Optimal sleep-state control of energy-aware M/G/1 queues

Misikir Eyob Gebrehiwot
Aalto University, Finland
misikir.gebrehiwot@aalto.fi

Samuli Aalto
Aalto University, Finland
samuli.aalto@aalto.fi

Pasi Lassila
Aalto University, Finland
pasi.lassila@aalto.fi

ABSTRACT

We study the problem of optimally controlling the use of sleep states in an energy-aware M/G/1 queue. In our model, we consider a family of policies where the server upon becoming idle can wait for a random period before entering, potentially randomly, any of a finite number of possible sleep states to save energy. The server becomes busy again after a possibly random number of jobs have arrived. However, jobs are served only after a random setup time. This kind of an energy-aware queuing system has been analyzed in recent papers under specific assumptions regarding the cost metrics and the distributions of the random variables. In this paper, we consider an essentially more general model. Notably we show that the optimal control of the idle time and sleep states is deterministic and does not benefit from randomization: either the system only uses the idle state and no sleep states, or the idle state is not used at all and the server immediately goes to some fixed sleep state and waits until a fixed number of jobs have arrived before starting the setup. We prove this result for two popular cost metrics, namely weighted sum of energy and response time (ERWS) and their product ERP.

Categories and Subject Descriptors

D.2.8 [**Metrics**]: Performance; D.4.8 [**Performance**]: Queueing theory; B.1.2 [**Control Structure Performance Analysis and Design Aids**]:

General Terms

Performance, Theory

Keywords

Performance-energy trade-off, $M/G/1$, Setup delay

1. INTRODUCTION

An increasing demand for green ICT has inspired the queueing community to consider energy-aware queueing systems. In many cases, it is no longer enough to optimize just the performance costs but one should also take into account the energy costs. An idle server (waiting for an arriving job to be processed) in the server farm of a typical data center may consume as much as 60% of the peak power. From the energy point of view, such an idle server should be switched off until a new job arrives. However, from the performance point of view, this is suboptimal since it typically takes a rather long time to wake the server up. Thus, there is a clear trade-off between the performance and energy aspects. The performance aspects of similar vacation models have been analyzed in, e.g., [18, 5, 15].

The two main metrics used in the literature to analyze the performance-energy trade-off in energy-aware queueing systems are ERWS [1, 4, 3, 19, 2, 16, 9] and ERP [8, 17, 12, 11, 13, 6]. Both of them are based on the expected response time, $\mathrm{E}[T]$, and the expected power consumption per time unit, $\mathrm{E}[P]$. The former one, ERWS, is defined as their weighted sum,

$$w_1\mathrm{E}[T] + w_2\mathrm{E}[P], \quad w_1, w_2 \geq 0, \tag{1}$$

and the latter one, ERP, as their product,

$$\mathrm{E}[T]\mathrm{E}[P]. \tag{2}$$

In this paper, we focus on the optimal control of a single energy-aware server. We assume that jobs arrive according to a Poisson process and service times are IID with a general distribution. In addition to the *off* state, we assume that the server has multiple intermediate *sleep* states, for which the deeper the sleep state, the smaller the power consumption but also the longer the setup delay. The possible control actions are as follows:

(i) When the server becomes idle, it will start a timer denoted by I, which is an IID random variable with a general distribution (including even the deterministic special cases $I = 0$ and $I = \infty$), and wait until either a new job arrives or the timer expires. If there is an arrival before the timer expiration, the server will immediately start a new busy period and no further control actions are possible until the server next becomes idle again.[1]

(ii) However, if the timer expires before a new arrival, the server will independently and randomly choose one of

[1]Note that if $I = \infty$, then the system is an ordinary M/G/1 queue.

VALUETOOLS 2014, December 09-11, Bratislava, Slovakia

DOI 10.4108/icst.valuetools.2014.258149

the possible sleeping states, say i, and a switch-on threshold, say k, from a given distribution $\mathbf{p} = (p_{ik})$. The server will stay in the chosen sleep state i until the queue length reaches the chosen threshold k. At that time, the server is switched on so that after a setup delay denoted by D_i, which is an independent random variable with a general distribution depending on i, the system will start a new busy period and no further control actions are possible until the server next becomes idle again.

A control policy in this system is defined by giving the distribution of the idling time I and the sleep-state/threshold distribution $\mathbf{p}$. The target is to find the optimal distributions that minimize the chosen cost metric.

