HAL
open science

Analysis of an M/G/1 queue with repeated inhomogeneous vacations – Application to IEEE 802.16e power saving

Sara Alouf, Eitan Altman, Amar Prakash Azad

To cite this version:

Sara Alouf, Eitan Altman, Amar Prakash Azad. Analysis of an M/G/1 queue with repeated inhomogeneous vacations – Application to IEEE 802.16e power saving. [Research Report] 2008, pp.33. inria-00266552v1

HAL Id: inria-00266552

https://inria.hal.science/inria-00266552v1

Submitted on 24 Mar 2008 (v1), last revised 2 Apr 2008 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Analysis of an $M/G/1$ queue with repeated inhomogeneous vacations Application to IEEE 802.16e power saving

Sara Alouf — Eitan Altman — Amar Prakash Azad

N° ????

March 2008

Thème COM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--????--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
SOPHIA ANTIPOLIS - MÉDITERRANÉE

Analysis of an $M/G/1$ queue with repeated inhomogeneous vacations Application to IEEE 802.16e power saving

Sara Alouf , Eitan Altman , Amar Prakash Azad

Thème COM — Systèmes communicants
Équipe-Projet Maestro

Rapport de recherche n° ???? — March 2008 — 30 pages

Abstract: This report presents a method for analyzing a queueing model with repeated inhomogeneous vacations. At the end of a vacation, the server goes on another vacation, possibly with a *different* probability distribution, if during the previous vacation there have been no arrivals. In order to get an insight on the influence of parameters on the performance, we choose to study a simple $M/G/1$ queue (Poisson arrivals and general independent service times) which has the advantage of being tractable analytically. The theoretical model is applied to the problem of power saving for mobile devices in which the sleep durations of a device correspond to the vacations of the server. Various system performance metrics such as the frame response time and the economy of energy are derived. A constrained optimization problem is formulated to maximize the economy of energy achieved in power save mode, with constraints as QoS conditions to be met. An illustration of the proposed methods is shown with a WiMAX system scenario to obtain design parameters for better performance. Our analysis allows us not only to optimize the system parameters for given traffic intensity but also to propose parameters that provide the best performance under worst case conditions.

Key-words: $M/G/1$ queue with repeated inhomogeneous vacations, power save mode, system response time, constrained optimization, numerical analysis

Centre de recherche INRIA Sophia Antipolis – Méditerranée
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Analyse d'une file $M/G/1$ avec vacances repetées et non-homogènes
Application au mode veille du standard IEEE 802.16e

Résumé : Dans ce rapport, nous analysons une file d'attente dans laquelle le serveur prend des vacances répétées tant que la file est vide. Les vacances peuvent suivre des lois de distribution différentes. Nous considérons une file $M/G/1$ dont la politique de service est exhaustive, ce qui nous permettra de résoudre le modèle de fa con analytique. Nous appliquons ce modèle à l'étude du mode veille disponible chez les équippements sans-fil mobiles. Les périodes de veille d'un équippement correspondent ainsi aux vacances du serveur. Nous trouvons analytiquement le temps de séjour dans le système ainsi que l'économie en énergie (le gain) que le mode veille apporte. Plusieurs problèmes d'optimisation sous contrainte du gain sont alors proposés. Pour illustrer le modèle étudié, nous considérons le mode veille du standard IEEE 802.16e qui fait partie de la famille WiMAX. Nous évaluons numériquement les performances du système et calculons les valeurs optimales des paramètres du protocole afin d'obtenir les meilleures performances dans le cas pire.

Mots-clés : file $M/G/1$ avec vacances répétées non-homogènes, mode veille, temps de séjour, optimisation sous contrainte, analyse numérique

1 Introduction

Power save/sleep mode operation is the key point for energy efficient uses of mobile devices driven by limited battery lifetime. Current standards of Mobile communication such as WiFi, 3G and WiMAX have provisions to operate the mobile station in power save mode in case of low uses scenarios. A mobile operating in power save or sleep mode saves the battery energy and enhances lifetime but it also introduces unwanted delay in serving data packets arriving during a sleep duration. Though energy is a major aspect for handheld devices, delays may also be crucial for various QoS services such as voice and video traffic. Mobility extension of WiMAX [5] is one of the most recent technologies where the sleep mode operation is discussed in detail and is being standardized.

The IEEE 802.16e standard [5] defines 3 types of power saving classes.

- Type I classes are recommended for connections of Best-Effort (BE) and Non-Real Time Variable Rate (NRT-VR) traffic. Under the sleep mode operation, sleep and listen windows are interleaved as long as there is no downlink traffic destined to the node. During listen windows, the node checks with the base station whether there is any buffered downlink traffic destined to it in which case it leaves the sleep mode. Each sleep window is twice the size of the previous one but it is not greater than a specified final value. A node may awaken in a sleep window if it has uplink traffic to transmit.

- Type II classes are recommended for connections of Unsolicited Grant Service (UGS) and Real-Time Variable Rate (RT-VR) traffic. All sleep windows are of the same size as the initial window. Sleep and listen windows are interleaved as in type I classes. However, unlike type I classes, a node may send or receive traffic during listen windows if the requests handling time is short enough.

- Type III classes are recommended for multicast connections and management operations. There is only one sleep window whose size is the specified final value. At the expiration of this window, the node awakens automatically.

The related operational parameters including the initial and maximum sleep window sizes can be negotiated between the mobile node and the base station.

The sleep mode operation of IEEE 802.16e, more specifically the type I power saving class, has received an increased attention recently. In [11], the base station queue is seen as an $M/GI/1/N$ queueing system with multiple vacations; an embedded Markov chain models the successive (increasing in size) sleep windows. Solving for the stationary distribution, the dropping probability and the mean waiting time of downlink packets are computed. Analytical models for evaluating the performance in terms of energy consumption and frame response time are proposed in [12, 13] and supported by simulation results. While [12] considers incoming traffic solely, both incoming and outgoing traffic are considered in [13]. In [4], the authors evaluate the performance of the type I power saving class of IEEE 802.16e in terms of packet delay and power consumption through the analysis of a semi-Markov chain.

Power save mode in systems other than the IEEE 802.16e have also been studied; hereafter we cite some of these studies. In [1], the authors evaluate the energy consumption of various access protocols for wireless infrastructure networks. The sleep mode operation of Cellular Digital Packet Data (CDPD) has been investigated through simulations in [10] and analytically in [9]. To efficiently support short-lived sessions

such as web traffic, a bounded slowdown method – that is similar to type I power saving classes in the IEEE 802.16e – is proposed for the IEEE 802.11 protocol in [8]. Last, the power saving mechanism for the 3G UMTS system is evaluated in [14].

In this report, we propose a queueing-based modeling framework that is general enough to study many of the power save operations described in standards and in the literature. In particular, our model enables the characterization of the performance of type I and type II power saving classes as defined in the IEEE 802.16e standard [5]. The system composed of the base station, the wireless channel and the mobile node is modeled as an $M/G/1$ queue with repeated inhomogeneous vacations. Traffic destined to the mobile node awaits in the base station as long as the node is in power save mode. When the node awakens, the awaiting requests start being served on a first-come-first-served basis. The service consists of the handling of a packet at the base station, its successful transmission over the wireless channel and its handling at the node. Analytical expressions for the distribution and/or the expectation of many performance metrics are derived yielding the expected packet transfer time and the expected gain in energy. We formulate an optimization problem so as to maximize the energy efficiency gain, constrained to meeting some QoS requirements. We illustrate the proposed optimization scheme through four application scenarios.

Although we have motivated our modeling framework using power saving operation in wireless technologies, it is useful whenever the system can be modeled by a server with repeated vacations. The structure of the idle period is general enough to accommodate a large variety of scenarios.

There has been a very rich literature on queues with vacations, see e.g. the survey by Doshi [2]. Our model resembles the one of server with repeated vacations: a server goes on vacation again and again until it finds the queue non-empty. To the best of our knowledge, however, all existing models assume that the vacations are identically distributed whereas our setting applies to inhomogeneous vacations and can accommodate the case when the duration of a vacation increases in the average if the queue is found empty.

The rest of the report is organized as follows. Section 2 describes our system model whose analysis is presented in Sect. 3. Our modeling framework is applied to the power saving mechanism in a WiMAX standard through four scenarios in Sect. 4. Section 5 formulates several performance and optimization problems whose results are shown and discussed in Sect. 6. Section 7 concludes the report and outlines some perspectives.

2 System Model and Notation

Consider an $M/G/1$ queue in which the server goes on vacation for a predefined period once the queue empties. At the end of a vacation period, a new vacation initiates as long as no request awaits in the queue. We consider the exhaustive service regime, i.e., once the server has started serving customers, it continues to serve the queue until the queue empties. Request arrivals are assumed to form a Poisson process, denoted $N(t), t \geq 0$, with rate λ. Let σ denote a generic random variable having the same (general) distribution as the queue service times.

Note that the queue size at the beginning of a busy period impacts the duration of this busy period and is itself impacted by the duration of the last vacation period. Because arrivals are Poisson (a non-negative Lévy input process would have been enough), the queue regenerates each time it empties and the cycles are i.i.d. Each regeneration cycle consists of:

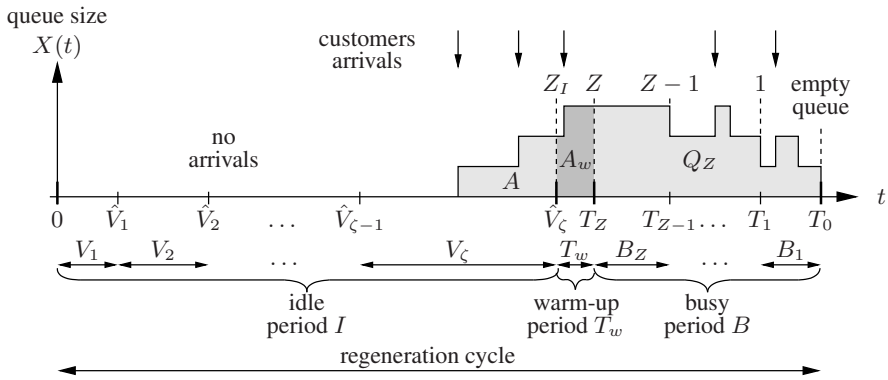


Figure 1: Sample trajectory of the queue size during a regeneration cycle.

1. an *idle* period; let I denote a generic random variable having the same distribution as the queue idle periods, a generic idle period I consists of ζ vacation periods denoted $V_1, \ldots, V_\zeta$;

2. a *warm-up* period; it is a fixed duration denoted T_w during which the server is warming up to start serving requests;

3. a *busy* period; let B denote a generic random variable having the same distribution as the queue busy periods.

The distribution of V_i may depend on i, so the repeated vacations are *not* identically distributed. They are however assumed to be independent.

