Providing quality of service guarantees in multiclass IEEE 802.16e sleep mode

Georgios Paschos
CERTH (The Center for Research and Technology Hellas)
Email: gpasxos@uth.gr

Petteri Mannersalo
VTT Technical Research Centre of Finland
Email: petteri.mannersalo@vtt.fi

***Abstract*—We consider the sleep mode algorithm of IEEE 802.16e mobile networks where a mobile station may switch off the transceiver and go into sleep mode in order to save power. Depending on the underlying application, different algorithms are defined in the standard. In practice, a hybrid combination of the sleep mode algorithms will be used in a mobile. In this paper, we build a novel performance model that captures the behavior of a single sleep mode algorithm. Moreover, we provide an accurate approximation for the hybrid case. Finally, the model is used to select the standard-compliant sleep window strategy which satisfies a given (approximative) delay constraint and minimizes the energy usage.**

I. Introduction

IEEE 802.16 standard is one of the main carriers of broadband wireless access. The user is expected to roam freely while accessing high speed connectivity and Internet services. Nevertheless, energy consumption and battery life set constraints for usability of wireless services. In an effort to make wireless technology more efficient and handy, sleep mode algorithms are often associated with wireless protocol standards. For this reason, the IEEE 802.16e extension proposed in [1], defines sleep mode algorithms and handover management.

There are three power saving classes (PSCs) defined in the standards. The first one (type-I) applies to data transmissions and it utilizes an exponential algorithm for the sleep duration. This class can be also used when the mobile is idle, i.e., it has no applications running. The second (type-II) has a fixed sleep duration and is used by isochronous applications like VoIP. The third (type-III) is usually initiated by the base station (BS) in case of multicast transmissions or periodic ranging.

The standalone mode of type-I has been studied extensively. The first work on type-I sleep mode algorithm is by Xiao [2]. Zhang and Fujise [3] extended the analysis to include outgoing traffic. Moving away from purely Poissonian arrivals, Zhang [4] considered the difference between interarrival time and interevent time and proposed hyper-Erlang distribution to model the arrival process. In [5], Almhana et al. are using hyperexponential distribution for arrivals. This distribution has more degrees of freedom, and can be used to simulate multimodal arrivals (e.g., inside a flow and between two flows). Moreover, hyperexponential distribution can model the case that the arrival rate of a Poisson process is unknown and chosen among a finite set of values (see Azad et al. [6]). Another effort with more general non-Poissonian arrival processes is by Hsu and Feng [7]. They apply the Residual life theorem in their analysis. Unfortunately, analyzing the residual times is actually quite involved and intractable (see section III).

All the above efforts are concentrated on type-I PSC. Nevertheless, as described in the standards, a mobile station (MS) can maintain an arbitrary number of sleep mode instances per class resulting in a hybrid scenario. Then the MS turns off the transceiver only on common unavailability periods of all instances. More importantly, a packet assigned to type-I class can also be monitored in an availability interval of another class. This hybrid case is expected to appear quite often when terminals have active VoIP applications and/or terminals are mobile and quality checks are established by a periodic ranging procedure. Very few analytical attempts have been made in this direction and to our knowledge, there are no results for the general hybrid case. Chen et al. [8] provide means to decide the optimal synchronization of arbitrarily many instances of type-II and type-III based on Chinese remainder theorem. In [7], performance analysis is made for each class alone but not for the combination of them. Kong and Tsang propose a class selection mechanism in [9], keeping only one class operating at a time.

The parameters of the sleep mode algorithms can be optimized to adapt to traffic load. The standards actually provide room for optimizing the behavior by tuning two variables *initial_sleep_window* and *final_sleep_window*. To our knowledge, all work in this area has been focusing on a single PSC. Alouf et al. [10] introduce an optimization tool to obtain the optimal settings fulling given average delay and energy guarantees. In [5], the authors calculate the distribution of delay and propose a control method for adaptation on the arrivals. The above approaches aim for finding the optimal tradeoff between delay and power consumption using adaptive algorithms.

In this paper, we put forward an optimization scheme that works on the class of policies admissible by the standards. The results can be used by an operator to customize a network that is 802.16e standard compliant. We deviate from the earlier studies by 1) relaxing the assumption of non-responsiveness in the listening windows, 2) providing results for the distribution of delay in closed form without the above assumption, 3) providing approximate results for the hybrid multiclass case for the first time in the literature and 4) proposing an optimization method that provides quality of service guarantees in a standard-compliant system.

