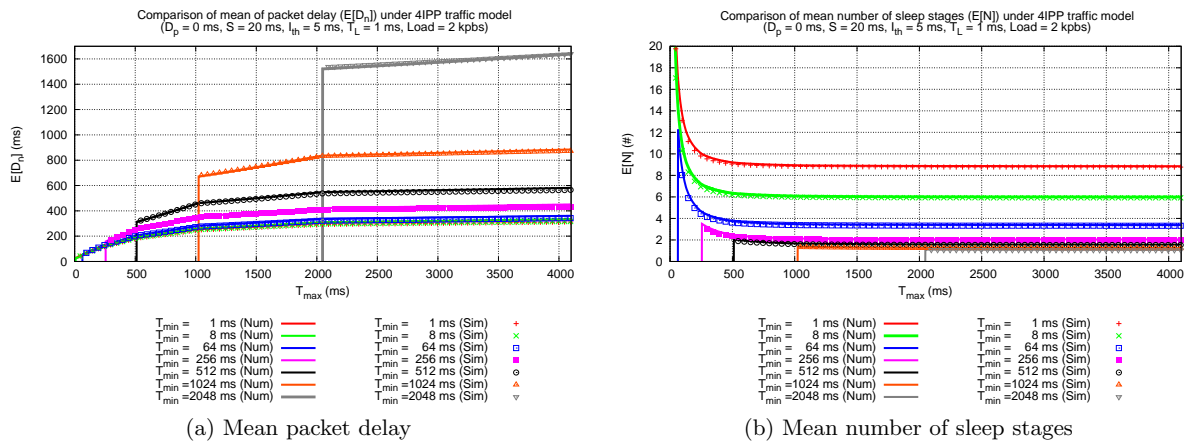


Figure 11: Performance of the 4IPP traffic model

6. ACKNOWLEDGMENTS

This work was supported in part by funds from the Excellent Research Projects of the National Taiwan University under Grant 97R0062-06 and the Wireless Broadband Communications Technology and Application Project of the Institute for Information Industry subsidized by ROC Ministry of Economy Affairs.

7. REFERENCES

- [1] C. Baugh, J. Huang, R. Schwartz, and D. Trinkwon. Traffic Model for 802.16 TG3 MAC/PHY Simulations. IEEE 802.16.3c-01/30r1, March 2001.
- [2] M.-F. Chen. Modeling, Analysis, and Design of Adaptive Power Saving Algorithms for 802.16 Sleep Mode Operations. National Taiwan University Master Thesis, July 2008.
- [3] G. Dong, C. Zheng, H. Zhang, and J. Dai. Power Saving Class I Sleep Mode in IEEE 802.16e System. In *Proceedings of International Conference on Advanced Communication Technology*, pages 1487–1491, February 2006.
- [4] W. Fischer and K. Meier-Hellstern. The Markov-Modulated Poisson Process (MMPP) Cookbook. *Performance Evaluation*, 18(2):149–171, September 1993.
- [5] K. Han and S. Choi. Performance Analysis of Sleep Mode Operation in IEEE 802.16e Mobile Broadband Wireless Access Systems. In *Proceedings of Vehicular Technology Conference*, pages 1141–1145, May 2006.
- [6] Y.-H. Han, S.-G. Min, and D. Jeong. Performance Comparison of Sleep Mode Operations in IEEE 802.16e Terminals. In *Proceedings of ICCS 2007*, pages 441–448, July 2007.
- [7] J. Huang. Generalizing 4IPP Traffic Model for IEEE 802.16.3. IEEE 802.16.3c-00/58, December 2000.
- [8] IEEE Std 802.16e-2005. *Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*. February 2006.
- [9] M.-G. Kim, M. Kang, and J. Y. Choi. Performance Evaluation of the Sleep Mode Operation in the IEEE 802.16e MAC. In *Proceedings of International Conference on Advanced Communication Technology*, pages 602–605, February 2007.
- [10] L. Kong and D. H. Tsang. Performance Study of Power Saving Classes of Type I and II in IEEE 802.16e. In *Proceedings of IEEE Conference on Local Computer Networks (LCN)*, pages 20–27, November 2006.
- [11] Z. Niu, Y. Zhu, and V. Benetis. A Phase-Type Based Markov Chain Model for IEEE 802.16e Sleep Mode and its Performance Analysis. In *Proceedings of ITC 2007*, pages 791–802, 2007.
- [12] J.-B. Seo, S.-Q. Lee, N.-H. Park, H.-W. Lee, and C.-H. Cho. Performance Analysis of Sleep Mode Operation in IEEE 802.16e. In *Proceedings of Vehicular Technology Conference*, pages 1169–1173, September 2004.
- [13] J. Xiao, S. Zou, B. Ren, and S. Cheng. An Enhanced Energy Saving Mechanism in IEEE 802.16e. In *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM)*, pages 1–5, November 2006.
- [14] Y. Xiao. Energy Saving Mechanism in the IEEE 802.16e Wireless MAN. *IEEE Communication Letters*, 9(7):595–597, July 2005.
- [15] Y. Xiao. Performance Analysis of an Energy Saving Mechanism in the IEEE 802.16e Wireless MAN. In *Proceedings of Consumer Communications and Networking Conference (CCNC)*, pages 406–410, January 2006.
- [16] Y. Zhang and M. Fujise. Energy Management in the IEEE 802.16e MAC. *IEEE Communication Letters*, 10(4):311–313, April 2006.
- [17] S. Zhu and T. Wang. Enhanced Power Efficient Sleep Mode Operation for IEEE 802.16e Based WiMAX. In *Proceedings of IEEE Mobile WiMAX Symposium*, pages 43–47, March 2007.