Let $X(t)$ denote the queue size at time t. It will be useful to define the following instants relatively to the beginning of a generic cycle (in other words, $t = 0$ at the beginning of the generic cycle):

- $\hat{V}_i$ refers to the end of the ith vacation period, for $i = 1, \ldots, \zeta$; observe that the idle period ends at $\hat{V}_\zeta$; we have $\hat{V}_i = \sum_{j=1}^{i} V_j$ and $I = \hat{V}_\zeta = \sum_{i=1}^{\zeta} V_i$;

- T_Z refers to the beginning of the busy period B; we define $Z := X(T_Z)$ as the queue size at the beginning of a busy period;

- T_i refers to the first time the queue size *decreases* to the value i (i.e. $X(T_i) = i$) for $i = Z-1, \ldots, 0$; observe that the cycle ends at T_0.

The times $\{T_i\}_{i=Z,Z-1,\ldots,0}$ delimit Z subperiods in B, as can be seen in Fig. 1. We can write $B = \sum_{i=1}^{Z} B_i$ where $B_i = T_{i-1} - T_i$.

The random variable Z is in fact the number of arrivals from $t = 0$ until time T_Z, even though all of the arrivals occur between $\hat{V}_{\zeta-1}$ and T_Z. Introduce Z_I as the number of requests that have arrived up to time $\hat{V}_\zeta$ (i.e. during period I) and Z_w as the number of arrivals during the warm-up period T_w. Hence $Z = Z_I + Z_w$. Observe that $X(I) = Z_I$.

A possible trajectory of $X(t)$ during a regeneration cycle is depicted in Fig. 1 where we have shown the notation introduced so far. The introduction of the notation A, A_W and Q_Z is deferred until Sect. 3.5.

3 Analysis

This section is devoted to the analysis of the queueing system presented in Sect. 2. We will characterize the distribution of ζ and Z, derive the expectations of ζ, I, Z, B_Z and $X(t)$ and the second moments of I and Z, and last compute the system response time. The gain from idling the server is introduced in the special case when the model is applied to study the power save operation in wireless technologies; see Sect. 4.

3.1 The Number of Vacations

To compute the distribution of ζ, the number of vacation periods during an idle period, we first observe that the event $\zeta \geq i$ is equivalent to the event of no arrivals during $\hat{V}_{i-1} = \sum_{k=1}^{i-1} V_k$.

Let A_k denote the event of no arrivals during the period of time V_k, and let A_k^c denote the complementary event. Denoting by $L_k(s) = \mathrm{E}[\exp(-sV_k)]$ the Laplace Stieltjes transform (LST) of V_k, we can readily write

$$\begin{aligned} P(\zeta = 1) &= P(A_1^c) = \mathrm{E}[\mathbb{1}\{A_1^c\}] = \mathrm{E}[\mathrm{E}[\mathbb{1}\{A_1^c\}|V_1]] \\ &= \mathrm{E}\left[1 - \exp(-\lambda V_1)\right] = 1 - L_1(\lambda), \qquad (1) \\ P(\zeta = i) &= \prod_{k=1}^{i-1} P(A_k)\, P(A_i^c) \\ &= \left(\prod_{k=1}^{i-1} L_k(\lambda)\right)(1 - L_i(\lambda)), \qquad (2) \\ P(\zeta \geq i) &= \prod_{k=1}^{i-1} P(A_k) = \prod_{k=1}^{i-1} L_k(\lambda), \qquad (3) \end{aligned}$$

for $i > 1$, where we have used the fact that arrivals are Poisson with rate λ. The product $\prod_{k=a}^{b} L_k(\lambda)$ is defined as equal to 1 for any $b < a$.

Using (3), the expected number of vacations in an idle period is given by

$$\mathrm{E}[\zeta] = \sum_{i=1}^{\infty} iP(\zeta = i) = \sum_{i=1}^{\infty} P(\zeta \geq i) = \sum_{i=1}^{\infty} \prod_{k=1}^{i-1} L_k(\lambda). \qquad (4)$$

3.2 The Idle Period

Recall that the idle period is $I = \sum_{i=1}^{\zeta} V_i$. It can be rewritten as

$$I = \sum_{i=1}^{\infty} V_i \mathbb{1}\{\zeta \geq i\}.$$

Since the vacation period V_i does not depend on the event of no arrivals during $\hat{V}_{i-1}$, we have for a Poisson arrival process

$$\mathrm{E}[I] = \sum_{i=1}^{\infty} \mathrm{E}[V_i] \prod_{k=1}^{i-1} L_k(\lambda), \qquad (5)$$

where we have used (3). We shall also need the second moment which we derive next. Let us write $I^2 = I_a + 2I_b$ with

$$\begin{aligned} I_a &:= \sum_{i=1}^{\infty} V_i^2 \mathbb{1}\{\zeta \geq i\}, \\ I_b &:= \sum_{i=1}^{\infty}\sum_{j=1}^{i-1} V_i V_j \mathbb{1}\{\zeta \geq i\} \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{i-1} V_i V_j \prod_{k=1}^{i-1} \mathbb{1}\{A_k\}. \end{aligned}$$

Observe that in I_b, only $\mathbb{1}\{A_j\}$ and V_j depend on each other. Using

$$\begin{aligned} \mathrm{E}[V_j \mathbb{1}\{A_j\}] &= \mathrm{E}\Big[\mathrm{E}\left[V_j \mathbb{1}\{A_j\}|V_j\right]\Big] = \mathrm{E}\Big[V_j P(A_j|V_j)\Big] \\ &= \mathrm{E}\Big[V_j \exp(-\lambda V_j)\Big] = -\left.\frac{dL_j(s)}{ds}\right|_{s=\lambda}, \end{aligned}$$

and the LST of V_i introduced earlier, we find after some calculus

$$\begin{aligned} \mathrm{E}[I_a] &= \sum_{i=1}^{\infty} \mathrm{E}\left[V_i^2\right] \prod_{k=1}^{i-1} L_k(\lambda) \qquad (6) \\ \mathrm{E}[I_b] &= \sum_{i=1}^{\infty} \mathrm{E}[V_i] \prod_{k=1}^{i-1} L_k(\lambda) \sum_{j=1}^{i-1} \frac{1}{L_j(\lambda)} \left.\frac{-dL_j(s)}{ds}\right|_{s=\lambda}. \end{aligned}$$

Last, $\mathrm{E}[I^2] = \mathrm{E}[I_a] + 2\mathrm{E}[I_b]$.

3.3 The Initial Queue Size in Busy Periods

The number of requests waiting in the queue at the beginning of a busy period is $Z = Z_I + Z_w$. Since the arrival process is Poisson, it is obvious that Z_w, the number of arrivals during a warm-up period T_w, is a Poisson variable with parameter λT_w. We then have

$$\begin{aligned} \mathrm{E}[Z_w] &= \lambda T_w, \qquad (7) \\ \mathrm{E}\left[Z_w^2\right] &= (\mathrm{E}[Z_w])^2 + \mathrm{V}ar[Z_w] = \lambda^2 T_w^2 + \lambda T_w. \qquad (8) \end{aligned}$$

In order to compute the distribution of Z_I, we will first compute the joint distribution of Z_I and ζ, the number of vacations in an idle period. Observe that Z_I takes value in $\mathbb{N}^*$. We can write

$$\begin{aligned} &P(Z_I = j, \zeta = i) \\ &= P\Big(j \text{ arrivals in } V_i, \mathbb{1}\{A_1, \ldots, A_{i-1}\}\Big) \\ &= P(j \text{ arrivals in } V_i) \prod_{k=1}^{i-1} P\left(\mathbb{1}\{A_k\}\right) \\ &= \mathrm{E}\left[\exp(-\lambda V_i)\frac{(\lambda V_i)^j}{j!}\right] \prod_{k=1}^{i-1} L_k(\lambda). \end{aligned}$$

Therefore

$$\begin{aligned} P(Z_I = j) &= \sum_{i=1}^{\infty} P(Z_I = j, \zeta = i) \\ &= \sum_{i=1}^{\infty} \mathrm{E}\left[\exp(-\lambda V_i)\frac{(\lambda V_i)^j}{j!}\right]\prod_{k=1}^{i-1} L_k(\lambda). \end{aligned}$$

The expected number of arrivals at the end of an idle period is then

$$\begin{aligned} \mathrm{E}[Z_I] &= \sum_{j=1}^{\infty} jP(Z_I = j) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} j\mathrm{E}\left[\exp(-\lambda V_i)\frac{(\lambda V_i)^j}{j!}\right]\prod_{k=1}^{i-1} L_k(\lambda) \\ &= \sum_{i=1}^{\infty} \mathrm{E}\left[\lambda V_i \exp(-\lambda V_i)\sum_{j=0}^{\infty}\frac{(\lambda V_i)^j}{j!}\right]\prod_{k=1}^{i-1} L_k(\lambda) \\ &= \lambda\sum_{i=1}^{\infty}\mathrm{E}[V_i]\prod_{k=1}^{i-1} L_k(\lambda) \\ &= \lambda\mathrm{E}[I] \end{aligned} \tag{9}$$

where we have used (5) to write the last equality. The second moment will also be required. It can be written

$$\begin{aligned} \mathrm{E}\left[Z_I^2\right] &= \sum_{j=1}^{\infty} j^2 P(Z_I = j) \\ &= \sum_{i=1}^{\infty} \mathrm{E}\left[\lambda V_i \exp(-\lambda V_i)\sum_{j=0}^{\infty}(j+1)\frac{(\lambda V_i)^j}{j!}\right]\prod_{k=1}^{i-1} L_k(\lambda) \\ &= \sum_{i=1}^{\infty} \mathrm{E}\left[\lambda V_i \exp(-\lambda V_i)\sum_{j=1}^{\infty} j\frac{(\lambda V_i)^j}{j!}\right]\prod_{k=1}^{i-1} L_k(\lambda) + \mathrm{E}[Z_I] \\ &= \sum_{i=1}^{\infty} \mathrm{E}\left[\lambda^2 V_i^2 \exp(-\lambda V_i)\sum_{j=0}^{\infty}\frac{(\lambda V_i)^j}{j!}\right]\prod_{k=1}^{i-1} L_k(\lambda) + \mathrm{E}[Z_I] \\ &= \lambda^2\sum_{i=1}^{\infty}\mathrm{E}\left[V_i^2\right]\prod_{k=1}^{i-1} L_k(\lambda) + \mathrm{E}[Z_I] \\ &= \lambda^2\mathrm{E}[I_a] + \lambda\mathrm{E}[I] \end{aligned} \tag{10}$$

where we have used (6) and (9) to write the last equality.