The rest of the paper is organized as follows. In section II the sleep mode algorithms are described for all power saving classes. In sections III and IV, the modeling assumptions and performance analysis are presented. In section V the protocol parameters are optimized to fulfill QoS constraints. Finally, the paper is concluded in section VI.

II. Power saving classes

A. Type-I

Before entering the sleep mode under type-I PSC, the MS sends a request message to BS to ask for permission to transit into sleep mode. This message negotiates the parameters initial_sleep_window ($s_{\min}$), final_sleep_window ($s_{\max}$) and listening_window (l). After an initial sleeping period, the MS sets the transceiver to listening mode for a fixed period l and receives the traffic indication message MOB-TRF-IND sent by the BS. The message indicates whether there has been incoming traffic addressed to the MS during its absence. If MOB-TRF-IND is negative, then the MS continues its sleep mode by doubling the sleeping duration. If positive, the MS exits the sleep mode and serves the incoming packets which are stored at the BS.

The exponentially increasing sleeping durations $s_j^{(1)}$, $j = 1, 2, \ldots$, are defined by $s_j^{(1)} = \min\{2^{j-1}s_{\min}, s_{\max}\}$. For the listening windows $l_j^{(1)}$, we have $l_1^{(1)} = l_{\text{init}}$ and $l_j^{(1)} = l, j \geq 2$, where l_{init} denotes the initialization time.

The choice of parameters $s_{\min}$, $s_{\max}$, l_{init} and l is relative to desired quality (tradeoff between delay and energy). Usually, l_{init} and l are parameters defined by the technology used. Then $s_{\min}$ and $s_{\max}$ are the only customizable parameter to improve the performance of the system. From [6], it is inferred that in case of exponential interarrival times and $l_{\text{init}} = 0$ it always makes sense to choose a fixed length for the sleeping periods, i.e., $s_{\min} = s_{\max} = h(\lambda)$ where h is some function of the arrival rate λ. However, using $s_{\max} > s_{\min}$, as the standards indicate, provides on one hand a dynamic adaptation (no exact knowledge of λ is needed) and covers cases of mixed arrivals. In this paper, we show that if $l_{\text{init}} > 1$ then $s_{\min} = s_{\max}$ is not always optimal.

B. Type-II and type-III

The procedure for type-II is similar to type-I with the only difference that type-II is periodic. As such, the initial and final sleeping windows are equal. Type-III differs from the other classes by being defined only for one sleep period. However, the BS might repeat the initialization of an instance of this class in a periodic manner. An application of this can be found in mobile 802.16e networks, where periodic estimation of link quality takes place by means of a procedure called periodic ranging. When the MS moves from point to point, the BS should make sure to maintain a type-III instance in order to track the MS at pre-scheduled times. The period of ranging in this case depends on the velocity of the terminal.

In analogy to the previous subsection, the sleeping and listening windows for classes $k = 2, 3$ are defined by $s_j^{(k)} = s^{(k)}$ and $l_j^{(k)} = l^{(k)}$, where $j = 1, 2, \ldots$ for $k = 2$ and $j = 1$ for $k = 3$. Note that these parameters are not customizable variables since they depend on the corresponding application.

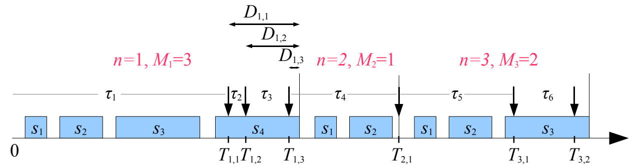

Fig. 1. System model under Assumption 1 (queueing neglected).

III. Model definition and assumptions

Consider type-I arrivals defined by a process with i.i.d. packet interarrival times $\tau_i, i = 0, 1, \ldots$. The interval between two consecutive busy periods is called idle period (or sleep mode period) and we denote by M_n the number of packets arriving in idle period n. We index the packet arrival times in idle period n as $T_{n,i}, i = 1, \ldots, M_n$, where $T_{n,1}$ is the time from the beginning of the idle period to the arrival of the first packet and the $T_{n,i}$ with $i > 1$ are the interarrival times of the packets arriving in idle period n. The time from an arrival in the idle period to the end of the corresponding idle period is called (sleep mode) delay and denoted by $D_{n,i}, i = 1, \ldots, M_n$. In a real system, each idle period is followed by a busy period of finite length during which new packets may arrive into queue. In this paper we will neglect this and assume an instantaneous service of the packets (Assumption 1 below). Figure 1 illustrates the system.

In the analysis of the system there is a number of issues to be accounted for, namely, 1) queueing issues, 2) arrival process, 3) hybrid case and 4) residuals times, which we express with the following assumptions.

