Performance Analysis of Sleep Mode Operation 
In IEEE802.16e

June-Bae Seo, Seung-Que Lee, Nam-Hoon Park  
Department of Wireless System Research  
ETRI  
Deajeon, Korea  
{jbseo, sqlee, nhpark}@etri.re.kr

Hyong-Woo Lee, Choong-Ho Cho  
Depart. Electronics and Information Engineering  
Depart. Computer Science  
Chungnam, Korea  
{hwlee, chcho}@korea.ac.kr

I. INTRODUCTION

As a promising solution of a fixed Broadband Wireless Access Systems(BWA), recently IEEE 802.16 wireless MAN has been standardized. It provides network access from buildings through external antennas communicating with central radio base stations. Due to wireless medium covering broad geographic areas without expensive installation cost, this may offer ubiquitous broadband access[1]-[3]. In this standard, a common MAC protocol is used with physical layer specifications dependent on the spectrum of use and the associated regulations. Additionally, IEEE802.16e is a standard in progress for BWA in order to support mobility, which may fill the gap between fixed wireless local area networks and mobile cellular systems. In [2], the sleep mode operation to save the power of an MS is proposed. In the scenario, an MS has two modes, such as, awake- and sleep-mode. Before entering the sleep mode, an MS in the awake mode sends the sleep request message to a serving BS. After receiving the sleep response message which notifies whether the sleep request of the MS is approved or not, an MS can determine to enter the sleep mode. In the sleep response message, some parameters, such as, the start time of sleep mode, the minimum, $T_0$, and the maximum, $T_{bmax}$, of sleep interval and the listening interval are presented as units of MAC frames. As the first sleep interval, the minimum of sleep interval, $T_0$, is used and the consecutive sleep intervals which will be denoted by $T_i$, is equal to $2^{i} \cdot T_0$, until $T_i$ reaches $T_{bmax}$. When an MS enters the next sleep over $T_{bmax}$, the maximum of sleep interval is repeatedly used, until it will awaken. During the listening interval, $T_L$, at each end of sleep interval, an MS examines the traffic indication message broadcasted, for confirming whether traffic is addressed to the MS or not. If no traffic is addressed to the MS, this MS will return to the sleep mode again with the sleep timer described above. In addition, an MS can terminate its sleep mode any time. During each sleep interval, an MS can turn off its transmitter- and receiver-unit in order to conserve the power and a serving BS buffers incoming traffic, i.e., packet Protocol Data Unit(PDUs), in order to convey them at the end of the sleep interval. From practical viewpoint, it is reasonable because a connection is alive. For more detailed operation of the sleep mode, refer to [2]. In the literature, one can find simulation study[4] and analytical work using a supplementary variable technique[5] for the sleep mode operation of Cellular Digital Packet Data (CDPD). In contrast, we model the sleep mode operation of IEEE802.16e as $M/GI/1/N$ queueing system with multiple vacations whose periods depend on the previous one, by constructing Hessenberg matrix representing a $M/G/1$ queueing system[6][7]. In this work, we examined the dropping probability and mean waiting times of the arriving packet PDUs in the queue of a serving BS, while an MS is sleeping and how long an MS can turn off its transmitter and receiver units. This work is organized as follows. In section II, details of the analysis is given. Numerical examples is discussed in section III. Finally, concluding remarks are given in section IV.

II. SYSTEM ANALYSIS

The system of interest is assumed to be a queue of size $C$ in a BS, of which is assigned each MS. We assume that the listening interval, $T_L$, is only used for sending the traffic indication message to MSs and the purpose of synchronization between MSs and BS. Hence, when there is no traffic addressed to an MS during the sleep interval and packets arrive
in the listening interval immediately following the sleep interval, these packets are buffered and the traffic indication message for these packets will be notified at the next listening interval. Accordingly, the transmission of traffic arrived during the previous sleep interval begins at the end of the listening interval. It is assumed that the times for negotiating the sleep mode between a BS and MSs, i.e., the time to exchange the sleep request and sleep response messages, is zero. We assume that the packet arrival process from network to a BS follows a Poisson process with rate, \( \lambda \). This may oversimplify dynamics of queueing behaviors at a BS. However, Poisson arrival process may be a basis for measuring a loose upper bound of a queueing behavior.

