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
salouf@sophia.inria.fr, altman@sophia.inria.fr, aazad@sophia.inria.fr

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Modeling techniques; G.3 [**Probability and Statistics**]: Queueing theory, Renewal theory, Stochastic processes; G.1.6 [**Numerical Analysis**]: Optimization—*Constrained optimization, Nonlinear programming*; C.2.5 [**Computer-Communication Networks**]: Local and Wide-Area Networks—*Access schemes*

General Terms

Performance, Theory

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 λ. Let σ 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):

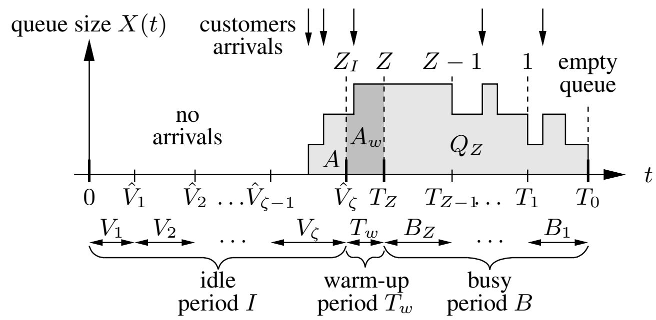


Figure 1: Sample trajectory of the queue size during a regeneration cycle.

(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 ζ vacation periods denoted $V_1, \ldots, V_\zeta$;

(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 ith vacation period, $i = 1, \ldots, \zeta$; 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 $i = Z-1, \ldots, 0$; the cycle ends at T_0.

The times $\{T_i\}_{i=Z,Z-1,\ldots,0}$ 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 $X(I) = Z_I$.

2. ANALYSIS

Number of Vacations. Let ζ 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 $\{V_k\}_{k=1,\ldots,i-1}$. Denoting by $L_k(s) = \mathrm{E}[\exp(-sV_k)]$ the Laplace Stieltjes transform of V_k, we have

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

SIGMETRICS'08, June 2–6, 2008, Annapolis, Maryland, USA.
ACM 978-1-60558-005-0/08/06.

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

The Idle and Busy Period. We can write

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

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

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

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 λT_w. We then have $\mathrm{E}[Z_w] = \lambda T_w$, and $\mathrm{E}\left[Z_w^2\right] = \lambda^2 T_w^2 + \lambda T_w$. It can be shown that $\mathrm{E}[Z_I] = \lambda \mathrm{E}[I]$ and $\mathrm{E}[Z_I^2] = \lambda^2 \mathrm{E}[I_a] + \lambda \mathrm{E}[I]$; refer to [1] for details. Z_I and Z_w are independent, and so

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

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

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

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

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

Let $N(t), t \geq 0$ be a Poisson process with rate λ. Let τ_i be the ith arrival epoch. Define $A(t) := \mathrm{E}[\alpha(t) | N(t) \geq 1]$ with

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

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

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

From Fig. 1 we can write

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

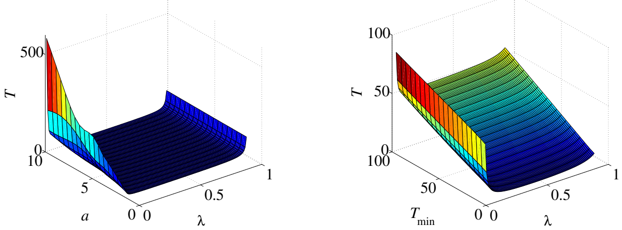
The following recursive equation holds

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

Hence $\mathrm{E}[Q_z] = \mathrm{E}[Z]\mathrm{E}[Q_1] + (\mathrm{E}[Z^2] - \mathrm{E}[Z])\mathrm{E}[B_1]/2$ where

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

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



(a) $l = 9$ and $T_{\min} = 2$ (b) $a = 2$ and $l = 9$

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

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

	Deterministic case				Exponential case			
λ	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%

3. APPLICATION TO POWER SAVING

Let S_i be a generic random variable that gives the actual time a node is sleeping during the ith vacation period. We then have $V_1 = S_1$ and $V_i = T_l + S_i$ for $i = 2, \ldots, \zeta$. 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) $S_i = a^{\min\{i-1,l\}} T_{\min}, \; i = 1, 2, \ldots,$ (deterministic case)
(ii) $\mathrm{E}[S_i] = a^{\min\{i-1,l\}} T_{\min}, \; i = 1, 2, \ldots,$ (exponential S_i).
Define the economy of energy $G := (E_{\text{no sleep}} - E_{\text{sleep}})/E_{\text{no sleep}}$. The following nonlinear program $\mathcal{P}$:

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

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: $\mathrm{E}[\sigma] = 1$, $\mathrm{E}[\sigma^2] = 2$, $T_l = T_w = 1$, $T_{\text{QoS}} = 50, 100$. 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 λ. Observe that the load $\rho = \lambda \mathrm{E}[\sigma]$ gives the probability of finding the queue busy. Hence, with probability $1 - \lambda \mathrm{E}[\sigma]$, 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.

4. 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/`.