Assumption 1: Service times are set to zero and there is no queueing taking place in the busy periods, i.e., the busy periods have zero length.

Relaxing Assumption 1 has been studied in [10] and [11]. Since the service times depend on scheduling from BS and in general are very short, it is reasonable to overlook queueing. Moreover, we are mostly interested in the idle periods of the system and particularly the tradeoff between delay and energy savings during the idle periods. Finally, the total busy period time and the total energy spent in busy periods will be the same for any sleep mode algorithm. For these reasons, we adopt Assumption 1, similar to the majority of approaches in the literature (see [2], [3], [4], [5], [6]). Nevertheless, note that the modeling of busy periods, despite making the analysis involved, would be useful in order to compute correctly the residual times.

Assumption 2: The arrival process is a Poisson point process.

This assumption is used extensively in the literature as well (see [2], [3], [4], [11]). It is extremely difficult to calculate the performance of the system for any other arrival process, even if the arrivals are considered to have i.i.d. interarrival times. Although calculating $D_{1,1}$ using $T_{1,1}$ is relatively straightforward (it is a piece-wise location and scale transformation of random variables), doing the same with τ_i requires the knowledge of residual times. The residual time to the first arrival at the n^{th} period depends on $D_{n-1,M_{n-1}}$ and the length of the $n-1$ busy period. Authors in [4] and [7] attempted to use the Residual life theorem from [12]. Note, however, that this theorem applies when the observation point is thrown uniformly into the renewal process. This is not true since we know that the observation point depends on the above mentioned random variables. The memoryless property of Poisson point process eliminates the need for calculating residual times.

Assumption 3: In a hybrid scenario, the type-I and periodic (type-II and type-III) sleep mode instances are mutually independent.

In a hybrid scenario with one instance of type-I class and many instances of type-II and type-III, the sleep mode process depends only on arrivals of packets belonging to type-I applications. Every time that such a packet is served, the type-I instance will restart while the rest of the instances will continue their periodic pattern unaffected. Assumption 3 is frequently invalidated whenever a type-I packet waiting in the BS is monitored in an availability period caused by a listening period of a periodic class. If the periodic instance was sleeping when the packet arrived, then the type-I instance gets synchronized with the periodic instance. We will use this assumption to provide approximate results for the hybrid scenario. The simulations show that the error caused is quite small for the parameter values used in real systems.

Assumption 4: (Not used in this paper.) An arrival during a listening window does not immediately trigger a waking up. Instead it is monitored as part of the next sleeping window.

This assumption has been used, e.g., in [2], [3], [4], [5], [6]. In this work we relax this assumption by monitoring immediately arrivals in listening periods. Packets arriving in a listening window have $M_n = 1$ and $D_{n,1} = 0$. Allowing these events to exist brings important difference since arrivals in a listening period have different properties from those in a sleeping period.

Our model using only Assumptions 1–3 is general in the sense that it can also capture the behavior of a system operating under Assumption 4 as well. Set for example $l' = 0$ and $s'_j = s_j + l$ to obtain delay results for a system where Assumption 4 holds. Related to Assumption 2, one could also to consider only interactive applications, see [6]. In this case, each user responds only after the reception of one action. In other words, τ_i and $T_{1,1}$ are identical. Using this assumption one could easily modify our analysis to compute the performance of any arrival process with i.i.d. interarrival times whose location and scale transformation is tractable.

IV. Performance analysis

A. Preliminaries

Assume that a sleep mode system satisfies Assumption 1, i.e., the busy periods have zero length. If the sequences $\{M_n\}_n$ and $\left\{\sum_{i=1}^{M_n} \mathbf{1}\{D_{n,i} > x\}\right\}_n$ are ergodic, then the distribution of the delay of an arbitrary packet is given by

$$\begin{aligned}\mathrm{P}(D > x) &= \lim_{n\to\infty} \frac{\sum_n \sum_{i=1}^{M_n} \mathbf{1}\{D_{n,i} > x\}}{\sum_n M_n} \\ &= \frac{\mathrm{E}\left[\sum_{i=1}^{M_1} \mathbf{1}\{D_{1,i} > x\}\right]}{\mathrm{E}[M_1]} \\ &= \frac{\mathrm{E}[\mathbf{1}\{D_1 > x\}(1 + N(T_1, T_1 + D_1 - x))]}{\mathrm{E}[1 + N(T_1, T_1 + D_1)]},\end{aligned}$$

where T_1 and D_1 are the packet arrival time and delay of the first packet to arrive with the condition that the previous service time was at $t = 0$, $\mathbf{1}\{.\}$ is the indicator function and $N(t_1, t_2)$ counts the number of packet arrivals in the time period (t_1, t_2). By Poisson arrivals (Assumption 2) and conditioning with respect to D_1

$$\mathrm{P}(D > x) = \frac{\mathrm{E}[\mathbf{1}\{D_1 > x\}(1 + \lambda(D_1 - x))]}{1 + \lambda \mathrm{E}[D_1]}.$$