Since Z_I and Z_w are independent random variables, we have (using (7)-(10))

$$\mathrm{E}[Z] = \lambda(\mathrm{E}[I] + T_w) \tag{11}$$

$$\begin{aligned} \mathrm{E}[Z^2] &= \mathrm{E}\left[Z_I^2\right] + 2\mathrm{E}[Z_I]\mathrm{E}[Z_w] + \mathrm{E}\left[Z_w^2\right] \\ &= \lambda^2(\mathrm{E}[I_a] + 2\mathrm{E}[I]T_w + T_w^2) + \lambda(\mathrm{E}[I] + T_w). \end{aligned} \tag{12}$$

For future use, we compute

$$\frac{\mathrm{E}[Z^2]}{\mathrm{E}[Z]} = \lambda \frac{\mathrm{E}[I_a] + \mathrm{E}[I]T_w}{\mathrm{E}[I] + T_w} + \lambda T_w + 1. \tag{13}$$

3.4 The Busy Period

Recall from Sect. 2 that a busy period is composed of Z subperiods. These periods are delimited by the times $\{T_i\}_{i=Z,Z-1,\ldots,0}$, the instants at which the queue size $X(t)$ first decreases to a given value $i = Z-1, \ldots, 0$, except for T_Z which denotes the beginning of the busy period; see Fig. 1. The busy period can be expressed as

$$B = \sum_{i=1}^{Z} B_i.$$

Observe that B_1 is nothing but the busy period of a simple $M/G/1$ queue without vacations. The busy periods $\{B_i\}_i$ are i.i.d. and have the same distribution as the busy period of an $M/G/1$ queue. Therefore

$$\begin{aligned} \mathrm{E}[B] &= \mathrm{E}\left[\mathrm{E}[B|Z]\right] \\ &= \mathrm{E}\left[Z\mathrm{E}[B_1]\right] \\ &= \mathrm{E}[Z]\mathrm{E}[B_1]. \end{aligned}$$

Considering the loss free $M/G/1$ queue, we know that the load $\rho := \lambda \mathrm{E}[\sigma]$ is equal to the server utilization $\mathrm{E}[B_1]/(\mathrm{E}[B_1] + 1/\lambda)$. Hence,

$$\mathrm{E}[B_1] = \frac{\mathrm{E}[\sigma]}{1-\rho} \tag{14}$$

and

$$\mathrm{E}[B] = \frac{\rho}{1-\rho}(\mathrm{E}[I] + T_w) \tag{15}$$

where we have used the equality (11); recall that $\mathrm{E}[I]$ is given in (5).

3.5 The Queue Size

In this section, we focus on deriving the expected queue size $\mathrm{E}[X(t)]$.

For convenience, and without loss of generality, we have let $t = 0$ at the beginning of a regeneration cycle. The queue is empty until the first customer arrival in the vacation V_ζ, so $X(t) = 0$ for $0 \le t \le \hat{V}_{\zeta-1}$. After the first arrival, the queue may only increase up to the time T_Z, so $X(t)$ is a non-decreasing step function for $\hat{V}_{\zeta-1} < t \le T_Z$. Also, we have by definition $X(I) = Z_I$ and $X(T_Z) = Z$. After time T_Z, the queue may decrease or increase according to whether a service has ended or a customer has arrived to the queue. We also have by definition that $X(T_i) = i$, for $i = Z, Z-1, \ldots, 0$.

Define

$$A := \int_{\hat{V}_{\zeta-1}}^{\hat{V}_\zeta} X(t)dt, \tag{16}$$

$$A_w := \int_{\hat{V}_\zeta}^{T_Z} X(t)dt,$$

$$Q_Z := \int_{T_Z}^{T_0} X(t)dt, \tag{17}$$

as the total area under the curve $X(t)$ for the idle, warm-up and busy periods respectively, as can be seen in Fig. 1. The subscript Z in Q_Z expresses the fact that the initial queue size is Z.

We can write

$$\mathrm{E}[X] = \frac{\mathrm{E}[A] + \mathrm{E}[A_w] + \mathrm{E}[Q_Z]}{\mathrm{E}[I] + T_w + \mathrm{E}[B]}. \tag{18}$$

The terms in the denominator have already been computed: $\mathrm{E}[I]$ in (5) and $\mathrm{E}[B]$ in (15). Observe that

$$\mathrm{E}[I] + T_w + \mathrm{E}[B] = \frac{\mathrm{E}[I] + T_w}{1-\rho}, \tag{19}$$

which allows to rewrite (18) as follows

$$\mathrm{E}[X] = (1-\rho)\frac{\mathrm{E}[A] + \mathrm{E}[A_w] + \mathrm{E}[Q_Z]}{\mathrm{E}[I] + T_w}. \tag{20}$$

We will now compute the expectations of A, A_w and Q_Z.

Computation of $\mathrm{E}[A]$

To compute the expectation of A, one needs to consider the joint distribution of the arrival process $N(t)$, the number of vacations ζ and the last vacation V_ζ. Observe that ζ and V_ζ are correlated, since the distribution of V_i depends on the value of i. We first compute the expectation with respect to the distribution of $N(t)$, conditioning on $V_\zeta = t$. Let τ_i be the ith arrival epoch. Define

$$A(t) := \mathrm{E}[\alpha(t)|N(t) \geq 1],$$

with

$$\alpha(t) := \int_0^t N(s)ds = \sum_{i=1}^{N(t)} (t - \tau_i) = tN(t) - \sum_{i=1}^{N(t)} \tau_i.$$

According to this definition, the area A defined in (16) is equal to $A(V_\zeta)$. Our objective is thus to compute $\mathrm{E}[A] = \mathrm{E}[A(V_\zeta)]$.

For a given $N(t)$, shuffle the $N(t)$ arrival epochs and denote the resulting variables as the "original" variables. It is known that for a given $N(t)$, (i) the "original" variables are uniformly distributed in $(0, t)$ (i.e. their expected value is $t/2$), (ii) the distribution of the arrival epochs is the distribution of the order statistics corresponding to the "original" variables, and (iii) the sum of the order statistics is the sum of the "original" variables. Therefore, we get that

$$\begin{aligned}
\mathrm{E}[\alpha(t)] &= \mathrm{E}[\mathrm{E}[\alpha(t)|N(t)]] \\
&= \mathrm{E}\left[\mathrm{E}\left[tN(t) - \sum_{i=1}^{N(t)} \tau_i \,\middle|\, N(t)\right]\right] \\
&= \mathrm{E}[tN(t) - N(t)t/2] \\
&= \mathrm{E}[N(t)]t/2 \\
&= \lambda t^2/2,
\end{aligned}$$

and

$$\begin{aligned} A(t) &= \frac{\mathrm{E}[\alpha(t)]}{P(N(t) \geq 1)} \\ &= \frac{\lambda t^2}{2(1-\exp(-\lambda t))} \\ &= \frac{\lambda}{2} t^2 \sum_{k=0}^{\infty} \exp(-k\lambda t). \end{aligned}$$

Hence,

$$\begin{aligned} \mathrm{E}[A] &= \sum_{i=1}^{\infty} P(\zeta = i)\mathrm{E}[A(V_i)] \\ &= \frac{\lambda}{2} \sum_{i=1}^{\infty} P(\zeta = i) \sum_{k=0}^{\infty} \mathrm{E}\left[V_i^2 \exp(-k\lambda V_i)\right] \\ &= \frac{\lambda}{2} \sum_{i=1}^{\infty} \left(\prod_{k=1}^{i-1} L_k(\lambda) \right) (1 - L_i(\lambda)) \sum_{k=0}^{\infty} \left. \frac{d^2 L_i(s)}{ds^2} \right|_{s=k\lambda} \end{aligned} \tag{21}$$

where we have used the LST of V_i and (2).

Computation of $\mathrm{E}[A_w]$

Recall that there are Z_I customers at the beginning of the warm-up period and that $N(t)$ is a Poisson arrival process. We can readily write

$$A_w = Z_I T_w + \int_0^{T_w} N(t)dt,$$

yielding

$$\begin{aligned} \mathrm{E}[A_w] &= E[Z_I]T_w + \int_0^{T_w} \lambda t dt \\ &= \lambda T_w (E[I] + T_w/2), \end{aligned} \tag{22}$$

where we have used (9).

Computation of $\mathrm{E}[Q_Z]$

From the definition (17) and as seen in Fig. 2, the following recursive equation holds

$$\begin{aligned} Q_Z &= \int_{T_Z}^{T_{Z-1}} X(t)dt + \int_{T_{Z-1}}^{T_0} X(t)dt \\ &= \Big((Z-1)\, B_Z + Q_1 \Big) + Q_{Z-1} \end{aligned}$$

whose solution is (recall that the $\{B_i\}_i$ are i.i.d.)

$$Q_Z = Z\, Q_1 + \sum_{i=1}^{Z-1} i B_1 = Z\, Q_1 + Z(Z-1)B_1/2.$$

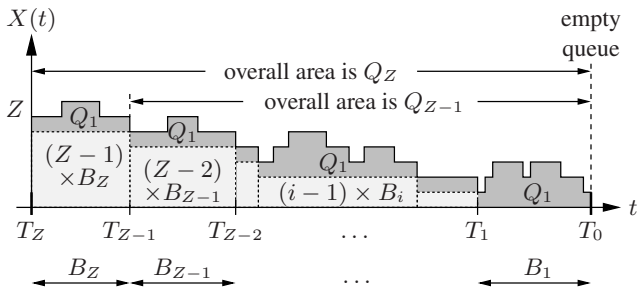


Figure 2: Structure of Q_Z.

Hence,

$$\begin{aligned} \mathrm{E}[Q_Z] &= \mathrm{E}[\mathrm{E}[Q_Z|Z]] \\ &= \mathrm{E}[Z\mathrm{E}[Q_1] + Z(Z-1)\mathrm{E}[B_1]/2] \\ &= \mathrm{E}[Z]\mathrm{E}[Q_1] + (\mathrm{E}[Z^2] - \mathrm{E}[Z])\mathrm{E}[B_1]/2 \end{aligned} \tag{23}$$

The terms $\mathrm{E}[Z]$, $\mathrm{E}[Z^2]$ and $\mathrm{E}[B_1]$ have been derived in (11), (12) and (14) respectively. It remains to compute $\mathrm{E}[Q_1]$ to complete the derivation of $\mathrm{E}[Q_Z]$.

Consider the $M/G/1$ queue without vacation. Its queue size is denoted $X_{M/G/1}(t)$ ans its expected sojourn time is denoted $T_{M/G/1}$. We know that

$$T_{M/G/1} = \frac{\mathrm{E}[X_{M/G/1}]}{\lambda} = \mathrm{E}[\sigma] + \frac{\lambda \mathrm{E}[\sigma^2]}{2(1-\rho)}.$$

where the first equality derives from Little's law, and the second equality comes from the Pollaczek-Khintchine formula (see for instance [7]). Applying renewal theory, we can write

$$\mathrm{E}[X_{M/G/1}] = \frac{\mathrm{E}[Q_1]}{1/\lambda + \mathrm{E}[B_1]}, \qquad \text{where} \quad Q_1 = \int_0^{B_1} X_{M/G/1}(t)dt.$$

Thus, using $1-\rho = (1/\lambda)/(1/\lambda + \mathrm{E}[B_1])$ (loss free system), it comes that

$$\mathrm{E}[Q_1] = \frac{T_{M/G/1}}{1-\rho} = \frac{\mathrm{E}[\sigma]}{1-\rho} + \frac{\lambda \mathrm{E}[\sigma^2]}{2(1-\rho)^2}. \tag{24}$$

Using (14) and (24), we can rewrite (23) as follows

$$\mathrm{E}[Q_Z] = \frac{\mathrm{E}[Z]}{1-\rho}\left(T_{M/G/1} + \frac{\mathrm{E}[\sigma]}{2}\left(\frac{\mathrm{E}[Z^2]}{\mathrm{E}[Z]} - 1\right)\right). \tag{25}$$

The derivation of all elements of $\mathrm{E}[X]$ in (20) is now completed.

