Queueing Delay and Energy Efficiency Analyses of Sleep Based Power Saving Mechanism

Fan ZHU†, Yiqun WU†, Nonmembers, and Zhisheng NIU†, Fellow

SUMMARY In wireless networks, sleep mode based power saving mechanisms can reduce the energy consumption at the expense of additional packet delay. This letter analyzes its packet queueing delay and wireless terminals’ energy efficiency. Based on the analysis, optimal sleep window size can be derived to optimize terminal energy efficiency with delay constraint.

key words: sleep mode, queueing delay, energy efficiency, sleep window size selection

1. Introduction

In wireless networks, sleep mode based power saving mechanisms have been adopted to save mobile terminals’ energy. In the sleep mode operation, optimizing the power saving performance while satisfying the requirement of Quality of Service, especially the delay requirement, is an important problem. To achieve the optimal solution, packets’ average queueing delay and terminal’s energy efficiency, i.e., throughput per unit energy consumption, should be analyzed first.

When a terminal is in sleep window, the base station (BS) buffers the downlink packets addressing to the terminal. After the sleep window, the terminal wakes up and detects the existence of buffered packets during the listening window. If packets exist, the terminal stays awake until all buffered packets addressed to it are received. Otherwise, it starts the next sleep window. In 802.16e networks [1], the sleep window increases exponentially up to the maximum window.

Figure 1 shows a cycle of sleep and awake durations for a terminal under the power saving mechanism. Time $T_j$ represents the $j$th sleep window and $T_l$ represents the listening window. It is assumed that the first packet in this sleep-award cycle arrives during the $j$th sleep window. All packets in the sleep-award cycle can be divided by their arrival time into two categories. Packets of category A arrive during the sleep duration and packets of category B arrive during the awake duration. The square and triangle in Fig. 1 represent packets’ arrival and served time respectively. For a packet of category A, its queueing delay $d_A^j$ is the sum of two parts: the delay in the sleep duration, or the “sleep delay” $d_s^j$, and the delay in the awake duration. While the terminal is being served in the awake duration, the terminal also has packets of category B destined to it arriving at the BS and added into the buffer. Before ending the awake duration, the terminal is supposed to have received all the packets buffered at the BS. For a packet of category B, there is no sleep delay.

Focusing on the terminal’s sleep duration, average packet sleep delay $d_A^j$ and energy consumption in sleep duration were investigated in [2], [3] and [4]. By considering both sleep and awake duration, the queueing delay of category A packets was analyzed in [5] and [6]. But the arrival of category B packets was not considered. References [7] and [8] analyzed the average queueing delay taking both category A and B packets into consideration. Reference [7] assumed that all the packets can be served in a listening window $T_l$ and [8] assumed that just a single packet is served during a unit time, or the service time follows the deterministic distribution with mean value 1. By assuming that the service time follows general distribution, [9] proposed an M/GI/1/K queueing system with multiple vacations to analyze the energy consumption based on analyzing the time of awake duration. But the average queueing delay was not investigated in [9].

In this letter, we consider both category A and category B packets arriving in a repeated cycle of sleep and awake periods, and analyze average packet queueing delay by assuming a generally distributed service time. Then, we analyze the terminal’s energy efficiency in the sleep-award cycle. Based on the analysis, proper sleep window size could be derived to optimize terminal energy efficiency with delay constraint.

2. Analytical Model

2.1 Packet Queueing Delay

The assumptions in the model are listed as follows: 1) there is only downlink traffic; 2) the arrival interval follows the exponential distribution with mean value $1/\lambda$; 3) the service time follows the general distribution with mean value $1/\mu$ and variance coefficient $c_B^2$.

Let $d$ and $d'$ denote the packet queueing delay and the packet sleep delay respectively, and let $E[d]$ and $E[d']$ denote the corresponding expected value. Let $p_j$ denote the
the additional sleep delay window. For all the 
and probability of station’s waking up after the 
jth sleep window. 
conditioned that the terminal wakes up after the 
jth sleep window.
For all the 
probability of station’s waking up after the 
jth sleep window.

Theorem 2.1.2 in [10], \(w_j\) can be decomposed into the sum of two values: the delay 
arriving in the sleep and awake duration respectively, conditioned that the terminal wakes up after the 
jth sleep window.