The service time of a packet is assumed to be a general system before introducing a correlated arrival process into a system. The service time of a packet is assumed to be a sum of a sleep interval and a listening interval. It is assumed that the times for negotiating the sleep mode between a BS and MSs, i.e., the time to exchange the sleep request and sleep response messages, is zero. Each state probability is expressed as

\[
\phi_i = \phi_0 \prod_{j=0}^{i-1} \alpha_j, \quad 0 \leq i \leq k_{\text{max}}
\]

where \( \phi_0 \) is the probability vector of \( \phi_i \), \( 0 \leq i \leq k_{\text{max}} \).

Each state probability is expressed as

\[
\phi_j = \phi_0 \prod_{i=0}^{j-1} \alpha_i, \quad 1 \leq j \leq k_{\text{max}} - 1
\]

Using the normalized condition \( \sum_{i=0}^{k_{\text{max}}} \phi_i = 1 \), \( \phi_0 \) is obtained by

\[
\phi_0 = \frac{(1 - \alpha_{k_{\text{max}}})}{(1 - \alpha_{k_{\text{max}}})} \cdot \left[ 1 + \sum_{j=0}^{k_{\text{max}}-1} \prod_{i=0}^{j-1} \alpha_i + \prod_{i=0}^{k_{\text{max}}-1} \alpha_i \right]
\]

The expected remaining vacation period given that the present state is \( S_i \), will be denoted by \( \bar{V_i} \) and the expected remaining vacation period of the initial state, \( E[V] = \bar{V_0} \), is obtained by the following recursion.

\[
\bar{V}_0 = (1 - \alpha_0) \cdot \bar{V}_0 + \alpha_0 \cdot (\bar{V}_1 + \bar{V}_0)
\]

\[
\bar{V}_1 = \bar{V}_1 + \alpha_1 \cdot \bar{V}_2
\]

\[
\bar{V}_{k_{\text{max}}-1} = \bar{V}_{k_{\text{max}}-1} + \alpha_{k_{\text{max}}-1} \cdot \bar{V}_{k_{\text{max}}}
\]

\[
\bar{V}_{k_{\text{max}}} = \bar{V}_{k_{\text{max}}-1} + \alpha_{k_{\text{max}}} \cdot \bar{V}_{k_{\text{max}}}
\]

Finally, one can readily obtain

\[
E[V] = \sum_{i=0}^{k_{\text{max}}-1} \left[ \bar{V}_i - \prod_{j=0}^{i-1} \alpha_j + \frac{V_{k_{\text{max}}}}{1 - \alpha_{k_{\text{max}}}} \cdot \prod_{i=0}^{k_{\text{max}}-1} \alpha_i \right]
\]

Let \( \phi_i \) and \( P_j \) be the steady state probability that the server takes a vacation of \( V_i \), i.e., the server is in the state of \( S_i \), and the transition probability matrix of the embedded Markov chain in Fig.2, respectively. Then, \( \phi_i \) can be obtained by

\[
\bar{\phi} = \bar{\phi} \cdot P_j
\]
where \( j! = j \cdot (j-1) \cdots 2 \cdot 1 \).

Hence, the probability that \( j \) packets arrive in the queue during the server is on vacation is given by

\[
h_j = \sum_{i=0}^{k_{\text{max}}} \Pr[a=j|S=S_i] \cdot \Pr[S=S_i], \quad 0 \leq j \leq \infty
\]

\[
= \sum_{i=0}^{k_{\text{max}}} \left( \lambda V_i \right)^j \frac{e^{-\lambda V_i \phi_i}}{j!}
\]

(9)

where \( \phi_i \) is given in (4).

The equation (9) expresses the sum of the probabilities that \( j \) packets arrive during each vacation of \( V_i \), while the server is on the vacation, \( S_i \), over all \( i \). Without regard to the vacation interval, \( h_j \), means the probability that \( j \) packets arrive in the queue at an arbitrary time in steady state, while a server is on vacation. So far, we obtained the probability that \( j \) packets arrive in the queue, while the server is on vacation.

The probability that \( j \) packets arrive in a packet service times is given by

\[
g_j = \int_0^\infty \frac{(\lambda \cdot t)^j}{j} e^{-\lambda t} \cdot b(t) \, dt
\]

(10)

where \( b(t) \) is the probability density function (PDF) of a packet service time.

