



**HAL**  
open science

## M/G/1 queue with repeated inhomogeneous vacations applied to IEEE 802.16e power saving

Sara Alouf, Eitan Altman, Amar Prakash Azad

► **To cite this version:**

Sara Alouf, Eitan Altman, Amar Prakash Azad. M/G/1 queue with repeated inhomogeneous vacations applied to IEEE 802.16e power saving. 2008 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, Jun 2008, Annapolis, Maryland, United States. pp.451-452, 10.1145/1375457.1375516 . hal-00641082

**HAL Id: hal-00641082**

**<https://inria.hal.science/hal-00641082v1>**

Submitted on 12 Jul 2019

**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.

# M/G/1 Queue with Repeated Inhomogeneous Vacations Applied to IEEE 802.16e Power Saving

Sara Alouf, Eitan Altman, and Amar Prakash Azad  
INRIA, 2004 Route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex, France

## 1 Introduction and Model

We shall analyze and optimize the parameters of the following two types of power saving of IEEE 802.16e in presence of downlink traffic:

(i) Type I classes: 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. The initial sleep window is  $T_{\min}$ , the multiplicative factor is  $a$  ( $a = 2$  in the standard) and the final value  $a^l T_{\min}$  depends on the exponent  $l$ .

(ii) Type II classes: All sleep windows are of the same size as the initial window (i.e.  $a = 1$ ). Sleep and listen windows are interleaved as in type I classes.

To model these classes, we analyze an M/G/1 queue in which the server begins a vacation of random length each time that the system becomes empty. If the server returns from a vacation to find an empty queue, a new random vacation initiates; otherwise, the server works until the system empties (exhaustive service regime). Request arrivals are assumed to form a Poisson process with parameter  $\lambda$ . Let  $\sigma$  denote a generic random variable having the same (general) distribution as the queue service times. The queue regenerates each time it empties and the cycles are i.i.d. Each regeneration cycle consists of (see Fig. 1):

(i) 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  $\zeta$  vacation periods denoted  $V_1, \dots, V_\zeta$ ;

(ii) a *warm-up* period; it is a fixed duration denoted  $T_w$  during which the server is warming up to start serving requests;

(iii) a *busy* period; let  $B$  denote a generic random variable having the same distribution as the queue busy periods.

Let  $X(t)$  denote the queue size at time  $t$ . Define

- (i)  $\hat{V}_i$ : the end of the  $i$ th vacation period,  $i = 1, \dots, \zeta$ ; the idle period ends at  $\hat{V}_\zeta$ ; we have  $\hat{V}_i = \sum_{j=1}^i V_j$  and  $I = \hat{V}_\zeta = \sum_{i=1}^\zeta V_i$ ;
- (ii)  $T_Z$ : the start of the busy period  $B$ ; let  $Z := X(T_Z)$ , the queue size at the beginning of a busy period;
- (iii)  $T_i$ : the first time the queue size *decreases* to the value  $i$  (i.e.  $X(T_i) = i$ ) for  $i = Z - 1, \dots, 0$ ; the cycle ends at  $T_0$ .

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

$Z$  is in fact the number of arrivals from  $t = 0$  until time  $T_Z$ . 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  $X(I) = Z_I$ .

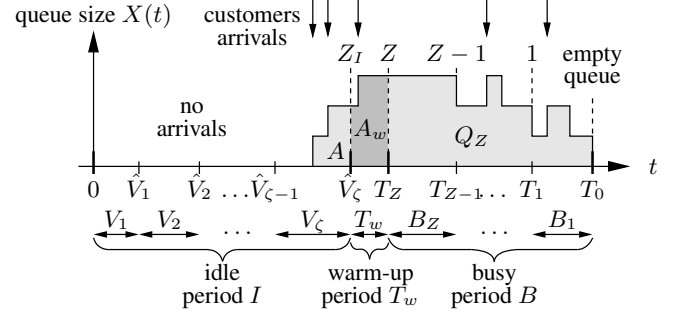


Figure 1: Sample trajectory of the queue size during a regeneration cycle.

