PERFORMANCE OF SLEEP-MODE MECHANISMS UNDER LIGHT-TRAFFIC CONDITIONS

K. De Turck[1], D. Fiems[1], S. Wittevrongel[1], H. Bruneel[1]

[1] *SMACS Research Group, Department TELIN*

[1] *Ghent University, Belgium*

kdeturck@telin.ugent.be

Modern standards for wireless telecommunication, such as IEEE 802.16e (WiMAX), foresee in several power-saving mechanisms. One of the main mechanisms is sleep-mode operation, which allows the mobile device to switch off the antenna during a negotiated time (it goes to *sleep*), thus reducing energy consumption. However, traffic that arrives at the base station (BS) incurs an extra delay, because its delivery can only start after the sleep period has ended. The sleep-mode mechanism in WiMAX has provisions to vary the lengths of subsequent sleep periods, which allows to exploit correlation in the arriving traffic. We depart from our earlier analyses and study the model of the sleep-mode mechanism specifically under light-traffic conditions. This results in relatively short and simple formulas that give a lot of insight.

Keywords: light-traffic approximation, power-saving mechanisms, wireless communications, IEEE WiMAX

1. INTRODUCTION

We study the performance of the energy saving mechanism defined in the IEEE 802.16e [1] standard for Broadband Wireless Access networks. Consider the downstream communication between a Mobile Station (MS) and its serving Base Station (BS) over a wireless link. To allow the MS to extend its battery life, the standard provides a *sleep-mode mechanism* that allows the MS to turn off its radio interface for a certain time whenever the BS has no packets in its queue. Specifically, if the BS's queue is empty, the MS starts a sleep period during which it remains powered down and thus cannot be reached by the BS. After this period, the MS is briefly reactivated to check whether there are packets waiting for it at the BS. If not, the MS initiates a second sleep period, a third, and so on. However, if any packets arrived at the BS during the last sleep period, the MS remains powered and enters *awake mode*. This allows the BS to transmit all the packets in its buffer exhaustively, after which the whole procedure is repeated. In case of IEEE 802.16e, an exponential update strategy is used, where the sleep periods double in length every time until a certain maximum is reached. However, in this paper, we allow more general sequences of sleep period lengths.

Quite a few authors have already investigated the performance of the IEEE sleep-mode mechanism. We refer to [2] for an overview. Our paper was the first to take into account traffic correlation.

The structure of the paper is as follows. We elaborate on the model in section 2. The light-traffic technique is explained in section 3. We give the main results in section 4. Finally, we numerically verify the exact results we obtained in our previous paper [2] with the light-traffic results we derived here in section 5.

2. MODEL AND PROBLEM STATEMENT

We model the BS as a discrete-time single-server queue with infinite capacity and a first-come-first-served discipline. Time is divided into fixed-length intervals called *slots* such that changes in the system can occur at slot boundaries only. The service (i.e. transmission) times of the packets, expressed as a number of slots, are independent and have common probability generating function (pgf) $S(z)$ and mean $\mathrm{E}[S]$. Packets arrive in the queue according to a D-BMAP with M phases. This arrival process is characterised by the values $a(k, j|i)$, $(k \geq 0, i, j \in \{1, \cdots, M\})$, denoting the probability that, given phase i in a slot, there are k packet arrivals and the phase switches to j in the next slot. These probabilities can conveniently be arranged in the $M \times M$ probability generating matrix (pgm) $\mathbf{A}(z)$ with entries $[\mathbf{A}(z)]_{ij} = \sum_{k=0}^{\infty} a(k, j|i) z^k$. The stationary probabilities of the phases are given by the entries of a row vector $\mathbf{a}$ that satisfies $\mathbf{a} = \mathbf{a}\mathbf{A}(1)$ and $\mathbf{a}\mathbf{1} = 1$, where $\mathbf{1}$ is a column vector of ones of appropriate length. The mean number of arrivals per slot is given by $\lambda = \mathbf{a}\mathbf{A}'(1)\mathbf{1}$.

The *sleep-period update strategy* is denoted by a sequence $(t_1, t_2, \ldots)$ where the integers t_n $(n \geq 1)$ are the lengths of the subsequent sleep periods of the same idle period. Let us also define $\tau_n = \sum_{i=1}^{n} t_i$ $(n \geq 1)$ and $\tau_0 = 0$ indicating the starting slots of the sleep periods relative to the start of the idle period. We assume that the lengths of the sleep periods become equal above a certain value J, i.e. $t_n = t_J$, for $n \geq J$.