Our purpose is to unify and generalize some results presented in two recent papers, [6] and [14]. Gandhi et al. [6] considered this model, however, without an idling option (so that only $I = 0$ and $I = \infty$ are possible alternatives) and assuming exponential service times and deterministic setup delays. They showed that the optimal policy for the ERP metric is deterministic with threshold $k^* = 1$ (i.e., there is i^* such that $p_{i^*,1} = 1$). In this paper, we prove that the optimal policy still remains deterministic even under the idling option and for any distributions for service times and setup delays. We also show that the optimal threshold may not necessarily be $k = 1$ for the ERP metric when general service and setup times are considered. In addition, we prove that the optimal policy is deterministic also for the ERWS metric with the optimal threshold being not necessarily $k = 1$.

Maccio and Down [14] also considered this model but they restricted themselves to a single sleep state and deterministic rules to choose the corresponding threshold k. They showed, however, explicitly only for the ERWS metric and assuming exponential service times and setup delays, that the optimal idling time is either $I = 0$ or $I = \infty$. In addition, they gave a heuristic argument for the claim that the result would be valid for a more general cost metric, namely

$$\sum_{j=1}^{M} w_j \mathrm{E}[T]^{a_j} \mathrm{E}[P]^{b_j}, \quad w_j, a_j, b_j \geq 0 \; \forall j, \tag{3}$$

which covers both common metrics, ERWS and ERP. In this paper, we prove that the optimal idling time is, indeed, either $I = 0$ or $I = \infty$ for the ERWS and ERP cost metrics even if we allow general distributions for service times and setup delays, multiple sleep states, and randomized rules to choose the sleep state i and the corresponding threshold k. In addition, we prove that the result is valid for the general cost metric (3) if there is only a single term in the sum, i.e., for the cost metric

$$w\mathrm{E}[T]^a \mathrm{E}[P]^b, \quad w, a, b \geq 0, \tag{4}$$

which is a slightly generalized version of the ERP metric. However, if there are multiple terms in (3), we demonstrate, by constructing a counter-example, that the optimal idling time may be different from $I = 0$ and $I = \infty$.

The rest of the paper is organized as follows. In Section 2, we introduce the system model and its analysis is in Section 3. Cost metric optimization is carried out in Section 4, and an extension of idling time considerations is given in Section 5. We give numerical examples in Section 6 and conclude the paper in Section 7.

2. MODEL

We consider an energy-aware M/G/1-FIFO queue, where service requests (also referred to as jobs) arrive according to a Poisson process with rate λ and service times S may have any distribution with mean $\mathrm{E}[S]$. Let $\rho = \lambda \mathrm{E}[S]$ denote the system load. In addition to the ordinary *busy* and *idle* states, the server has an *off* state and $n-1$ different intermediate *sleep* states ($sleep_1, \ldots, sleep_{n-1}$). For the sake of notational convenience, the states *off* and *idle* are also referred to as $sleep_0$ and $sleep_n$, respectively. With this notation, let P_i [P_{busy}] denote the (deterministic) power consumption of the server when in state $sleep_i$ [*busy*]. We assume that

$$0 = P_{\mathrm{off}} = P_0 < P_1 < \ldots < P_n = P_{\mathrm{idle}} < P_{\mathrm{busy}}.$$

While transition from state *idle* to *busy* is immediate, that from any *sleep* state induces a random setup delay. Let D_i denote the setup delay from state $sleep_i$ to *busy*. Setup delays D_i may have any distribution but we make a natural assumption that the mean values satisfy

$$\mathrm{E}[D_{\mathrm{off}}] = \mathrm{E}[D_0] > \mathrm{E}[D_1] > \ldots > \mathrm{E}[D_n] = \mathrm{E}[D_{\mathrm{idle}}] = 0.$$

All setup delays are assumed to be independent of each other. The (deterministic) power consumption during any *setup* delay is denoted by P_{setup}.

As already described in Section 1, the power consumption is controlled by a policy, which is defined by specifying the following two distributions: (i) the distribution for the idling time I (with the special cases $I = 0$ and $I = \infty$ included) and (ii) the distribution $\mathbf{p} = (p_{ik}; i \in \{0, \ldots, n-1\}, k \in \{1, 2, \ldots\})$. Let Π_{mixed} denote the family of all such policies.[2]

In the sequel, we will also use the following short-hand notations:

$$p_i = \sum_{k=1}^{\infty} p_{ik}, \quad q_k = \sum_{i=0}^{n-1} p_{ik}.$$

Recall that when the idling time timer expires in state *idle*, the following *sleep* state will be i and the corresponding switch-on threshold k with probability p_{ik}. Initially we assume, as in [14], that the idling time timer is reset (to 0) only when it expires. Thus, if the server is *idle* and a new job arrives before the timer expires, the server moves from state *idle* to *busy* and the idling time timer is not reset but the accumulated idling time is *remembered* for the next time the server becomes *idle* again. In Section 5, we argue that the optimality results remain the same even if the idling time timer is reset after *any* idle period.