Let \(T_j\) denote the time of 
jth sleep phase. In IEEE 802.16e networks, let \(l\) denote the maximum exponent of 
sleep window, and \(t_j\) can be expressed as

\[
t_j = \begin{cases} 
T_12^{j-1} + T_i, & j \leq l; \\
T_12^{j-l} + T_i, & j \geq l + 1 
\end{cases}
\]

in which \(T_1\) is the initial sleep window and \(T_i\) is the listening window.

Let \(S_j^a\) denote the expected time of awake duration, conditioned that the terminal wakes up after the 
jth sleep window. From the knowledge of queueing theory, \(S_j^a\) is the mean time of “\(k – busy\) period” with \(k\) equals to \(m_j\) [11].

Therefore, \(S_j^a\) is given by \(S_j^a = m_j^a\).

The number of packets arriving in \(S_j^a\) is

\[
m_j^a = \lambda S_j^a.
\]

For an M/G/1 queue, the average delay \(w_{j,1}\) is

\[
w_{j,1} = \frac{2\mu + (c_0^2 - 1)\lambda}{2\mu(\mu - \lambda)}. \tag{6}
\]

Since the arrival process is a stationary process, we have the average sleep delay of packets arriving in \(t_j\) as

\[
w_{j,2} = \frac{t_j}{2}. \tag{7}
\]

The value of \(p_j\) is the probability of there being no packet arriving during the \(1_{st}, 2_{nd}, ..., (j-1)_{th}\) sleep phase while there existing packets arriving during the \(j_{th}\) sleep phase. In IEEE 802.16e networks, when \(j \leq l\), we have \(p_j\) as

\[
p_j = e^{-\lambda[(2^{j-1} - 1)T_1 + (j-1)T_i]}(1 - e^{-\lambda[(2^{j-1} - 1)T_1 + T_i]}); \tag{8}
\]

when \(j \geq l + 1\), we have \(p_j\) as

\[
p_j = e^{-\lambda[(2^{l-1} - 1)T_1 + (j-l-1)2^{l-1} + (j-1)T_i]}(1 - e^{-\lambda[(2^{l-1} - 1)T_1 + T_i]}). \tag{9}
\]

By substituting Eqs. (3), (5), (6), (7) into (1), we have

\[
E[d] = \frac{\sum_{j=1}^{\infty} \frac{p_j T_j}{1-e^{-\mu T_j}}}{\sum_{j=1}^{\infty} \frac{p_j T_j^2}{1-e^{-\mu T_j}}}. \tag{10}
\]

By substituting Eqs. (3) and (7) into Eq. (2), we have

\[
E[d^a] = \frac{\sum_{j=1}^{\infty} \frac{p_j T_j^2}{2(1-e^{-\mu T_j})}}{\sum_{j=1}^{\infty} \frac{p_j T_j}{1-e^{-\mu T_j}}}. \tag{11}
\]

2.2 Energy Efficiency

Let \(\eta\) denote terminal’s energy efficiency and we have

\[
\eta = \frac{\sum_{j=1}^{\infty} p_j(m_j^s + m_j^a)}{\sum_{j=1}^{\infty} p_j E_j}. \tag{12}
\]
in which $E_j$ is the energy consumption during a cycle of sleep and awake durations conditioned that terminal wakes up after the $j$th sleep window.

$E_j$ consists of three parts: energy consumed on being sleep in the sleep duration, on listening in the sleep duration and on listening in the awake duration. Let $E_s$ and $E_l$ denote the energy consumption per unit of time on being sleep and listening. Let $S_s^j$ denote the sum of sleep windows, $S_l^j$ denote the sum of listening windows. Then we have

$$E_j = E_s S_s^j + E_l S_l^j + E_s S_a^j. \quad (13)$$

$S_s^j$ is given by

$$S_s^j = \sum_{i=1}^{j} T_i. \quad (14)$$

The terminal has listened $j$ times before awaking and we have

$$S_l^j = j T_l. \quad (15)$$

By substituting Eqs. (3), (5), (13), (14) and (15) into (12), the energy efficiency is given by

$$\eta = \frac{\lambda \mu \sum_{j=1}^{\infty} p_j \frac{t_j}{1-e^{-\lambda t_j}}}{\sum_{j=1}^{\infty} p_j [(\mu - \lambda) E_s \sum_{i=1}^{j} T_i + (\mu - \lambda) j E_l T_l + \frac{\lambda E_l}{1-e^{-\lambda T_l}}]} \quad (16)$$

Note that the energy efficiency is only related to the mean service time and has no relation with the distribution of service time.