We cite here the results we obtained in [2] regarding the virtual waiting time and the energy consumption. We also derived the distributions of the queue content and the packet delay, but in this paper we focus on the two aforementioned quantities. The vector generating function $\mathbf{W}(z)$ is a row vector of M generating functions with entries defined as follows:

$$[\mathbf{W}(z)]_i = \sum_{k=0}^{\infty} \Pr[w = k, b = i] z^k, \tag{1}$$

where b and w are random variables denoting the phase and the virtual waiting time at the beginning of a random slot in stationary regime. The virtual waiting time at a time instant is defined as the time until the server is ready to serve a virtual packet that arrives exactly at that time instant. We found in [2] that

$$\mathbf{W}(z) = \mathbf{w} \sum_{n=1}^{\infty} \mathbf{A}(0)^{\tau_{n-1}} (z^{t_n} - 1)(z\mathbf{I} - \mathbf{A}(S(z)))^{-1}, \tag{2}$$

where $\mathbf{w}$ is a $1 \times M$ vector that can be determined by a variety of techniques.

We define the energy consumption E of the base station as the probability that during a random slot the antenna is switched on. Note that the antenna is switched on during every busy slot and also during every last slot of a sleep period. We found that

$$E = \rho + \mathbf{w} \sum_{n=0}^{\infty} \mathbf{A}(0)^{\tau_n} \mathbf{1}. \tag{3}$$

Although the above sketched solution technique is very powerful, there are a couple of drawbacks to it. Firstly, computing the vector $\mathbf{w}$ is a numerically complex process, especially if the number of background states M is large. Secondly, the exact influence of arrival correlation is not evident from the equations. We try to circumvent these problems by considering a Taylor expansion in the region that –for this application– is most interesting from a practical point of view, namely where the traffic is low.

3. LIGHT-TRAFFIC TECHNIQUE

Although perhaps less well-known than their heavy-traffic counterparts, light-traffic approximations have a long history, and very different approaches have been developed to tackle a wide variety of systems. For an excellent overview, see e.g. [4].

It is very important how light-traffic conditions are reached: different ways of reaching low traffic can give different results. Here we take the so-called π-thinning approach: we start from a 'normal' arrival process, and then consider the family of related processes where each arrival is retained with a probability π, independent from anything else. Light-traffic conditions are then reached in the limit $\pi \to 0$.

We want to obtain the mean virtual waiting time and the mean power consumption under light traffic. Generally, these quantities are studied by relating the distributions at different time instants and then, by invoking stationarity, we get an equation for the desired quantity. Under light traffic, we can take the more direct approach of averaging over all sample paths.

For this technique to work, we assume that the operation of the system has started at time $t = -\infty$, and hence, assuming that the system is stable, has reached stationarity in slot 0. Let $\mathcal{S}$ be the set of sample paths. A sample path s for this model records the arrival times and the service durations of each packet, resulting in two sequences $i_1, i_2, \cdots$ and $j_1, j_2, \cdots$, as well as a variable g signifying the number of slots until the service of the very first packet starts. Note that g is uniformly distributed between 0 and $t_J - 1$. Let ψ be the quantity of interest, for example the virtual waiting time W at the beginning of slot 0 or the energy consumption E during slot 0. Then we can compute the expectation of this quantity under π-thinning as follows:

$$\mathrm{E}_\pi[\psi] = \sum_{s \in \mathcal{S}} P_\pi(s) f(s), \tag{4}$$

where $P_\pi(s)$ measures the probability of a sample path under π-thinning, and function $f(.)$ computes the quantity ψ for given sample path s. As the variable g and the service times are independent of the arrival process, and moreover not subject to π-thinning, average over them separately. From henceforth we only include the arrival times in the sample path.