For the mean delay we get

$$\begin{aligned}\mathrm{E}[D] &= \int_0^\infty \mathrm{P}(D > x)\,\mathrm{d}x \\ &= \frac{\mathrm{E}\left[\int_0^{D_1}(1 + \lambda(D_1 - x))\mathrm{d}x\right]}{1 + \lambda \mathrm{E}[D_1]} = \frac{\mathrm{E}[D_1] + \frac{\lambda}{2}\mathrm{E}\left[D_1^2\right]}{1 + \lambda \mathrm{E}[D_1]},\end{aligned}$$

where Fubini's theorem was used to exchange the order of integration and expectation.

By the PASTA (Poisson Arrivals See Time Averages) property, $\mathrm{P}(\text{System sleeping}) = \mathrm{P}(D > 0)$. Thus the mean power usage is given by

$$\mathrm{E}[P] = P_{\mathrm{L}} - \mathrm{P}(D > 0)\,(P_{\mathrm{L}} - P_{\mathrm{S}}),$$

where P_{L} and P_{S} denote the power usage during listening and sleep mode, correspondingly. The mean power consumption is relative to the use of sleep mode only and it does not consider power used for transmission of packets. For a sleep algorithm that matches perfectly the idle periods we would have P_{S} and for no sleeping we have P_{L} consumption respectively.

In the earlier studies, the results are usually performance formulae for the mean delay or the mean power usage in an arbitrary sleep mode period. If one is interested in delay characteristics of a random packet or overall power usage, then lengths of the idle periods or equivalently the number of packets arriving there should also be accounted for. E.g., an arbitrary packet is more likely to enter the system during a long idle period than a short one (Inspection paradox). The above results are correct in this case. Note also, that the analysis holds also for hybrid scenario if one would be able to calculate the characteristics of D_1.

B. Type-I

Recall that in type-I PSC, the lengths of the sleeping intervals are exponentially increasing up to the maximum length $s_{\max}$ if no packets arrive whereas the listening intervals remain unchanged, i.e., $s_j^{(1)} = \min\{2^{j-1}s_{\min}, s_{\max}\}$ for $j = 1, 2, \ldots,$

and $l_1^{(1)} = l_{\mathrm{init}}$ and $l_j^{(1)} = l$ for $j = 2, 3, \ldots$. Setting $s_0^{(1)} = l_0^{(1)} = 0$, it is convenient to define the cumulative time function

$$v_j^{(1)} = \sum_{k=0}^{j-1} \left(s_k^{(1)} + l_k^{(1)} \right) + l_j^{(1)}.$$

Moreover, define the first state where the maximum sleep interval is used by $m = \sup\{k + 1 : s_k^{(1)} < s_{\max}\}$.

For any piecewise continuous function $g : \mathbb{R} \to \mathbb{R}$,

$$\mathrm{E}[g(D_1)] = \sum_{j=1}^{\infty} \int_{v_j}^{v_j+s_j} g(v_j + s_j - t)\lambda e^{-\lambda t} \mathrm{d}t$$
$$= \sum_{j=1}^{\infty} e^{-\lambda v_j} G(s_j) = \sum_{j=1}^{m-1} e^{-\lambda v_j} G(s_j) + \frac{e^{-\lambda v_m} G(s_m)}{1 - e^{-\lambda(l_m+s_m)}}, \tag{1}$$

where $G(s) = \int_0^s g(s-t)\lambda e^{-\lambda t}\mathrm{d}t$. The infinite sum is always divided in two terms, a finite sum with $m - 1$ terms and a geometric sum which is given in a closed form expression.