3.6 The expected sojourn time

Let T denote the expected system response time or, equivalently, the expected time a customer spends in the queue. It is straightforward to write T using Little's formula

$$T = \frac{\mathrm{E}[X]}{\lambda}, \tag{26}$$

where $\mathrm{E}[X]$ is given in (20). After the replacement of the elements of $\mathrm{E}[X]$ with their respective expressions, the expected sojourn time can be rewritten

$$\begin{aligned} T &= (1-\rho)\frac{\mathrm{E}[A]}{\mathrm{E}[Z]} + (1-\rho)\frac{\mathrm{E}[A_w]}{\mathrm{E}[Z]} + (1-\rho)\frac{\mathrm{E}[Q_Z]}{\mathrm{E}[Z]} \\ &= (1-\rho)\frac{\mathrm{E}[A]}{\mathrm{E}[Z]} + (1-\rho)T_w\left[\frac{\mathrm{E}[I]+T_w/2}{\mathrm{E}[I]+T_w}\right] + T_{M/G/1} + \frac{\mathrm{E}[\sigma]}{2}\left[\frac{\mathrm{E}[Z^2]}{\mathrm{E}[Z]} - 1\right] \\ &= \frac{1/\lambda - \mathrm{E}[\sigma]}{\mathrm{E}[I]+T_w}\mathrm{E}[A] + T_w\frac{\mathrm{E}[I]+T_w/2}{\mathrm{E}[I]+T_w} + \frac{\rho\mathrm{E}[I_a]}{2(\mathrm{E}[I]+T_w)} + T_{M/G/1} \end{aligned} \tag{27}$$

where we have used (11), (13), (22) and (25). Observe that the first three terms of (27) are the contribution of the vacation and warm-up periods to the expected sojourn time.

As the rate $\lambda \to 1/\mathrm{E}[\sigma]$ (recall that the stability condition enforces that $\lambda\mathrm{E}[\sigma] < 1$), we must have $P(\zeta = 1) \to 1$ (thus $L_1(\lambda) \to 0$) whatever the distribution of the vacations. There will then be only one vacation period in most idle periods. Therefore, at large input rates, the largest contribution to the sojourn time is expected to come from the waiting time when the server is active (queueing delays).

4 Application to Power Saving

The model analyzed in Sect. 3 can be used to study energy saving schemes used in wireless technologies. Consider the system composed of the base station, the wireless channel and the mobile node. When the energy saving mechanism is disabled, the system can be seen as an $M/G/1$ queue; and when it is enabled, the system can be modeled as an $M/G/1$ queue with vacations. The server goes on vacations repeatedly until the queue is found non-empty. This models the fact that the mobile node goes to sleep by turning off the radio as long as there is no packets destined to it.

In practice, the mobile needs to turn on the radio to check for packets. The amount of time needed is called the *listen* window and is denoted T_l. During a listen window, the mobile can be informed of any packet that has arrived *before* the listen window. Any arrival during a listen window can only be notified in the following listen window. To comply with this requirement, we will make all but the first vacation periods start with a listen window T_l. The last listen window is included in the warm-up period T_w (in practice we will make $T_w = T_l$).

Let S_i be a generic random variable representing the time for which a node is sleeping during the ith vacation period. We then have $V_1 = S_1$ and $V_i = T_l + S_i$ for $i = 2, \ldots, \zeta$. In this report, we are assuming T_l to be a constant. As for the $\{S_i\}_i$, four cases will be considered as detailed further on. Figure 3 (resp. 4) maps the state of an $M/G/1$ queue (resp. an $M/G/1$ queue with repeated vacations) to the possible states of a mobile node.

4.1 The Energy Gain under Power Saving

The performance metric defined in this section complements the ones derived in Sect. 3, but is specific to applications in wireless networks, and more precisely, to energy saving mechanisms. In this section, we will derive the gain in energy at a node should the power save mechanism be activated.

Having in mind the possible node states, we can distinguish between four possible levels of energy consumption, that are, from highest to lowest,

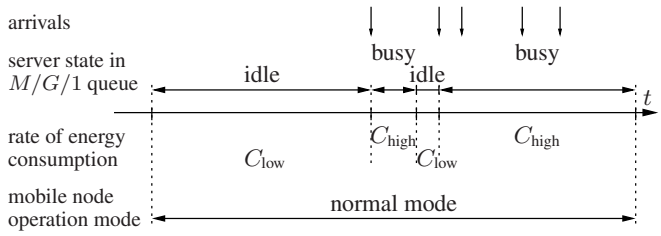


Figure 3: Mapping the $M/G/1$ queue to the normal mode of a mobile node.

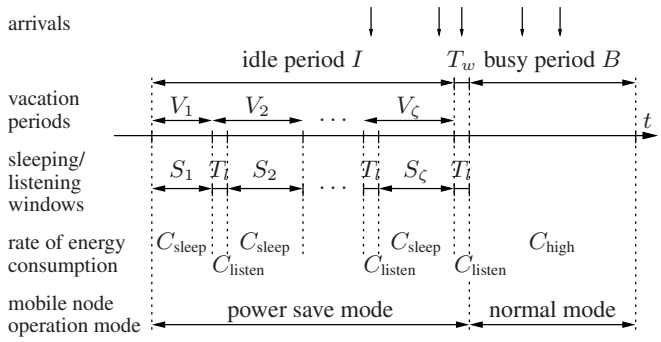


Figure 4: Mapping the $M/G/1$ queue with repeated vacations to the possible states of a mobile node.

- C_{high}; experienced during exchanges of packets,
- C_{listen}; experienced when checking for downlink packets,
- C_{low}; the lowest level observed when the mobile node is inactive, but not in sleep state,
- C_{sleep}; the lowest level observed when the mobile node is in sleep state.

When the power save mechanism is not activated, the energy consumption per unit of time is C_{low} in idle periods (whose expectation is $1/\lambda$) and is equal to C_{high} during the busy periods (whose expectation is $\mathrm{E}[B_1]$). The energy consumption rate can be written

$$E_{\text{no sleep}} := \rho\, C_{\text{high}} + (1-\rho) C_{\text{low}} \tag{28}$$

where $\rho = \lambda \mathrm{E}[\sigma] = \mathrm{E}[B_1]/(1/\lambda + \mathrm{E}[B_1])$ (loss free system).

Consider now the case when the power save mechanism is activated. During busy periods (that are on average equal to $\mathrm{E}[B]$), the energy consumption per unit of time is C_{high}. During idle periods, the consumption is C_{listen} in listen windows (whose length is T_l) and is equal to C_{sleep} the rest of the idle period. Observe that there are on average $\mathrm{E}[\zeta]-1$ listen windows in each idle period; see Fig. 4. The energy consumption rate is

$$\begin{aligned} E_{\text{sleep}} \quad := \quad & \frac{\mathrm{E}[B]}{\mathrm{E}[I]+T_w+\mathrm{E}[B]} C_{\text{high}} + \frac{T_l(\mathrm{E}[\zeta]-1)+T_w}{\mathrm{E}[I]+T_w+\mathrm{E}[B]} C_{\text{listen}} \\ & + \frac{\mathrm{E}[I]-T_l(\mathrm{E}[\zeta]-1)}{\mathrm{E}[I]+T_w+\mathrm{E}[B]} C_{\text{sleep}} \end{aligned}$$

Observe that $\mathrm{E}[B]/(\mathrm{E}[I] + T_w + \mathrm{E}[B]) = \rho = \lambda \mathrm{E}[\sigma]$ because we have assumed an unlimited queue (no overflow losses).

The economy in energy per unit of time should a node enable its power saving mechanism is $E_{\text{no sleep}} - E_{\text{sleep}}$. The relative economy, or the *energy gain* is defined as

$$\begin{aligned} G &:= \frac{E_{\text{no sleep}} - E_{\text{sleep}}}{E_{\text{no sleep}}} \\ &= \frac{1-\rho}{\rho + (1-\rho)\frac{C_{\text{low}}}{C_{\text{high}}}} \left(\frac{C_{\text{low}}}{C_{\text{high}}} - \frac{T_l(\mathrm{E}[\zeta]-1) + T_w}{\mathrm{E}[I] + T_w} \frac{C_{\text{listen}}}{C_{\text{high}}} \right. \\ &\qquad \left. - \frac{\mathrm{E}[I] - T_l(\mathrm{E}[\zeta]-1)}{\mathrm{E}[I] + T_w} \frac{C_{\text{sleep}}}{C_{\text{high}}} \right) \end{aligned} \tag{29}$$

where we have used (15). We expect the battery lifetime to increase by the same factor.

In practice $C_{\text{sleep}} \ll C_{\text{high}}$ so that $\frac{C_{\text{sleep}}}{C_{\text{high}}}$ can be neglected. Letting $T_w = T_l$, the lifetime gain reduces to

$$G = \frac{(1-\rho)\left(\frac{C_{\text{low}}}{C_{\text{high}}} - \frac{T_l \mathrm{E}[\zeta]}{\mathrm{E}[I]+T_l}\frac{C_{\text{listen}}}{C_{\text{high}}}\right)}{\rho + (1-\rho)\frac{C_{\text{low}}}{C_{\text{high}}}}. \tag{30}$$

The energy consumption rate when the power save mechanism is activated is rewritten

$$E_{\text{sleep}} = C_{\text{high}} \left(\rho + \frac{(1-\rho)T_l \mathrm{E}[\zeta]}{\mathrm{E}[I] + T_l} \frac{C_{\text{listen}}}{C_{\text{high}}} \right). \tag{31}$$

All performance metrics found so far have been derived as functions of

- *network* parameters: such as the load ρ, the input rate λ, and the first and second moments of the service time ($\mathrm{E}[\sigma]$ and $\mathrm{E}[\sigma^2]$);
- *physical* parameters: such as the consumption rates C_{low}, C_{high} and C_{listen}, neglecting C_{sleep};
- *combined* physical and network parameters: such as the listen window T_l and warm-up period T_w;
- the LSTs of the vacation periods and their first and second moments.

In the following we will specify the distribution of the sleep windows $\{S_i\}_i$ so as to compute explicitly $\{L_i(s)\}_i$, $\{\mathrm{E}[V_i]\}_i$ and $\{\mathrm{E}[V_i^2]\}_i$.

4.2 Sleep Windows are Deterministic

We will first consider that the sleep windows $\{S_i\}_i$ are deterministic. More precisely, let

$$S_i = a^{\min\{i-1,l\}} T_{\min}, \quad i = 1, 2, \ldots,$$

where $T_{\min}$ is the initial sleep window size, a is a multiplicative factor, and l is the final sleep window exponent or equivalently the number of times the sleep window could be

increased. We call $T_{\min}$, a and l the *protocol* parameters. The LSTs of the vacations periods and their first and second moments can be rewritten

$$L_i(s) = \begin{cases} \exp(-T_{\min}s), & i = 1 \\ \exp(-(a^{\min\{i-1,l\}}T_{\min} + T_l)s), & i = 2, 3, \ldots, \end{cases}$$

$$\mathrm{E}[V_i^n] = \begin{cases} T_{\min}^n, & i = 1 \\ (a^{\min\{i-1,l\}}T_{\min} + T_l)^n, & i = 2, 3, \ldots, \end{cases}$$

for $n = 1, 2$.