With these, \( g_j \) and \( h_j \), we obtain the queue length distribution in an arbitrary time in steady state described in [4][5]. Provided that \( j \) packets arrive and are accepted while the server is on vacation, these packets will be served at the end of the listening interval immediately following the previous vacation interval. During this listening interval, packets can arrive. Therefore, the probability that \( i \) packets arrive during the listening interval, when \( j \) packets arrived during the previous vacation interval, is expressed as

\[
\psi_i = \frac{1}{1-h_0} \sum_{j=0}^{i} h_j \cdot \tilde{a}_{i-j}, \quad 1 \leq i \leq C-1
\]

\[
\psi_C = \frac{1}{1-h_0} \sum_{j=0}^{i} h_j \cdot \tilde{a}_{i-j}
\]

\[
= \frac{1}{1-h_0} \left( 1 - \sum_{j=1}^i \sum_{j=1}^i h_j \cdot \tilde{a}_{i-j} \right)
\]

(11)

where the probability that \( i \) packets arrive during the listening interval, \( \tilde{a}_{i} \), is expressed as

\[
\tilde{a}_i = \frac{(\lambda T_L)^i}{i!} e^{-\lambda T_L}
\]

(12)

The steady state probability that \( j \) packets are left in the BS at a service completion epoch, namely, a departure epoch, of a packets is denoted by \( \pi_j \). Then, \( \pi_j \) satisfy the following equations:

\[
\pi_j = \pi_0 \sum_{i=0}^{j+1} \varphi_i \cdot g_{j-i+1} + \sum_{i=0}^{j} \pi_i \cdot g_{j-i+1}, \quad 0 \leq j \leq C-2
\]

\[
\pi_{C-1} = \pi_0 \sum_{i=0}^{C-1} \varphi_i \cdot g_{C-i} + \sum_{i=0}^{C-1} \pi_i \cdot g_{C-i}
\]

(13)

where \( g_k = \sum_{i=k}^{\infty} g_i \) and the normalizing condition \( \sum_{i=0}^{C-1} \pi_i = 1 \).

With some modifications on (20)~(21) in [6], the probabilities that there are \( j \) packets in the system at an arbitrary time in steady state, \( \pi_j \), are given by

\[
\pi_j = \frac{\pi_j \lambda}{E[V] \pi_0 + E[B]}, \quad 0 \leq j \leq C-1
\]

\[
\pi_C = 1 - \frac{\lambda}{E[V] \pi_0 + E[B]}
\]

(14)

where \( E[B] \) is the mean of a packet service time.

The dropping probability of packets in the queue, \( P_d \), and the average waiting times of packets, \( E[W] \), are respectively obtained by

\[
P_d = \frac{\pi_C^*}{\bar{N}} \left( \frac{\pi_C^*}{\lambda (1-\pi_C^*)} - E[B] \right)
\]

(15)

where \( \bar{N} \) is the average number of packets in the queue expressed as

\[
\bar{N} = \sum_{i=0}^{C} i \cdot \pi_i^*
\]

(16)

The mean delay period, \( E[D] \), which consists of an initial busy period produced by packets arriving in a vacation and subsequent busy periods generated by initial and following successive busy periods can be derived by the followings.

\[
1 - \rho = \frac{E[V]}{E[V] + E[D]}
\]

\[
\rho = \lambda \cdot (1 - \pi_C^*) E[B]
\]
The equations (17) and (18) mean the probability that an MS is in a vacation and the throughput of this system, respectively. Accordingly, substituting (18) into (17), one can obtain the mean delay period expressed as

$$E[D] = \frac{\lambda (1 - \pi_c^*) E[B] E[V]}{1 - \lambda (1 - \pi_c^*) E[B]}$$  \hspace{1cm} (19)$$

In order to examine the percentage of turn-off times of an MS, we define this as

$$\text{Turn off} (\%) = \frac{E[V^*]}{E[D] + E[V]} \times 100$$  \hspace{1cm} (20)$$

where the parameter, $E[V^*]$, is the expected vacation interval without the listening interval, which is expressed as

$$E[V^*] = T_0 + \sum_{j=1}^{k_{\max} - 1} \left[ \tilde{T}_j \cdot \alpha_0 \cdot \prod_{i=1}^{j-1} \tilde{\alpha}_i \right] + \frac{\tilde{T}_{k_{\max}}}{1 - \tilde{\alpha}_{k_{\max}}} \cdot \alpha_0 \cdot \prod_{i=1}^{k_{\max}-1} \tilde{\alpha}_i \hspace{1cm} (21)$$

where $\tilde{T}_j = 2^j \cdot T_0$, $\tilde{\alpha}_i = e^{-\lambda \tilde{T}_i}$ for $1 \leq i \leq k_{\max}$.