Such probabilities are functions of π, of which we can make a Taylor expansion. Under light traffic, we consider only coefficients in, say, the n lowest powers of π. As we will see, this leads to an enormous reduction in the amount of sample paths we need to consider. Note that the probability matrix generating function of the arrival process under π-thinning is given by:

$$\mathbf{A}_\pi(z) = \mathbf{A}(1 - \pi + \pi\ z), \tag{5}$$

which gives rise to the following Taylor-expansion in $\pi = 0$:

$$\begin{aligned}\mathbf{A}_\pi(z) &= \sum_{n=0}^{\infty} \frac{1}{n!}\mathbf{A}^{(n)}(1)\pi^n(z-1)^n \\ &= \sum_{k=0}^{\infty} z^k \sum_{n=k}^{\infty}(-1)^{n-k}\binom{n}{k}\frac{1}{n!}\mathbf{A}^{(n)}(1)\pi^n \\ &= \sum_{k=0}^{\infty} z^k \mathbf{C}_k(\pi). \end{aligned}\tag{6}$$

Note that the probability of a sample path can written as an infinite product of transition matrices $\mathbf{C}_k(\pi)$, premultiplied by vector $\mathbf{a}$ and postmultiplied by vector $\mathbf{1}$, where each k signifies the number of arrivals during a certain slot. From this we observe that the probability of a sample path with n arrivals is a power series in π where the lowest non-zero coefficient is in π^n.

We state, without proof the following theorem, which shows the formula for a second order light-traffic approximation of a quantity ψ. The proof follows the lines of the expositions in [3] and [5], but must be adapted to D-BMAP arrivals. It is quite technical due to a tricky interchange of limits.

Theorem 1. *The expectation of a quantity ψ with associated function $f(.)$ under π-thinning has the following second-order approximation:*

$$\mathrm{E}_\pi[\psi] = f(\{\}) + \pi\lambda \sum_{i=-\infty}^{+\infty} \left(f(\{i\}) - f(\{\})\right) \tag{7}$$

$$+\ \pi^2 \sum_{i_1,i_2=-\infty}^{+\infty} R(|i_1 - i_2|)\left(f(\{i_1,i_2\}) - f(\{i_1\}) - f(\{i_2\}) + f(\{\}\right) + \mathcal{O}(\pi^3), \tag{8}$$

where $\{i_1, i_2, \cdots\}$ denotes a sample path with arrivals in slots i_1, i_2 and so on. The function $R(.)$ is defined as

$$\begin{aligned} R(0) &= \frac{1}{2}\mathbf{a}\mathbf{A}''(1)\mathbf{1}, \\ R(n) &= \frac{1}{2}\mathbf{a}\mathbf{A}'(1)\mathbf{A}(1)^{n-1}\mathbf{A}'(1)\mathbf{1}, \quad \textit{where } n \geq 1. \end{aligned}\tag{9}$$

We can view $R(.)$ as a sort of autocorrelation function.

4. MAIN RESULTS

The functions $f_E(s)$ and $f_W(s)$ that determine respectively the energy consumption and the unfinished work during slot 0 for a given sample path s follow quite straightforwardly from the system equations. For two arrivals, the complete definitions are already quite lengthy, and hence we omit them in favor of the results themselves.

After some tedious manipulations, we find from formula (8) that the energy consumption is approximated by

$$\begin{aligned}\mathrm{E}_\pi[E] = {} & \frac{1}{t_J} + \pi\lambda\left\{\mathrm{E}[S] + J - 1 - \frac{\mathrm{E}[S] + \tau_{J-1}}{t_J}\right\} + \\ & +2\pi^2 \sum_{j=0}^{J-1} X_j \left(\frac{\tau_{J-1} - \tau_j}{t_J} - J + 1 + j\right) - R(0)\left(\frac{\tau_{J-1} - \tau_j}{t_J} - J + 1\right) + \mathcal{O}(\pi^3),\end{aligned} \tag{10}$$

where the X_js are defined as

$$X_0 = \frac{1}{t_J}\sum_{i=0}^{t_J-1}\sum_{k=1}^{\infty}\Pr[S=k]\sum_{n=0}^{k+i} R(n) \tag{11}$$

$$X_j = \frac{1}{t_J}\sum_{i=0}^{t_J-1}\sum_{k=1}^{\infty}\Pr[S=k]\sum_{m=0}^{k-1} R(k+m+i+\tau_{j-1}), \quad \text{if } 0 < j < J. \tag{12}$$