We will study two cases so as to model type I and type II saving classes as defined in the IEEE 802.16e standard (see Sect. 1).

Scenario D-I

This scenario is inspired by type I power saving classes. We consider $a > 1$ which implies that the first $l + 1$ sleep windows are all distinct. In particular, the value $a = 2$ is consistent with IEEE 802.16e type I power saving classes.

Scenario D-II

In order to mimic the type II power saving classes of the IEEE 802.16e, we set $a = 1$ in this scenario. Letting $a = 1$ equates the length of all sleep windows. Observe that we could have alternatively let $l = 0$; the resulting sleep windows would then be the same, namely $S_i = T_{\min}$ for any i.

Recall from Sect. 1 that in type II classes, a node may send or receive traffic during listen windows if the requests handling time is short enough. Hence, our model applies to these classes only if we assume that no request is sufficiently small to be served during a listen window T_l.

4.3 Sleep Windows are Exponentially Distributed

As an alternative to deterministic sleep windows, we explore in this section the situation when the sleep window S_i is exponentially distributed with parameter μ_i, for $i = 1, 2, \ldots$. Similar to what was done in Sect. 4.2, we let

$$\mathrm{E}[S_i] = \frac{1}{\mu_i} = a^{\min\{i-1,l\}}T_{\min}, \quad i = 1, 2, \ldots. \tag{32}$$

The LSTs of the $\{V_i\}_i$ and their first and second moments are given below.

$$L_i(s) = \begin{cases} \dfrac{1}{1 + T_{\min}s}, & i = 1 \\ \dfrac{\exp(-sT_l)}{1 + a^{\min\{i-1,l\}}T_{\min}s}, & i = 2, 3, \ldots, \end{cases}$$

$$\mathrm{E}[V_i] = \begin{cases} T_{\min}, & i = 1 \\ a^{\min\{i-1,l\}}T_{\min} + T_l, & i = 2, 3, \ldots, \end{cases}$$

$$\mathrm{E}[V_i^2] = \begin{cases} 2T_{\min}^2, & i = 1 \\ 2a^{2\min\{i-1,l\}}T_{\min}^2 + 2a^{\min\{i-1,l\}}T_{\min}T_l + T_l^2, & i = 2, 3, \ldots. \end{cases}$$

Like in Sect. 4.2, we consider two cases inspired by the first two types of IEEE 802.16e power saving classes.

Scenario E-I

Similarly to what is considered in scenario D-I, we consider multiplicative factors that are larger than 1, in other words, the values $\{\mu_i\}_{i=1,\ldots,l+1}$ are different. When $a > 1$, the sleep windows increase in average over time. For $T_l = 0$ we can find closed-form expressions for all metrics derived in Sect. 3. However, when $T_l > 0$, the expected area $\mathrm{E}[A]$ can only be computed numerically, because of the infinite series composed of the second derivatives of the LSTs; see (21).

Scenario E-II

The last case considered in this report is when the sleep windows are i.i.d. exponential random variables. This can be achieved by letting either $a = 1$ or $l = 0$ in (32). Hence $\mu_i = 1/T_{\min}$ for any i. The LSTs of the $\{V_i\}_i$ and their first and second moments simplify to

$$
\begin{aligned}
L_i(s) &= \begin{cases} \dfrac{1}{1+T_{\min}s} & i=1 \\[2ex] \dfrac{\exp(-sT_l)}{1+T_{\min}s} & i=2,3,\ldots \end{cases} \\
\mathrm{E}[V_i] &= \begin{cases} T_{\min} & i=1 \\[2ex] T_{\min}+T_l & i=2,3,\ldots \end{cases} \\
\mathrm{E}[V_i^2] &= \begin{cases} 2T_{\min}^2 & i=1 \\[2ex] 2T_{\min}^2+2T_{\min}T_l+T_l^2 & i=2,3,\ldots \end{cases}
\end{aligned}
$$

5 Exploiting the Analytical Results

Our model is useful for evaluating performance measures as a function of various network parameters (such as the input rate), and allows us to identify the protocol parameters that mostly impact the system performance. Instances of the expected system response time T and the expected energy gain G are provided in Sect. 6.1.

Beside performance evaluation, we will use our analytical model to solve a large range of optimization problems. Below we propose some optimization problems adapted to various degrees of knowledge on the parameters defining the traffic statistics.

1. **Direct optimization** This approach is useful when the traffic parameters information (e.g. the arrival rate) are directly available, or when we can measure or estimate them. An optimization problem can thus be formulated to maximize the system performance (e.g. the energy gain); see Sect. 5.1 for details.

2. **Average performance.** Given that we know the probability distribution of the traffic parameters then we may obtain the protocol parameters that optimize the expected system performance. This optimization analysis is detailed in Sect. 5.2.

3. **Worst case performance.** In the case where we do not have knowledge of even the statistical distribution of the network parameters, then we can formulate the worst case optimization problem which aims at guaranteeing the optimal performance under worst choice of network parameter. Though this is a more robust optimization approach, it yields a quite pessimistic selection of protocol parameters. Even if we do have knowledge of the statistical distribution, we may have to use a worst case performance in the case that that there is a strict bound on the value of some performance measure. The worst-case analysis will be further detailed in Sect. 5.3.

We propose a multiobjective formulation of the optimization problem, where the performance objectives are the energy consumption (or performance measures directly related to the energy consumption) and the response time. We formulate the multiobjective problem as a constrained optimization one: the energy related criterion will be optimized under a constraint on the expected sojourn time. When the traffic parameters are not directly known, two types of constraints on the expected sojourn time will be considered; in the first case the constraint is with respect to the average performance, and in the second case, it is on the worst case performance.

5.1 Constrained Optimization Problem

The objective is to optimize the protocol parameters defined earlier, namely, the initial window $T_{\min}$, the multiplicative factor a, and the exponent l. We define the following generic non-linear program:

$$\begin{array}{ll} \text{maximize} & G \\ \text{subject to} & T \leq T_{\text{QoS}} \end{array} \tag{33a}$$

or equivalently (recall (29))

$$\begin{array}{ll} \text{minimize} & E_{\text{sleep}} \\ \text{subject to} & T \leq T_{\text{QoS}} \end{array} \tag{33b}$$

where G is given in (30), E_{sleep} is given in (31) and T, the system response time, is given in (27). The program (33) maximizes the energy gain, or equivalently, minimizes the expected energy consumption rate, conditioned on a maximum system response time T_{QoS}. The value of T_{QoS} is application-dependent; it needs to be small for interactive multimedia whereas larger values are acceptable for web traffic.

The decision variables in the above optimization will correspond to one or more protocol parameters. For a given distribution of the sleep windows $\{S_i\}_i$, the expected number of vacations $\mathrm{E}[\zeta]$, the expected idle period $\mathrm{E}[I]$, and subsequently the gain G and the expected energy consumption rate E_{sleep} will depend on the protocol parameters $T_{\min}$, a and l and on the physical parameters C_{low}, C_{high} and C_{listen} (assumed fixed).

We propose four types of applications of the mathematical program (33).

1. In the first, denoted $\mathcal{P}_1$, the decision variable is the initial expected sleep window $T_{\min}$. The parameters a and l are held fixed.

2. The second mathematical program, denoted $\mathcal{P}_2$, has as decision variable the multiplicative factor a whereas $T_{\min}$ and l are given.

3. The decision variable of the third program, denoted $\mathcal{P}_3$, is the exponent l. The parameters $T_{\min}$ and a are given.

4. In the fourth program, denoted $\mathcal{P}_4$, all three protocol parameters are optimized. The corresponding energy gain G is the highest that can be achieved.

These four mathematical programs will be solved considering (i) deterministic and (ii) exponentially distributed sleep windows $\{S_i\}_i$. Instances are provided in Sect. 6.2.

5.2 Expectation Analysis

Assume that the statistical distribution of the arrival process is known. Then we may obtain the protocol parameters that optimize the *expected* system performance. One may want to optimize either the expected energy consumption in power save mode or the economy of energy achieved by activating the power save mode. These problems are not equivalent as was the case in (33) since the energy consumption in normal mode itself also depends on the arrival process.

As already mentioned, we consider two different constraints on the expected sojourn time corresponding to the situations in which the application is sensitive either to the worst case value (hard constraint) or the average value (soft constraint).

Hard Constraints

Here, the application is very sensitive to the delay, so we need to ensure that the constraint on the expected sojourn time is always satisfied no matter the value of λ.

The problem is to find the protocol parameter θ that achieves

$$\begin{array}{l} \min_\theta \sum_\lambda p(\lambda) E_{\text{sleep}}(\lambda, \theta) \\ \text{subject to } T(\lambda, \theta) \leq T_{\text{QoS}} \ \forall \lambda. \end{array} \tag{34}$$

Another problem is to find the protocol parameter θ that achieves

$$\begin{array}{l} \max_\theta \sum_\lambda p(\lambda) G(\lambda, \theta) \\ \text{subject to } T(\lambda, \theta) \leq T_{\text{QoS}} \ \forall \lambda. \end{array} \tag{35}$$

The problems (34) and (35) are not equivalent because G depends also on $E_{\text{no sleep}}$ which itself depends on λ; recall (28). Instances of (35) will be provided in Sect. 6.3.

Soft Constraints

In this optimization problem it is assumed that the application is sensitive only to the expected sojourn time rather than to its worst case value. The objective is to find θ that achieves

$$\begin{array}{l} \min_\theta \sum_\lambda p(\lambda) E_{\text{sleep}}(\lambda, \theta) \\ \text{subject to } \sum_\lambda p(\lambda) T(\lambda, \theta) \leq T_{\text{QoS}}. \end{array} \tag{36}$$

Alternatively, one may want to find θ that achieves

$$\begin{array}{l} \max \theta \sum_\lambda p(\lambda) G(\lambda, \theta) \\ \text{subject to } \sum_\lambda p(\lambda) T(\lambda, \theta) \leq T_{\text{QoS}}. \end{array} \tag{37}$$

Instances of (37) will be provided in Sect. 6.3.

5.3 Worst Case Analysis

When the actual input rate is unknown, then a worst case analysis can be performed to enhance the performance under the considered time constraint. Let θ represent the protocol parameter(s) over which we optimize.

Hard Constraints

Assume the constraint on the expected sojourn time has to be satisfied for any value of λ. The problem then is to find θ that achieves

$$\begin{array}{l} \min_\theta \max_\lambda E_{\text{sleep}}(\lambda, \theta) \\ \text{subject to } T(\lambda, \theta) \leq T_{\text{QoS}} \ \forall \lambda. \end{array} \tag{38}$$

In other words, we want to find the value of θ that improves the worst possible energy consumption.

A different problem consists of finding θ that improves the worst possible gain, namely,

$$\begin{array}{l} \max_\theta \min_\lambda G(\lambda, \theta) \\ \text{subject to } T(\lambda, \theta) \leq T_{\text{QoS}} \ \forall \lambda. \end{array} \tag{39}$$

Observe that the worst possible gain is the one obtained when the traffic input rate tends to 1. Thus $\min_\lambda G(\lambda, \theta) \approx 0$. Therefore, the above problem is meaningful only for a restricted range of small values of λ for which the worst energy gain is far above 0. Instances of (39) will be provided in Sect. 6.3.