III. NUMERICAL EXAMPLE

The parameters, such as, queue size and the mean packet service times are purposely chosen in order to noticeably visualize the dropping probability and the mean waiting time of packets and the percentage of turn-off times of an MS. In the followings, the mean packet service times is 40 msec, regardless of the service time distributions, and the units of arrival rate is packets per seconds. In Figs.3–6 and 9, the exponential service time distribution with the mean of 40 msec is used except Figs.7–8 and 10 where Erlang, exponential and hyper-exponential distributions are used for the service time distribution.

The effects of queue size : In Figs.3–4, the parameters, the larger queue size, the lower dropping probability and the higher waiting times are observed. Due to finite capacity, the dropping probability and mean waiting times reach a limit over the arrival rate of 50 packets/sec.

The effects of the maximum sleep interval : In Fig.5, in low arrival rate region, the larger the $k_{\max}$, the larger mean waiting times are observed. This means that an MS can frequently enter the sleep-mode with longer sleep interval in low arrival rate region, when a large $k_{\max}$ is given.

The effects of the minimum sleep interval : In Fig.6, as similar effects of $k_{\max}$, $T_0$ shows larger mean waiting times in low arrival rate. However, degree of vacancy in the mean waiting times by various $T_0$ is not higher, compared to those by $k_{\max}$. This results in binary exponential backoff behaviors of $k_{\max}$.

The effects of service time distribution with same mean and different variance : In Figs.7–8, a hyper-exponential and 3-stage Erlangian service time distribution are used as [7]. The mean service times of each distribution are equal to 40 msec and only variances are different. Hyper-exponential distribution shows the highest dropping probability and the largest waiting times. Over the arrival rate of about 25 packets/sec, 3-stage Erlangian distribution shows the largest mean waiting times. This means that the queue may become more easily empty, when the variance of a service time distribution is large. Accordingly, an MS can enter the sleep mode with short period sleep timer. When the minimum sleep interval, $T_0$, is large, the mean waiting times by these distributions over high arrival rate region converge to a value, because a larger $T_0$ reduces likelihood for an MS to enter the sleep mode, even with high variability of service times distribution. However, a service time distribution with large variance always shows large dropping probability.

In Figs.9–10, the percentage of turn-off times are examined. In Fig.9, a BS with small queue size can have an MS slept more frequently in high arrival rate region. However, this phenomena results in that a BS drops incoming packets due to small capacity and small queue size become easily empty, which makes an MS enter the sleep mode. In Fig.10, three different service time distributions are used.

IV. CONCLUSION

In this work, we examined the sleep mode operation in IEEE802.16e in terms of the dropping probability and the mean waiting times of packets in a BS’s buffer by assuming that packets arrival process follows Poisson process and that service times for a packet is a general distribution. In low arrival rate region, the large mean waiting times are observed, when high order binary exponential backoff is used. This results from that an MS tends to enter the sleep mode with high order backoff stage. In high arrival rate region, capacity of buffer should be large enough to provide low dropping probability, while a MS is in the sleep mode. Although the binary exponential backoff scheme to increase each sleep interval may be adequate for a correlated arrival process, our work may be a rough guideline to understand queueing behavior of a BS and to assign a partitioned capacity to an MS, while an MS is in the sleep mode.

REFERENCES


Figure 3. Packet Dropping probability with capacity = 5, 10, 15 and $T_0 = 5$ msec, $T_L = 5$ msec and $k_{max} = 5$.

Figure 4. Mean Waiting time with capacity = 5, 10, 15 and $T_0 = 5$ msec, $T_L = 5$ msec and $k_{max} = 5$.

Figure 5. The mean waiting times with $k_{max} = 3, 5, 7, 8$ and queue size = 10, $T_0 = 5$ msec and $T_L = 5$ msec.

Figure 6. The mean waiting times with $T_0 = 5$, 10, 15 and 20 msec, queue size = 10 and $T_L = 5$ msec.

Figure 7. Packet dropping probability with service time distributions of the same mean and different variances(Each of them, Erlang : 1.1 msec, Exp. : 1.6 msec, Hyper-Exp : 2.4 msec), queue size = 10, $T_0 = 5$ msec, $T_L = 5$ msec and $k_{max} = 7$.

Figure 8. The mean waiting times with service time distributions of the same mean and different variances with same parameters in Fig. 7.

Figure 9. Percentage of Turn off times with capacity = 5, 10, 15, $T_0 = 5$ msec, $T_L = 5$ msec and $k_{max} = 7$.

Figure 10. Percentage of Turn off times with different service time distributions of the same mean and different variances, queue size = 10, $T_0 = 5$ msec, $T_L = 5$ msec and $k_{max} = 7$. 