In this paper, functions $g(s, x) = \mathbf{1}\{s > x\}(1 + \lambda(s - x))$, $g_1(s) = s$ and $g_2(s) = s^2$ are used. Simple calculations give $G(s, x) = \lambda(s - x)^+$ and

$$G_1(s) = \frac{s\lambda - 1 + e^{-\lambda s}}{\lambda},$$
$$G_2(s) = \frac{s\lambda(s\lambda - 2) - 2e^{-s\lambda} + 2}{\lambda^2}.$$

Thus, for the type-I PSC, we have

$$\mathrm{P}\left(D^{(1)} > x\right) = \frac{\mathrm{E}[g(D_1, x)]}{1 + \lambda \mathrm{E}[g_1(D_1)]}, \tag{2}$$
$$\mathrm{E}\left[D^{(1)}\right] = \frac{\mathrm{E}[g_1(D_1)] + \frac{\lambda}{2}\mathrm{E}[g_2(D_1)]}{1 + \lambda \mathrm{E}[g_1(D_1)]},$$
$$\mathrm{E}\left[P^{(1)}\right] = P_{\mathrm{L}} - \frac{\mathrm{E}[g(D_1, 0)]}{1 + \lambda \mathrm{E}[g_1(D_1)]}(P_{\mathrm{L}} - P_{\mathrm{S}}),$$

which can be calculated by Equation (1).

In Figure 2, the distribution of delay for type-I class is shown. The analysis and discrete-event simulations provide identical results as expected because the solution for this case is exact.

C. Type-II and type-III

Assume that a periodic type-II sleep mode instance is constantly running. For the periodic ranging kind of applications, the analysis with type-III sleep mode processes would be exactly the same. We consider the (virtual) delay of type-I packets caused by a type-II sleep mode instance. Since the periodic type-II sleep mode process is independent of type-I arrivals, an arbitrary type-I packet arrives uniform randomly on a type-II sleep mode period containing $s^{(2)}$ units of sleeping and $l^{(2)}$ units of listening. Denote by $D^{(2)}$ the (virtual) delay of a type-1 packet caused by type-II sleep mode instance. Then

$$\mathrm{P}\left(D^{(2)} > x\right) = \frac{s^{(2)} - x}{s^{(2)} + l^{(2)}},$$

The same holds for periodically renewed type-III sleepmode processes.

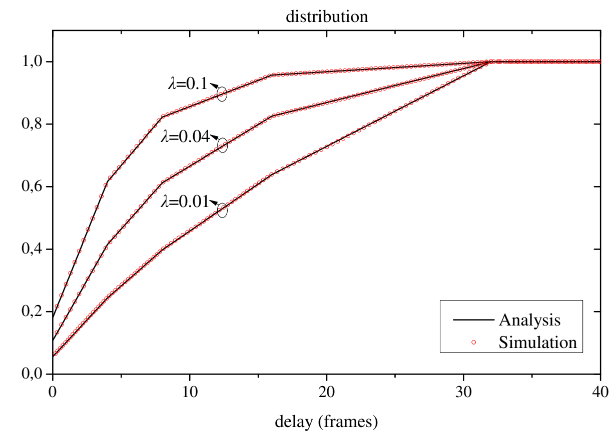

Fig. 2. Distribution of the type-I sleep mode delay. Settings: $s_{\min} = 4$, $s_{\max} = 32$, $l_{\mathrm{init}} = 2$, $l = 1$ and $\lambda = 0.01, 0.04, 0.1$.

D. Hybrid scenario

In the hybrid scenario, we combine the results from the previous subsections. Assume that an instance from type-I class always exists and then add $N - 1$ independent instances of type-II or type-III both called periodic instances. Now the random variable $D^{(1,\ldots,N)}$ corresponds to the hybrid case and $D^{(k)}(i)$ to the i^{th} of class k instance (each having different settings) considered in isolation.

Recall that the type-I instance is always restarted at the service time of a packet. When a type-I packet arrives to the system the following cases are possible:

1) arrival during a listening period of the type-I instance (immediate service) or the type-I instance starts listening before any of the periodic instances (service when the type-I instance turns active).
2) arrival during a sleep period of the type-I instance and a periodic instance is either listening (immediate service) or a sleep interval of a periodic instance is ending earlier than that of the type-1 instance (service when the periodic instance turns active)

Analyzing the system exactly is hard (if not even impossible). However, an approximation can be derived if we assume that instead of restarting the type-I instance at the service time determined by a periodic instance (case 2) we let the type-I run unchanged until it would have restarted in isolation. This assumption makes all sleep mode instances mutually independent and the distribution can be approximated by

$$\mathrm{P}\left(D^{(1,\ldots,N)} > x\right) \approx \mathrm{P}\left(D^{(1)} > x\right) \prod_{i=2}^{N} \mathrm{P}\left(D^{(2)}(i) > x\right).$$

Approximations for the mean delay and mean power usage can found by applying the above approximation to

$$\mathrm{E}\left[D^{(1,\ldots,N)}\right] = \int_0^{d_{\max}} \mathrm{P}\left(D^{(1,\ldots,N)} > x\right) \mathrm{d}x,$$
$$\mathrm{E}\left[P^{(1,\ldots,N)}\right] = P_{\mathrm{L}} - \mathrm{P}\left(D^{(1,\ldots,N)} > 0\right)(P_{\mathrm{L}} - P_{\mathrm{S}}),$$