Soft Constraints

Here, the application is not very sensitive to the delay, so it is acceptable that the constraint is respected by the average performance. The statistical distribution of the input rate, denoted $p(\lambda)$, is assumed to be known. The problem is to find θ that achieves

$$\begin{array}{l} \min_\theta \max_\lambda E_{\text{sleep}}(\lambda, \theta) \\ \text{subject to } \sum_\lambda p(\lambda) T(\lambda, \theta) \leq T_{\text{QoS}}. \end{array} \tag{40}$$

Again, a different objective can be desired, namely to maximize the worst gain. Like what was mentioned in the previous section, the problem is meaningful only when the rate λ is small.

$$\begin{array}{l} \max_\theta \min_\lambda G(\lambda, \theta) \\ \text{subject to } \sum_\lambda p(\lambda) T(\lambda, \theta) \leq T_{\text{QoS}}. \end{array} \tag{41}$$

Instances of (41) will be provided in Sect. 6.3.

6 Results and Discussion

We have performed an extensive numerical analysis to evaluate the performance of the system in terms of the expected system response time T given in (26) and the expected energy gain G given in (30); see Sect. 6.1. In addition we have solved the problems $\mathcal{P}_1$–$\mathcal{P}_4$ for given values of the protocol parameters held fixed; see Sect. 6.2. Instances of the problems (35), (37), (39) and (41) are also provided; see Sect. 6.3.

Physical and network parameters have been selected as follows:

$$\begin{array}{ll} C_{\text{low}}/C_{\text{high}} = 0.2 & \mathrm{E}[\sigma] = 1 \\ C_{\text{listen}}/C_{\text{high}} = 0.2 & \mathrm{E}[\sigma^2] = 2 \\ T_l = 1 & T_w = 1 \\ T_{\text{QoS}} = 50/100 & \end{array}$$

Unless otherwise specified, the protocol parameters are set to the *default* values: $T_{\min} = 2$, $a = 2$ and $l = 9$ in scenarios D-I and D-II, and $T_{\min} = 2$, $a = 1$ and $l = 0$ in scenarios E-I and E-II.

We have varied λ in the interval $(0, 1)$, $T_{\min}$ in $(1, 100)$, and a in $(1, 10)$. The parameter l takes integer values in the interval $(0, 10)$.

6.1 Performance Evaluation

We have evaluated numerically the expected sojourn time T and the expected energy gain G in all four scenarios defined in Sects. 4.2 and 4.3, varying the input rate λ and the three protocol parameters $T_{\min}$, a and l. Our results will be presented in the following sections. First, we discuss the impact of each of the three parameters on the performance of the system in terms of T and G: impact of $T_{\min}$ in Sect. 6.1.1, impact of a in Sect. 6.1.2, and impact of l in Sect. 6.1.3. Then, we comment on each of the performance metrics: comments on T are in Sect. 6.1.4, and comments on G are in Sect. 6.1.5.

6.1.1 Impact of the initial window size $T_{\min}$

We will first investigate the impact that the initial window size $T_{\min}$ has on the performance of the system. For reasons that will be made clear later, this parameter is foreseen to be the most important parameter in type I like power saving classes (scenarios D-I and D-II) and it is the unique parameter in type II like power saving classes (scenarios E-I and E-II).

Type I like power saving classes We set $a = 2$ and $l = 9$ in scenarios D-I and E-I. The results are graphically reported in Fig. 5.

Figures 5(a) and 5(b) respectively depict the expected sojourn time T against the traffic input rate λ and the multiplicative factor a when sleep windows are deterministic and exponentially distributed. The energy gain under the same conditions is depicted in Figs. 5(c) and 5(d).

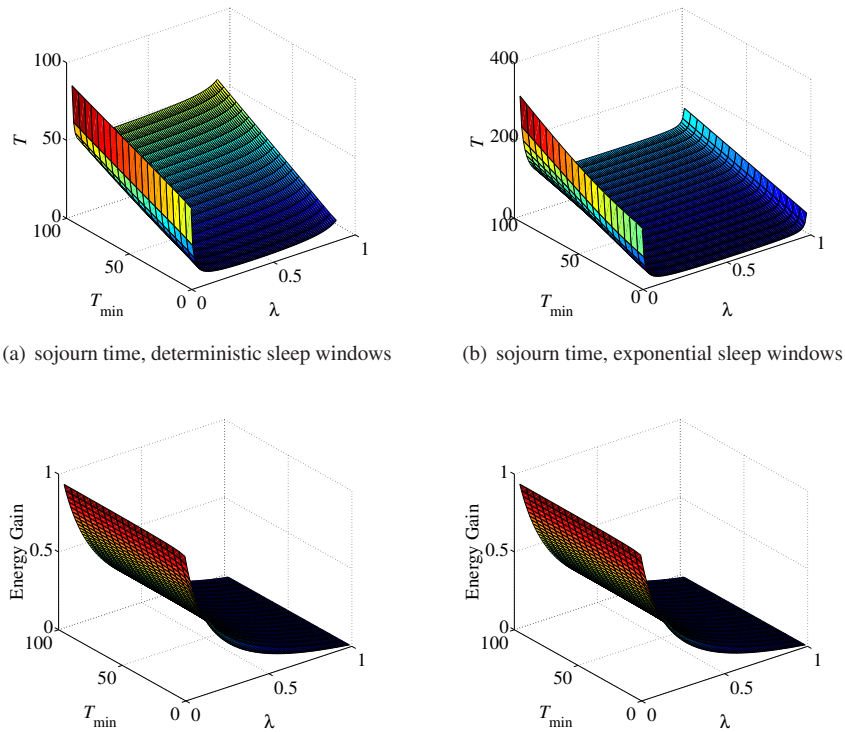
The size of the initial sleep window has a large impact on T for any value of λ. More precisely, T increases linearly with an increasing $T_{\min}$ for any λ; see Figs. 5(a), 5(b). As for the gain G, it is not impacted by $T_{\min}$, except for a small degradation at very small values of $T_{\min}$, hardly visible in Figs. 5(c) and 5(d).

Type II like power saving classes We set $a = 1$ and $l = 0$ in scenarios D-II and E-II. The results are graphically reported in Fig. 6.

Figures 6(a) and 6(b) respectively depict the expected sojourn time T against the traffic input rate λ and the multiplicative factor a when sleep windows are deterministic and exponentially distributed. The energy gain under the same conditions is depicted in Figs. 6(c) and 6(d).

About the impact of $T_{\min}$ on T and G, we can make similar observations to those made for type I like power saving classes, to the only exception that here the degradation of G at very small values of $T_{\min}$ is more visible, especially in Fig. 6(c).

Observe that a larger $T_{\min}$ yields a larger sleep time but it also reduces $\mathrm{E}[\zeta]$ which together explains why the impact on the energy gain is not significant.



(a) sojourn time, deterministic sleep windows

(b) sojourn time, exponential sleep windows

(c) energy gain, deterministic sleep windows

(d) energy gain, exponential sleep windows

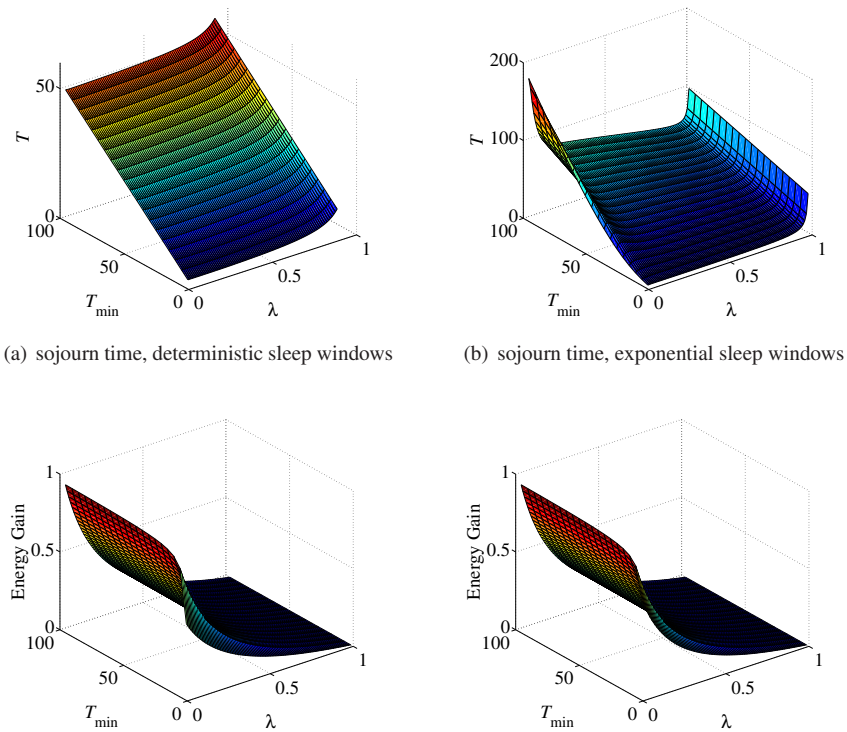
Figure 5: Impact of $T_{\min}$ on T and G in type I like power saving classes.

6.1.2 Impact of the multiplicative factor a

The second parameter used in type I like power saving classes (scenarios D-I and E-I) is the multiplicative factor a. In order to assess the impact of a on the performance of the system, we perform a numerical analysis in which the initial window size is $T_{\min} = 2$, the exponent is $l = 9$ and the multiplicative factor a is varied from 1 to 10. We evaluate the expected sojourn time T and the energy gain G both for deterministic (scenario D-I) and exponentially distributed (scenario E-I) sleep windows. We report the results in Fig. 7.

Figures 7(a) and 7(c) respectively depict the expected sojourn time T and the energy gain G against the traffic input rate λ and the multiplicative factor a when sleep windows are deterministic. The results obtained when the sleep windows are exponentially distributed are displayed in Figs. 7(b) and 7(d).

Interestingly enough, the multiplicative factor a does not impact the gain G. It impacts greatly T but only at very low input rates. Observe that T increases exponentially with an increasing a for small λ which is reflected in Figs. 7(a) and 7(b).



(a) sojourn time, deterministic sleep windows

(b) sojourn time, exponential sleep windows

(c) energy gain, deterministic sleep windows

(d) energy gain, exponential sleep windows

Figure 6: Impact of $T_{\min}$ on T and G in type II like power saving classes.