3. Numerical Results and Optimal Sleep Window Selection

3.1 Average Delay and Energy Efficiency

Simulations have been made in the scenario of IEEE 802.16e networks. The unit time is 1 ms as in [6]. The listening window $T_l$ is 1, the maximum sleep window is 1024 and the initial sleep window $T_1$ is 1, 2, 4, 8, 16 for different curves. The energy consumption per unit of time is set as $E_s = 50 \text{mw}$ and $E_l = 750 \text{mw}$. In the figures below, we use Ana to denote the analysis results and Sim to denote the simulation results.

In Fig. 2, the average sleep delay decreases as the arrival rate increases because a higher arrival rate will let a terminal transit to awake after a smaller number of sleep phase $j$. It also can be seen that the larger the initial sleep window is, the longer the sleep delay is.

Figures 3 and 4 illustrate the average packet queuing delay when the service time follows 1) the deterministic distribution with mean value equal to 1 and 2) the 2-stage hyper-exponential distribution with mean value $\frac{\alpha_1}{\mu_1} + \frac{\alpha_2}{\mu_2}$ equal to 1, respectively. For the same arrival rate and sleep window size, since the variance coefficient of the hyper-exponential distribution is larger than that of the deterministic distribution, it can be seen from Eq. (6) that the queuing delay in Fig. 4 is larger than that in Fig. 3.

Different from the sleep delay, the average queuing delay does not decrease as the arrival rate increases. The reason is that when the arrival rate is relatively small compared to the service rate, the sleep delay $w_{j,2}$ contributes
more to the queuing delay. However, when the arrival rate increases, the delay $w_{j1}$ increases and contributes more to the queuing delay. Since a relatively small arrival rate leads to large $w_{j2}$ but short $w_{j1}$, while a relatively large arrival rate leads to short $w_{j2}$ but large $w_{j1}$, the curves decrease first and increase later with the increase of arrival rate.

Figure 5 shows the terminal energy efficiency, which is only related to the mean value of service time, as illustrated in the discussion of Eq. (16). The energy efficiency increases as the arrival rate increases because the “energy overhead,” i.e., the energy consumed on listening window, decreases. Larger sleep windows lead to less energy overhead and achieve better energy efficiency performance than smaller windows. As the arrival rate increases, the energy efficiency approximates $\mu/E_1 = 1.33$ (packet/mJ).

3.2 Optimal Sleep Window Size Selection with Delay Constraint

In the sleep mode operation, one of the optimization problems is: how to select the optimal sleep window size to achieve the best energy saving performance while satisfying the requirement of Quality of Service, especially the delay requirement.

For IEEE 802.16e networks, Sects. 2.1, 2.2 and 3.1 provide the queuing delay and energy efficiency performance. For the optimization problem, given the traffic condition and the delay constraint, we can first get the feasible set of sleep windows based on the delay analysis. Then, according to the analysis of energy efficiency, the longest window should be chosen from the feasible set to achieve best energy saving performance. For example, when the arrival rate is 0.1 (packet/unit time) and the service time follows the deterministic distribution, if delay constraint is 8(unit time), the feasible set of initial sleep windows is 1, 2, 4 and 8. Then we choose $T_1 = 8$ to optimize the energy efficiency.

4. Conclusion

In this letter, an analytical model has been proposed for the sleep mode based power saving mechanism. Focusing on a cycle of sleep and awake duration, all the packets arriving in each cycle are considered to analyze packet queuing delay with packet service time following general distribution. The terminal’s energy consumed on both being sleep and being awake have been considered to compute the energy efficiency. Specifically, we have applied the model to IEEE 802.16e power saving mechanism. Based on the analysis, optimal sleep window size can be derived to achieve the best power saving performance while satisfying the delay constraint under different traffic conditions.

Acknowledgments

The authors would like to thank Fujitsu Research and Development Center for the continuous support.

References