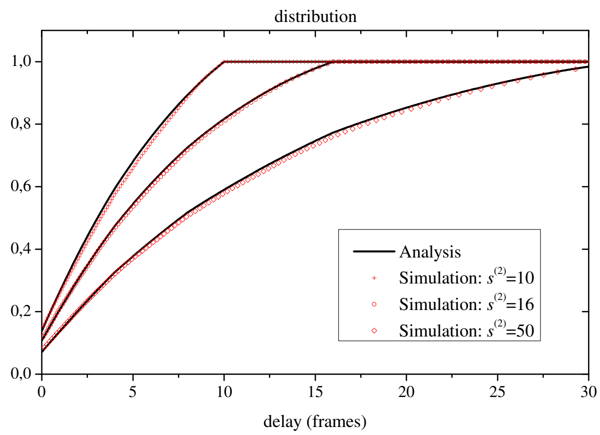

Fig. 3. Distribution of the multiclass sleep mode delay. Settings: $s_{\min} = 4$, $s_{\max} = 32$, $l_{\text{init}} = 2$, $l = 2$ and $\lambda = 0.01$.

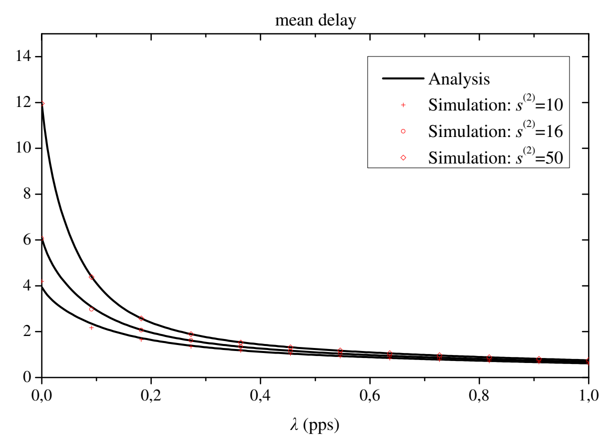

Fig. 4. Multiclass mean delay as a function of the arrival rate. Settings: $m = 4$, $s_{\min} = 4$, $s_{\max} = 32$, $l_{\text{init}} = 2$ and $l = 1$.

where $d_{\max} = \min\left(s_{\max}, s^{(2)}(2), \ldots, s^{(2)}(N)\right)$ is the maximum delay in the system.

Analysis of a hybrid scenario with more than one periodic instances is possible merely by elaborating on the approximation. Here we present results when there is one type-I and one periodic instance. Three scenarios are considered ($s^{(2)} = 10, 16, 50$): the first corresponds to case of high mobility and periodic ranging, the second to a VoIP application and the third to low mobility and periodic ranging. The behavior of the delay is shown in Figures 3 and 4. Figure 5 showcases the mean power consumption.

V. Optimization Method

In this section we describe how to use our model in order to optimize the settings of PSC-I by providing quality of service guarantees according to the delay constraints of the given application.

A. Type-I only

Let us assume initially that parameters l_{init} and l are fixed by the application. Also, we assume $s_{\max} = 2^{m-1} s_{\min}$.

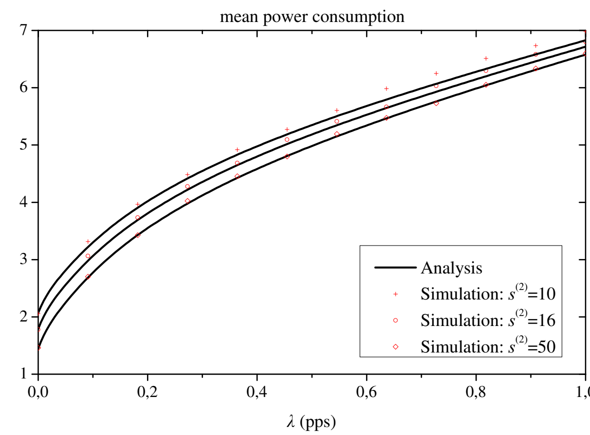

Fig. 5. Multiclass mean power consumption as a function of the arrival rate. Settings: $P_{\text{L}} = 10$, $P_{\text{S}} = 1$, $m = 4$, $s_{\min} = 4$, $s_{\max} = 32$, $l_{\text{init}} = 2$ and $l = 1$.