## 2 Analysis

**Number of Vacations.** Let  $\zeta$  be the number of vacation periods during an idle period. Observe that the event  $\zeta \geq i$  is equivalent to the event of no arrivals during any of the periods  $\{V_k\}_{k=1, \dots, i-1}$ . Denoting by  $L_k(s) = E[\exp(-sV_k)]$  the Laplace Stieltjes transform of  $V_k$ , we have

$$P(\zeta \geq i) = \prod_{k=1}^{i-1} L_k(\lambda),$$

for  $i > 1$ , where we have used the fact that arrivals are Poisson with rate  $\lambda$ .

**The Idle and Busy Period.** We can write

$$I = \sum_{i=1}^{\zeta} V_i = \sum_{i=1}^{\infty} V_i \mathbb{1}\{\zeta \geq i\} \Rightarrow E[I] = \sum_{i=1}^{\infty} E[V_i] \prod_{k=1}^{i-1} L_k(\lambda),$$

as the vacation period  $V_i$  does not depend on the event of no arrivals during  $\hat{V}_{i-1}$ . Define  $I_a := \sum_{i=1}^{\infty} V_i^2 \mathbb{1}\{\zeta \geq i\}$ , then

$$E[I_a] = \sum_{i=1}^{\infty} E[V_i^2] \prod_{k=1}^{i-1} L_k(\lambda).$$

The number of requests waiting in the queue at the beginning of a busy period is  $Z = Z_I + Z_w$ .  $Z_w$ , the number of arrivals during a warm-up period  $T_w$ , is a Poisson variable with parameter  $\lambda T_w$ . We then have  $E[Z_w] = \lambda T_w$ , and  $E[Z_w^2] = \lambda^2 T_w^2 + \lambda T_w$ . It can be shown that  $E[Z_I] = \lambda E[I]$  and  $E[Z_I^2] = \lambda^2 E[I_a] + \lambda E[I]$ ; refer to [1] for details.  $Z_I$  and  $Z_w$  are independent, and so

$$\begin{aligned} E[Z] &= \lambda(E[I] + T_w) \\ E[Z^2] &= \lambda^2(E[I_a] + 2E[I]T_w + T_w^2) + \lambda(E[I] + T_w). \end{aligned} \quad (1)$$

From Fig. 1 and using (1) we can write (note that  $B_1 = B_{M/G/1}$ )

$$E[B] = E[B_1]E[Z] = \frac{\rho}{1-\rho}(E[I] + T_w).$$

**The Queue Size and Sojourn Time.** Let

$$A := \int_{\hat{V}_{\zeta-1}}^{\hat{V}_{\zeta}} X(t)dt, A_w := \int_{\hat{V}_{\zeta}}^{T_Z} X(t)dt, Q_Z := \int_{T_Z}^{T_0} X(t)dt,$$

be defined as the total area under the curve  $X(t)$  for the idle, warm-up and busy periods respectively. Then

$$E[X] = \frac{E[A] + E[A_w] + E[Q_Z]}{E[I] + T_w + E[B]}. \quad T = \frac{E[X]}{\lambda}. \quad (2)$$

Let  $N(t), t \geq 0$  be a Poisson process with rate  $\lambda$ . Let  $\tau_i$  be the  $i$ th arrival epoch. Define  $A(t) := E[\alpha(t)|N(t) \geq 1]$  with

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

We thus need to compute  $E[A] = E[A(V_{\zeta})]$ . We have  $E[\alpha(t)] = \lambda t^2/2$ . Therefore,

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

$$E[A] = \frac{\lambda}{2} \sum_{i=1}^{\infty} \left( \prod_{k=1}^{i-1} L_k(\lambda) \right) (1 - L_i(\lambda)) \sum_{k=0}^{\infty} \frac{d^2 L_i(s)}{ds^2} \Big|_{s=k\lambda}.$$

From Fig. 1 we can write

$$E[A_w] = E[Z_I]T_w + \int_0^{T_w} \lambda t dt = \lambda T_w (E[I] + T_w/2).$$

The following recursive equation holds

$$Q_Z = \left( (Z-1)B_Z + Q_1 \right) + Q_{Z-1}.$$

Hence  $E[Q_Z] = E[Z]E[Q_1] + (E[Z^2] - E[Z])E[B_1]/2$  where

$$E[Q_1] = E[Q_{M/G/1}] = \frac{E[\sigma]}{1-\rho} + \frac{\lambda E[\sigma^2]}{2(1-\rho)^2}$$

which completes the computation of the response time  $T$ ; see (2).