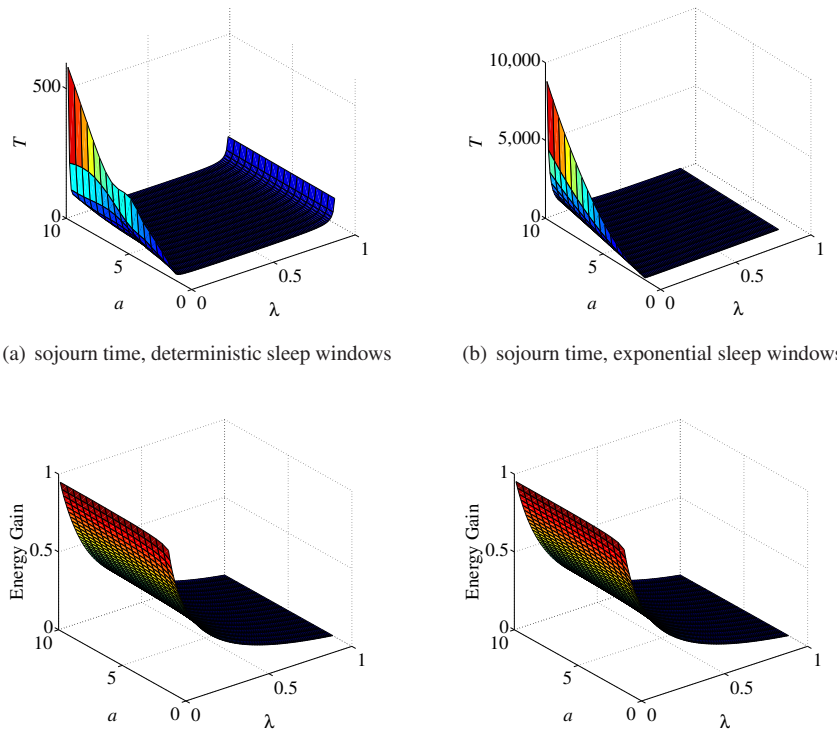
6.1.3 Impact of the exponent l

The third and last parameter used in type I like power saving classes (scenarios D-I and D-II) is the exponent l. In order to assess the impact of the maximum sleep window size on the performance of the system, we perform a numerical analysis in which the multiplicative factor is $a = 2$, the initial window size is $T_{\min} = 2$ and the exponent l is varied from 0 to 10. We evaluate the expected sojourn time T and the energy gain G both for deterministic (scenario D-I) and exponentially distributed (scenario E-I) sleep windows. We report the results in Fig. 8.

Figures 8(a) and 8(c) respectively depict the expected sojourn time T and the energy gain G against the traffic input rate λ and the exponent l when sleep windows are deterministic. The results obtained when the sleep windows are exponentially distributed are displayed in Figs. 8(b) and 8(d).

Alike the multiplicative factor, the exponent l has a large impact on T only for a very low traffic input rate, and has no impact on G whatever the rate λ.

Observe in Fig. 8(a) that T becomes almost insensitive to l beyond $l = 7$ (for small λ). Here the initial vacation window $T_{\min}$ is 2. We have computed T considering larger values of $T_{\min}$, and have observed that T saturates faster with l when the initial sleep window is larger. A similar behavior is observed in the exponential case for higher T; cf. Fig. 8(b).



(a) sojourn time, deterministic sleep windows

(b) sojourn time, exponential sleep windows

(c) energy gain, deterministic sleep windows

(d) energy gain, exponential sleep windows

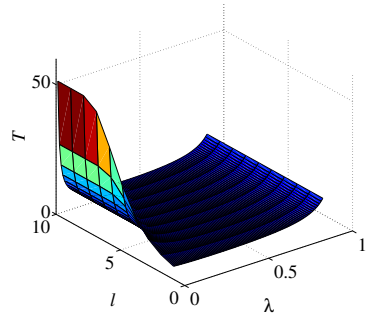
Figure 7: Impact of a on T and G with either deterministic or exponential $\{S_i\}_i$.

6.1.4 The expected sojourn time T

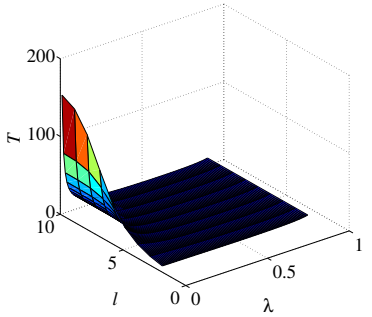
The numerical results of the expected sojourn time T are reported in Figs. 5–8, parts (a) and (b). As already mentioned, T is fairly insensitive to parameters l and a except for very small values of λ. However, T increases linearly as $T_{\min}$ increases.

In scenarios D-I, E-I and E-II, as λ increases, T first decreases rapidly then becomes fairly insensitive to λ up to a certain point beyond which T increases abruptly. This can easily be explained. The sojourn time is essentially composed of two main components: the delay incurred by the vacations of the server and the queueing delay once the server is active. As the input rate increases, the first component decreases while the second one increases. For moderate values of λ, both components balance each other yielding a fairly insensitive sojourn time. The large value of T at small λ is mainly due to the ratio $\mathrm{E}[I_a]/\mathrm{E}[I]$ (recall (27)), whereas the abrupt increase in T at large λ is due to the term $\frac{\lambda \mathrm{E}[\sigma^2]}{2(1-\rho)}$, which is the waiting time in the $M/G/1$ queue without vacations.

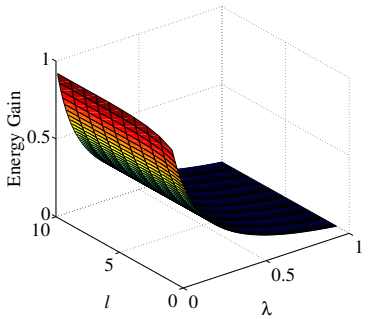
The situation in scenario D-II is different in that T is not large at small input rates λ. Recall that in this scenario, all sleep window are equal to a constant $T_{\min}$. As a consequence, the delay incurred by the vacations of the server is not as large as in the other scenarios. The balance between the two main components of the sojourn time stretches down to small values of λ.



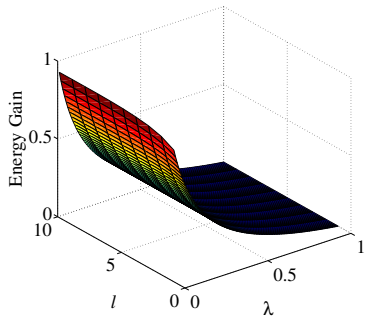
(a) sojourn time, deterministic sleep windows



(b) sojourn time, exponential sleep windows



(c) energy gain, deterministic sleep windows



(d) energy gain, exponential sleep windows

Figure 8: Impact of l on T and G with either deterministic or exponential $\{S_i\}_i$.

6.1.5 The expected energy gain G

The numerical results of the expected energy gain G are reported in Figs. 5–8, parts (c) and (d). As already mentioned, G is insensitive to parameters l and a for any λ, and sensitive to $T_{\min}$ up to a certain initial sleep window size.

The expected energy gain G decreases monotonically as λ increases which can be explained as follows. The larger the input traffic rate λ, the shorter we expect the idle time to be and hence the smaller the gain.

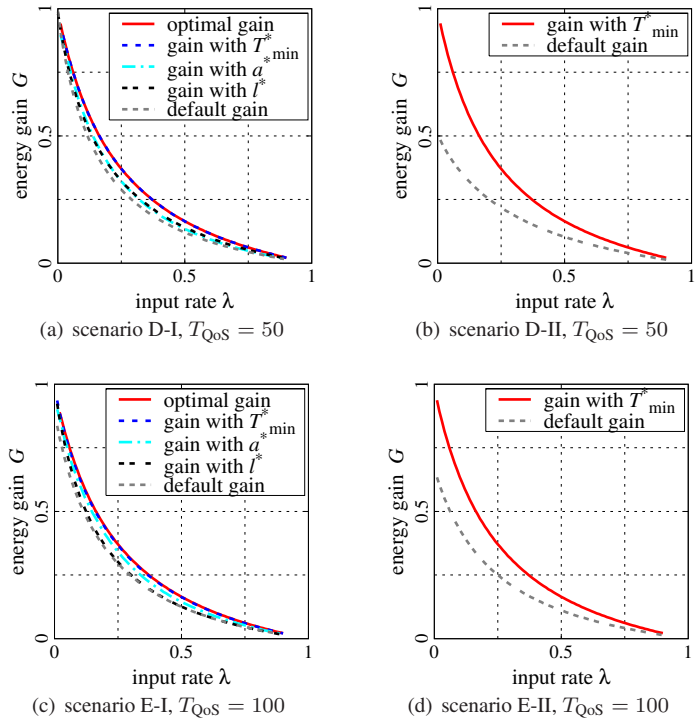
6.2 Constrained Optimization Problem

We have solved the constraint optimization program as depicted in sec. 5.1.

- $\mathcal{P}_1$ for $T^*_{\min}$ when $a = 2$ and $l = 9$ (default values) with $T_{\text{QoS}} = 50$ for scenario D-I and $T_{\text{QoS}} = 100$ for scenario E-I, and when $a = 1$ or $l = 0$ with $T_{\text{QoS}} = 50$ for scenario D-II and $T_{\text{QoS}} = 100$ for scenario E-II;
- $\mathcal{P}_2$ for a^* with $T_{\min} = 2$ and $l = 9$ (default values) with $T_{\text{QoS}} = 50$ for scenario D-I and $T_{\text{QoS}} = 100$ for scenario E-I;
- $\mathcal{P}_3$ for l^* when $T_{\min} = 2$ and $a = 2$ (default values) with $T_{\text{QoS}} = 50$ for scenario D-I and $T_{\text{QoS}} = 100$ for scenario E-I;

- $\mathcal{P}_4$ for $(T_{\min}, a, l)^*$ with $T_{\text{QoS}} = 50$ for deterministic sleep windows and $T_{\text{QoS}} = 100$ for exponential sleep windows.

The optimal gain achieved by the four programs $\mathcal{P}_1$–$\mathcal{P}_4$ and the gain obtained when using the default values are illustrated in Fig. 9 against the input rate λ, for deterministic (Figs. 9(a) and 9(b)) and exponential (Figs. 9(c) and 9(d)) sleep windows. The right-hand-side graphs depict the optimal gain (returned by program $\mathcal{P}_1$ when $a = 1$) and the gain achieved under the default protocol parameter ($T_{\min} = 2$).



(a) scenario D-I, $T_{\text{QoS}} = 50$ (b) scenario D-II, $T_{\text{QoS}} = 50$

(c) scenario E-I, $T_{\text{QoS}} = 100$ (d) scenario E-II, $T_{\text{QoS}} = 100$

Figure 9: Maximized/default gain versus the input rate λ.

The most relevant observation to be made on each of Figs. 9(a) and 9(c) is the match between the curve labeled "optimal gain" (result of program $\mathcal{P}_4$) and the curve labeled "gain with $T^*_{\min}$" (result of program $\mathcal{P}_1$). The interest of this observation comes from the fact that $\mathcal{P}_4$ involves a multivariate optimization whereas $\mathcal{P}_1$ is a much simpler single variate program.