Consequently, both the delay distribution $F_D(x) = \mathrm{P}(D \leq x)$ and the average power consumption $\mathrm{E}[P]$ are parameterized by m and $s_{\min}$ with $m \in \mathbb{N}$ and $s_{\min} \in \mathbb{R}^+$.

We are interested in solving the following problem:

$$\text{minimize}\,\{\mathrm{E}[P]\} \quad \text{s.t.} \quad F_D\left(d_{\text{thr}}\right) \geq a, \tag{3}$$

where the minimization is done with respect to m and $s_{\min}$. Here d_{thr} is the threshold for delay (expressed in frames) and a is a confidence interval (e.g. 0.95 for 5% confidence).

Remark 1: Denote by $D(m, s_{\min})$ the delay in a system with parameters m and $s_{\min}$. Then the following problem is equivalent to (3):

$$\begin{aligned} &\text{maximize}\,\{\mathrm{P}(D(m, s_{\min}) > 0)\}\,, \\ &\text{s.t. } \mathrm{P}(D\left(m, s_{\min}\right) > d_{\text{thr}}) = 1 - a \end{aligned} \tag{4}$$

The equality constraint in the above is sufficient because for fixed m and $s_{\min}$, function $F_D(x)$, is strictly increasing and continuous on $x \in (0, s_{\max})$. This is easy to show: For a fixed s function $G(s,x) = \lambda(s-x)^+$ is continuous and decreasing whenever $x \in (0,s)$. Applying equations (1) and (2) we see that $F_D(x)$ is continuous and increasing on $x \in (0, s_{\max})$.

The following simple algorithm (given here for clarification) solves problem (4).

Input: Parameters: $l_{\text{init}}, l, \lambda$
Output: $\{m^*, s^*_{\min}\}$
$s^*_{\min} = \infty$;
for $\underline{m = 1 : M}$ **do**
  Calculate $s^*_{\min}(m)$ using the equality constraint;
  if $\underline{s^*_{\min}(m) \leq s^*_{\min}}$ **then**
    $s^*_{\min} = s^*_{\min}(m)$;
    $m^* = m$;
  end
end

Algorithm 1: Optimizing the type-I class parameters.

Using algorithm 1, we first study the optimality in case of type-I only sleep mode. Due to memoryless property of the exponential arrivals one expects that $m = 1$ is always

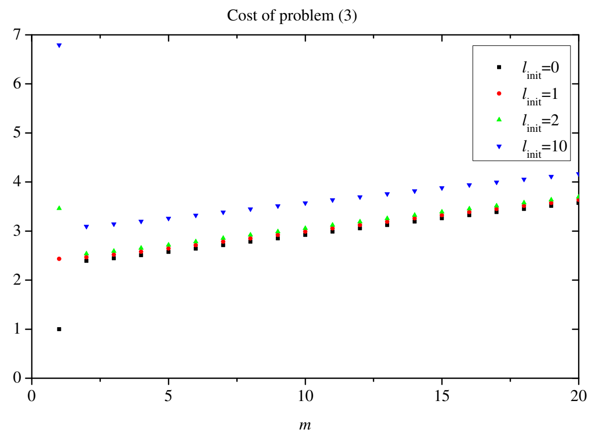

Fig. 6. Value of the target function in (3) as a function of m for type-I only. Settings: $l = 1$, $d_{\text{thr}} = 5$, $\alpha = 0.95$, $\lambda = 0.01$ and $l_{\text{init}} = 0, 1, 2, 10$.

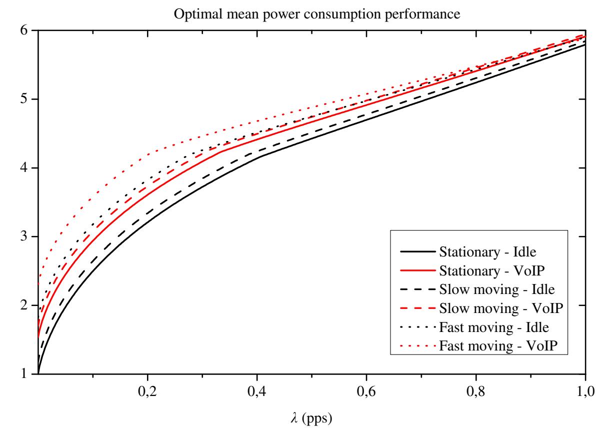

Fig. 7. Mean power consumption with QoS guarantees for delay in several application scenarios. Settings: $P_{\text{L}} = 10$, $P_{\text{S}} = 1$, $l_{\text{init}} = 2$, $l = 1$, $d_{\text{thr}} = 5$, $\alpha = 0.95$, for slow moving $s^{(2)} = 50$, for fast moving $s^{(2)} = 10$ and for VoIP $s^{(2)} = 16$.