### 3 Application to Power Saving

Let  $S_i$  be a generic random variable that gives the actual time a node is sleeping during the  $i$ th vacation period. We then have  $V_1 = S_1$  and  $V_i = T_l + S_i$  for  $i = 2, \dots, \zeta$ . The listen window  $T_l$  accounts for the time needed by a node to check for messages at the base station. We assume  $T_l = T_w$  and consider either

(i)  $S_i = a^{\min\{i-1, l\}} T_{\min}$ ,  $i = 1, 2, \dots$ , (deterministic case)

(ii)  $E[S_i] = a^{\min\{i-1, l\}} T_{\min}$ ,  $i = 1, 2, \dots$ , (exponential  $S_i$ ).

Define the economy of energy  $G := (E_{\text{no sleep}} - E_{\text{sleep}})/E_{\text{no sleep}}$ . The following nonlinear program  $\mathcal{P}$ :

$$\text{maximize } G \text{ subject to } T \leq T_{\text{QoS}}$$

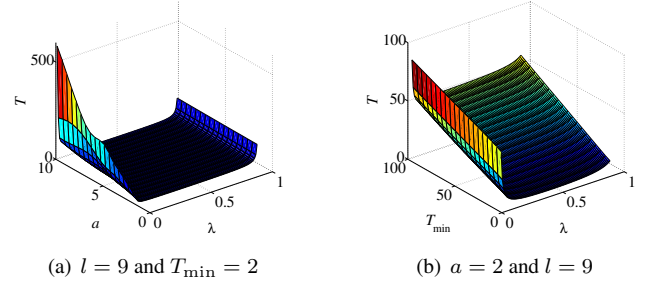


Figure 2: The expected system response time  $T$  when sleep windows are deterministic.

Table 1: Optimal protocol parameters obtained from  $\mathcal{P}$

$\lambda$	Deterministic case				Exponential case			
	$l$	$a$	$T_{\min}$	$G$	$l$	$a$	$T_{\min}$	$G$
0.02	1	1.5	72	89.7%	2	2.5	27	89.0%
0.03	5	1.5	82	85.7%	2	3.0	22	85.2%
0.05	3	2.0	92	78.3%	3	1.5	32	78.0%
0.10	1	5.0	92	63.6%	1	1.5	42	63.5%
0.20	1	5.0	92	44.0%	9	1.5	47	44.0%

is solved for the three protocol parameters: the exponent  $l$ , the multiplicative factor  $a$  and the initial expected sleep window  $T_{\min}$ .

The system parameters have been selected as follows:  $E[\sigma] = 1$ ,  $E[\sigma^2] = 2$ ,  $T_l = T_w = 1$ ,  $T_{\text{QoS}} = 50, 100$ . The ratio between the energy consumptions of a node in idle and active states is set to 0.2.

Numerical results for the sojourn time  $T$  in the deterministic case have been derived and are displayed graphically in Fig. 2. The sojourn time  $T$  appears to be fairly insensitive to parameters  $l$  and  $a$  except for very small values of  $\lambda$ . Observe that the load  $\rho = \lambda E[\sigma]$  gives the probability of finding the queue busy. Hence, with probability  $1 - \lambda E[\sigma]$ , an arriving request finds the server on vacation. This explains the high response time for low input rates. For very large input rates, the system response time is also high, but then it is mainly due to queueing delays. We have solved the nonlinear program  $\mathcal{P}$  for  $(l, a, T_{\min})^*$  with  $T_{\text{QoS}} = 50$  for deterministic sleep windows and  $T_{\text{QoS}} = 100$  for exponential sleep windows. The values of the optimal protocol parameters  $(l, a, T_{\min})^*$  are in Table 1.

### References

- [1] S. Alouf, E. Altman and A. P. Azad. Analysis of an  $M/G/1$  queue with repeated inhomogeneous vacations – Application to IEEE 802.16e power saving. INRIA Research Report, March 2008, available at <http://hal.inria.fr/inria-00266552/>.