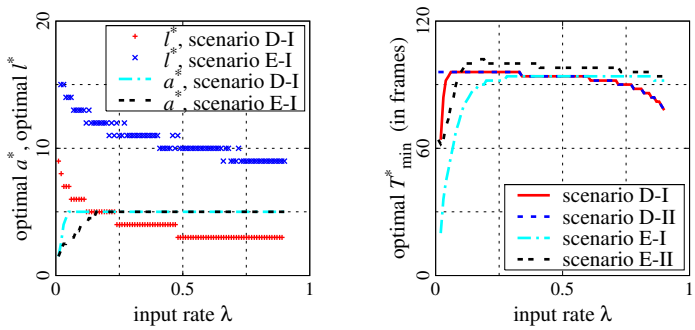
The explanation for this match is as follows. The program $\mathcal{P}_1$ is being solved for the optimal $T_{\min}$. It thus quickly reduces the number of vacations $\mathrm{E}[\zeta]$ to 1 (refer to Fig. 11) and thereby makes the role of both a and l insignificant. Hence, the energy gain maximized by $\mathcal{P}_1$ tends to the optimal gain returned by $\mathcal{P}_4$.

The values of the optimal protocol parameters returned by programs $\mathcal{P}_1$–$\mathcal{P}_3$ are given in Fig. 10 and Table 1. Those returned by program $\mathcal{P}_4$ can be found in Table 2.

Comparing the optimal values of $T_{\min}$ as returned by programs $\mathcal{P}_1$ and $\mathcal{P}_4$ in the deterministic case (cf. column 2 in Table 1 and column 2 in Table 2), it appears that they are very close to each other, confirming our argument that the single variate $\mathcal{P}_1$ is a very good approximation of the multivariate optimization involved in $\mathcal{P}_4$.

Table 1: Optimal values of the protocol parameters from programs $\mathcal{P}_1$–$\mathcal{P}_3$

	$T^*_{\min}$ from $\mathcal{P}_1$				a^* from $\mathcal{P}_2$		l^* from $\mathcal{P}_3$	
λ	D-I	D-II	E-I	E-II	D-I	E-I	D-I	E-I
0.02	64	96	20	62	1.5	2.0	15	8
0.03	84	96	36	64	2.5	2.5	15	7
0.04	92	96	44	70	4.0	2.5	14	7
0.05	94	96	50	74	4.5	3.0	14	6
0.10	96	96	76	96	5.0	4.0	13	6
0.20	96	96	92	100	5.0	5.0	12	5
0.40	94	94	94	100	5.0	5.0	11	4
0.60	94	94	94	98	5.0	5.0	10	3
0.80	88	88	94	96	5.0	5.0	9	3
0.90	78	78	92	94	5.0	5.0	9	3



(a) optimal a^* and l^* versus λ (b) optimal $T^*_{\min}$ versus λ

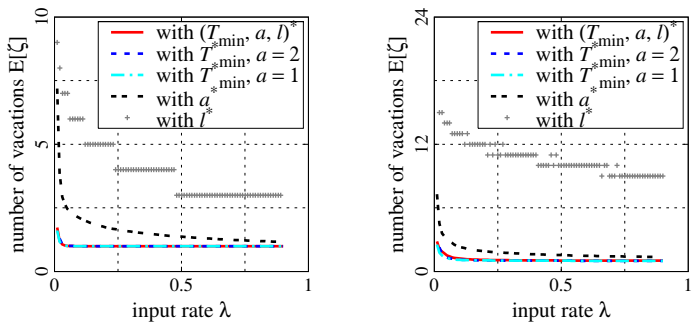
Figure 10: Optimal values of the protocol parameters from programs $\mathcal{P}_1$–$\mathcal{P}_3$.

When maximizing the gain by optimizing $T_{\min}$ (program $\mathcal{P}_1$; see Fig. 10(b)), we observe in all scenarios but scenario D-II that, optimally, $T_{\min}$ should first increase with the input rate λ then decrease with increasing λ for large values of λ. This observation is rather counter-intuitive and we do not have an explanation for it at the moment. Our intuition that $T_{\min}$ should decrease as λ increases is confirmed only in scenario D-II.

Looking at the expected number of vacations $\mathrm{E}[\zeta]$, should the optimal value $T^*_{\min}$ be used, it appears that $\mathrm{E}[\zeta]$ decreases asymptotically to 1 as λ increases; see Fig. 11. The reason behind this is the energy consumption during listen windows and warm-up periods. To maximize the energy gain, one could minimize the factor multiplying C_{sleep}, in other words minimize $\mathrm{E}[\zeta]$. As a consequence, if $T_{\min}$ is optimally selected, then the initial sleep window will be set large enough so that the server will rarely go for a second vacation period, thereby eliminating the unnecessary energy consumption incurred by potential subsequent listen windows. As a consequence, the multiplicative factor a and the exponent l will have a negligible effect on the performance of the system.

Table 2: Optimal values of the protocol parameters from program $\mathcal{P}_4$

	Deterministic case			Exponential case		
λ	$T_{\min}$	a	l	$T_{\min}$	a	l
0.02	72	1.5	1	27	2.5	2
0.03	82	1.5	5	22	3.0	2
0.04	92	1.5	4	22	3.0	2
0.05	92	2.0	3	32	1.5	3
0.10	92	5.0	1	42	1.5	1
0.20	92	5.0	1	47	1.5	9
0.40	92	1.5	1	47	1.5	8
0.60	92	1.5	1	47	1.5	7
0.80	87	1.5	1	47	1.5	1
0.90	77	1.5	1	42	2.0	6



(a) deterministic sleep windows (b) exponential sleep windows

Figure 11: Expected number of vacations $\mathrm{E}[\zeta]$ versus λ when the protocol parameters are optimally set.

6.3 Worst Case and Expectation Analysis

In this section, we report the results of a worst case and an expectation analysis, considering the expected energy gain as performance metric. We will solve the problems stated in (35), (37), (39) and (41). The decision variable is the initial sleep window size $T_{\min}$. Each problem is solved for each of the four scenarios defined in Sects. 4.2 and 4.3. We consider $a = 2$ and $l = 9$ in scenarios D-I and E-I. Recall that we necessarily have $a = 1$ and $l = 0$ in scenarios D-II and E-II. We consider that λ may take five different values. These values and the corresponding probabilities $p(\lambda)$ are given in Table 3. The values of the parameter $T_{\min}$ found for each of the problems are reported in Table 4.

7 Conclusion and Perspectives

In this report, we have analyzed the $M/G/1$ queue with repeated inhomogeneous vacations. In all prior work, repeated vacations are assumed to be i.i.d., whereas in our model the duration of a repeated vacation can come from an entirely different distribu-

Table 3: Distribution of the input rate λ

λ	0.02	0.05	0.1	0.2	0.5
$p(\lambda)$	0.3125	0.3125	0.1875	0.1250	0.0625

Table 4: Expectation/worst-case analysis: value of $T_{\min}$ (in number of frames)

	Expectation analysis		Worst-case analysis	
Scenario	hard constraint	soft constraint	hard constraint	soft constraint
D-I	65	92	64	64
D-II	96	97	94	94
E-I	22	50	21	21
E-II	69	79	62	62

tion. Using transform-based analysis, we have derived various performance measures of interest such as the expected system response time and the gain from idling the server. We have applied the model to study the problem of power saving for mobile devices. The impact of the power saving strategy on the network performance is easily studied using our analysis. We have formulated various constrained optimization problems aimed at determining optimal parameter settings. We have performed an extensive numerical analysis to illustrate our results, considering four different strategies of power saving having either deterministic or exponentially distributed sleep durations. We have found that the parameter that most impacts the performance is the initial sleep window size. Hence, optimizing this parameter solely is enough to achieve quasi-optimal energy gain.

In this report, we have focused on deriving the expected sojourn time. However, it is possible to derive stronger results in means of the distribution of the sojourn time using the decomposition properties obtained in [3] and the distributional form of Little's law [6]. The queue length decomposition property [3] states that the queue length in an $M/G/1$ queue with vacations at an arbitrary epoch (i.e. in stationary regime) is distributed as the independent sum of (i) the queue length in the corresponding $M/G/1$ queue without vacations and (ii) the queue length in the $M/G/1$ queue with vacations at an arbitrary epoch during a non-busy interval. Given that our vacations are inhomogeneous, a significant portion of the derivations shall need to be repeated. However, we think it is worthy to investigate this approach and plan to do so in the near future.

Other important research directions are considered. Namely,

Other traffic profiles. It is interesting to consider more bursty real time traffic as well as TCP traffic. We expect that much of this work may have to be performed through simulations as the queueing analysis may become intractable. It is important to examine how our optimized parameters perform when a new type of traffic is introduced, and whether our robust design for the worst case Poisson traffic maintains its robustness beyond the Poisson arrival processes.

Extensions of the protocol. So far our analysis enabled us to optimize parameters of the protocol. It is of interest to go beyond the optimization and to examine extensions or improvements of the protocol that would require to extend the theoretical framework as well. In particular we intend to examine rendering $T_{\min}$ dynamic, by choosing its value at the nth idle time as a function of the V_ζ (or of its expectation) in the $(n-1)$-th idle time.

References

[1] J. C. Chen, K. M. Sivalingam, and P. Agrawal. Performance comparison of battery power consumption in wireless multiple access protocols. *ACM Wireless Networks*, 5(6):445–460, December 1999.

[2] B. T. Doshi. Queueing systems with vacations - a survey. *Queueing Systems - Theory and Applications*, 1(1):29–66, 1986.

[3] S. W. Fuhrmann and R. B. Cooper. Stochastic decomposition in the M/G/1 queue with generalized vacation. *Operations Research*, 33(5):1117–1129, September-October 1985.

[4] K. Han and S. Choi. Performance analysis of sleep mode operation in IEEE 802.16e mobile broadband wireless access systems. In *Proc. of IEEE VTC 2006-Spring*, volume 3, pages 1141–1145, Melbourne, Australia, May 2006.

[5] IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems. *IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005 (Amendment and Corrigendum to IEEE Std 802.16-2004)*, 2006.

[6] J. Keilson and L. D. Servi. A distribution form of little's law. *Operations Research Letters*, 7(5):223–227, 1983.

[7] L. Kleinrock. *Queueing Systems: Theory*, volume 1. John Wiley and Sons, 1975.

[8] R. Krashinsky and H. Balakrishnan. Minimizing energy for wireless web access with bounded slowdown. In *Proc. of ACM MobiCom '02*, pages 119–130, Atlanta, Georgia, USA, September 2002.

[9] S. J. Kwon, Y. W. Chung, and D. K. Sung. Queueing model of sleep-mode operation in cellular digital packet data. *IEEE Transactions on Vehicular Technology*, 52(4):1158–1162, July 2003.

[10] Y. B. Lin and Y. M. Chuang. Modeling the sleep mode for cellular digital packet data. *IEEE Communication letters*, 3(3):63–65, March 1999.

[11] 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 *Proc. of IEEE VTC 2004-Fall*, volume 2, pages 1169–1173, Los Angeles, California, USA, September 2004.

[12] Y. Xiao. Energy saving mechanism in the 802.16e wireless MAN. *IEEE Communication letters*, 9(7):595–597, July 2005.

[13] Y. Xiao. Performance analysis of an energy saving mechanism in the IEEE 802.16e wireless MAN. In *Proc. of IEEE CCNC 2006*, volume 1, pages 406–410, January 2006.

[14] S. R. Yang and Y. B. Lin. Modeling UMTS discontinuous reception mechanism. *IEEE Transactions on Wireless Communications*, 4(1):312–319, January 2005.

Centre de recherche INRIA Sophia Antipolis – Méditerranée
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399