optimal (see [6]). However, solving the problem for different l_{init} we observe that the solution depends on l_{init}. More specifically, for $l_{\text{init}} = 0, 1$ the optimal is indeed $m = 1$ but for $l_{\text{init}} > 1$ making more than one steps seems to be optimal. The results are shown in Figure 6. Qualitatively, analogous behavior is seen when using average delay constraints instead of constraints for the distribution. The reason for this behavior is the initialization period which makes myopic optimization impossible. The system needs to adapt to two different modes, the first one that includes l_{init} and the second that lasts for the rest of the time. Therefore, using $m = 1$ can be suboptimal.

B. Hybrid scenario

In this subsection we provide results for the problem (3) in the case of multiple sleep mode classes. More specifically, we look at a terminal running a VoIP application and trying to save power in the periodic idle segments. This is a problem similar to the case of mobile networks (e.g. GSM) and thus of particular interest. We also consider a mobile terminal that requires ranging procedure in order to trace the signal quality. This demonstrates the impact of mobility in the sleep mode algorithm.

By applying algorithm 1 to the approximate results from Section IV, a good solution for the operation of the standards can be obtained. Particularly, in Figure 7 we show the minimum power consumption achieved in all cases for several arrival rates while respecting the quality constraints for delay.

VI. Conclusion

In this paper we provide an approximate solution to the problem of multiclass sleep mode for IEEE 802.16e mobile networks. The approximation conformance seems reasonably good in most cases tested. Using this approximation, one can analyze optimal behavior of the current standards in case of terminals with ongoing VoIP applications and/or performing periodic ranging with the BS, while respecting the Quality of Service constraints for delay. The results are optimistic in the sense that the sleep mode performance is only mildly deteriorated when having more than one sleep instances.

Acknowledgment

The work of P. Mannersalo was supported by TEKES as part of ABI project and the Future Internet program of TIVIT. The work of G.S. Paschos was supported by the project STREP-FP7-INFO-ICT-224218: OPNEX.

References

[1] IEEE 802.16e Workgroup, "Part 16: Air Interface for Mobile Broadband Wireless Access Systems-Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands", Feb. 2005.

[2] Y. Xiao, "Energy saving mechanism in the IEEE 802.16e wireless MAN", IEEE Communications Letters, vol. 9, no. 7, pp. 595-597, 2005.

[3] Y. Zhang and M. Fujise, "Energy management in the IEEE 802.16e MAC", IEEE Communications Letters, vol. 10, no. 4, pp. 311-313, 2006.

[4] Y. Zhang, "Performance modeling of energy management mechanism in IEEE 802.16e mobile WiMAX", in proc. of IEEE WCNC, pp. 3205-3209, 2007.

[5] J. Almhana and Z. Liu and C. Li and R. McGorman, "Traffic Estimation and Power Saving Mechanism Optimization of IEEE 802.16e Networks", in proc. of IEEE ICC, pp. 322-326, 2008.

[6] A. Azad, S. Alouf, E. Altman, V. Borkar and G. S. Paschos,"Optimal sampling for state change detection with application to the control of sleep mode", in proc. of IEEE CDC, 2009.

[7] Y.-P. Hsu and K.-T. Feng, "Performance Modeling of power saving classes with multiple connections for broadband wireless networks", in proc. of IEEE WCNC, pp. 1477-1482, 2008.

[8] T.-C. Chen and Y.-Y. Chen and J-C. Chen, "‘An efficient energy saving mechanism for IEEE 802.16e Wireless MANs", IEEE Transactions on Wireless Communications, vol. 7, no. 10, pp. 3708-3712, 2008.

[9] L. Kong and D. H. K. Tsang, "Performance study of power saving classes of type I and II in IEEE 802.16e", in proc. of IEEE LCN, pp. 20-27, 2006.

[10] S. Alouf and E. Altman and A. Azad, "M/G/1 queue with repeated inhomogeneous vacations applied to ieee 802.16e power saving", in proc. of SIGMETRICS, pp. 451-452, 2008.

[11] Y. Park and G. U. Hwang, "Performance modeling and analysis of the sleep-mode in IEEE 802.16e WMAN", in proc. of IEEE VTC, pp. 2801-2806, 2007.

[12] L. Kleinrock, “Queueing Systems, Vol. I: Theory”, John Wiley & Sons, 1975.