This expression for $P[E]$ has an intuitive basis. The constant term tells what happens when there are no arrivals at all: the system will forever be in the sleep period with length t_J, and hence have an energy consumption $\frac{1}{t_J}$. In the first order derivative we see the amount of *extra* energy consumption induced by one arrival. The second order derivative shows the influence of the overlap of two packet 'life times' (service time plus subsequent sleep periods). The terms X_j are such that $\pi^2 X_j$ gives the probability of having a sample path with two arrivals that have exactly j sleep periods in between. For the mean virtual waiting time we find:

$$\begin{aligned}\mathrm{E}_\pi[W] = {} & \frac{t_J - 1}{2} + \pi\lambda\left\{\frac{1}{2}\mathrm{E}[S(S-1)] + T_{J-1} - \tau_{J-1}\frac{t_J - 1}{2}\right\} + \\ & +2\pi^2\left\{\sum_{j=0}^{J-1} X_j\left(\frac{t_J - 1}{2}\tau_{J-1} - T_{J-1} + \mathrm{E}[S]\right) + Y\,\mathrm{E}[S] - X_0\,\mathrm{E}[S]^2 + \right. \\ & \left. + \sum_{j=1}^{J-1} X_j\left(\frac{t_J-1}{2}(\mathrm{E}[S] - \tau_j) - T_j - \mathrm{E}[S]t_j\right)\right\} + \mathcal{O}(\pi^3),\end{aligned} \tag{13}$$

where

$$Y = \frac{1}{t_J}\sum_{i=0}^{t_J-1}\sum_{k=1}^{\infty}\Pr[S=k]\sum_{n=0}^{\tau_{J-1}+k-1} nR(n) \tag{14}$$

$$T_j = \sum_{i=1}^{j}\frac{t_i(t_i-1)}{2} \tag{15}$$

5. NUMERICAL EXAMPLE

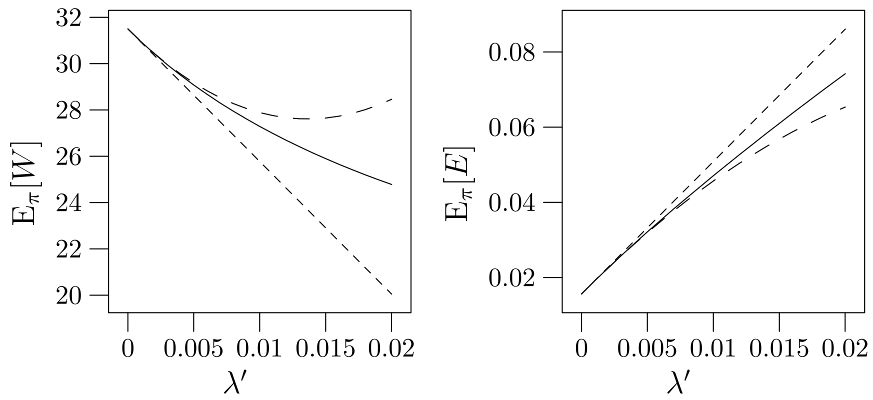

Fig. 1. Plots showing the exact curve of $\mathrm{E}_\pi[E]$ *and* $\mathrm{E}_\pi[W]$ *vs.* $\lambda' := \lambda\pi$ *as well as their first and second order approximations (in straight lines, dashes and long dashes respectively).*

We verified the obtained light-traffic limits with our earlier results for an on/off arrival source and for sleep strategy $(4, 8, 16, 32, 64, \cdots)$ and found indeed an agreement.

6. CONCLUSION

Light-traffic approximations are a very interesting tool for the analysis of power-saving protocols. We can not only reconstruct our previous work under light traffic, but all sorts of generalizations seem possible as well. It is remarkable that the first order approximation is insensitive to traffic correlation. It would be interesting to find out whether this is due to either the applied thinning scheme or the Markovian traffic correlation, or whether this can be established more generally.

REFERENCES

1. *IEEE 802.16e-2005* Part 16: Air interface for fixed and mobile broadband wireless access systems // 2006.
2. *De Turck K., De Vuyst S., Fiems D., Wittevrongel S.* Performance analysis of the IEEE 802.16e sleep mode for correlated downlink traffic // Telecommunication systems 2008. V. 39. No 2. P. 145–156.
3. *Reiman M., Simon B.* Open queueing systems in light traffic // Mathematics of operations research, 1989. V. 14. No. 1. P. 26–59.
4. *Kovalenko I.* Rare events in queueing theory. A survey // Queueing systems 1994. V. 16. No 1. P. 1–49.
5. *Tsoukatos. K.* Heavy and light traffic regimes for $M|G|\infty$ traffic models // PhD thesis. 1999.