3. ANALYSIS

In this section, we derive the mean response time $\mathrm{E}[T]$ and the mean power consumption $\mathrm{E}[P]$ for the policies belonging to Π_{mixed}. The results are given below in Theorems 1 and

[2]This is an extension of the family Π_{mixed} defined in [6] allowing the idling option before the transit to some *sleep* state.

2, which generalize the corresponding earlier results given in [6] and [14].

The special case $I = \infty$ corresponds to an ordinary M/G/1 queue for which

$$\mathrm{E}[T_{\mathrm{M/G/1}}] = \mathrm{E}[S] + \frac{\lambda \mathrm{E}[S^2]}{2(1-\rho)},$$
$$\mathrm{E}[P_{\mathrm{M/G/1}}] = \rho P_{\mathrm{busy}} + (1-\rho) P_{\mathrm{idle}}.$$

Consider now any policy for which $\mathrm{E}[I] < \infty$. In this case the idling time timer expires sooner or later, and every time it expires, the system regenerates itself.

Let C denote a regeneration cycle between two consecutive timer expirations. Let T_{busy}, T_{idle}, T_{sleep_i}, and T_{setup_i} denote the *aggregate* time during one cycle that the system is in the *busy* state, in the *idle* state, in *sleep*$_i$ state ($i \in \{0, \ldots, n-1\}$), and in *setup*$_i$ state ($i \in \{0, \ldots, n-1\}$), respectively. Thus,

$$C = T_{\mathrm{busy}} + T_{\mathrm{idle}} + \sum_{i=0}^{n-1} (T_{\mathrm{sleep}_i} + T_{\mathrm{setup}_i}).$$

Since FIFO is a work-conserving service discipline, we know that

$$\mathrm{E}[T_{\mathrm{busy}}] = \rho \mathrm{E}[C]. \tag{5}$$

In addition, since the idling time timer is reset only in the beginning of a new cycle, we have $T_{\mathrm{idle}} = I$, and consequently

$$\mathrm{E}[T_{\mathrm{idle}}] = \mathrm{E}[I]. \tag{6}$$

It is also easy to see that

$$\mathrm{E}[T_{\mathrm{sleep}_i}] = \sum_{k=1}^{\infty} p_{ik} \frac{k}{\lambda}, \tag{7}$$

$$\mathrm{E}[T_{\mathrm{setup}_i}] = \sum_{k=1}^{\infty} p_{ik} \mathrm{E}[D_i]. \tag{8}$$

Combining these together, we get

$$\mathrm{E}[C] = \frac{1}{1-\rho} \left(\mathrm{E}[I] + \sum_{k=1}^{\infty} q_k \frac{k}{\lambda} + \sum_{i=0}^{n-1} p_i \mathrm{E}[D_i] \right), \tag{9}$$

where, as mentioned earlier,

$$p_i = \sum_{k=1}^{\infty} p_{ik}, \quad q_k = \sum_{i=0}^{n-1} p_{ik}.$$

Note that the mean cycle length $\mathrm{E}[C]$ is insensitive to the shape of the idling time distribution depending just on its mean value $\mathrm{E}[I]$. In addition, for the special case where $p_{ik} = 1$ for some i and k, we clearly have

$$\mathrm{E}[C] = \frac{1}{1-\rho} \left(\mathrm{E}[I] + \frac{k}{\lambda} + \mathrm{E}[D_i] \right). \tag{10}$$

THEOREM 1. *For any policy belonging to* Π_{mixed}*, the mean power consumption is given by*

$$\mathrm{E}[P] = \mathrm{E}[P_{\mathrm{M/G/1}}] + \frac{1}{\mathrm{E}[C]} \left(\sum_{i=0}^{n-1} \sum_{k=1}^{\infty} p_{ik} \frac{k}{\lambda} (P_i - P_{\mathrm{idle}}) + \sum_{i=0}^{n-1} p_i \mathrm{E}[D_i] (P_{\mathrm{setup}} - P_{\mathrm{idle}}) \right), \tag{11}$$

where $\mathrm{E}[C]$ *is given in (9).*

Proof: By utilizing the theory of regenerative processes, we know that

$$\mathrm{E}[P] = \frac{\mathrm{E}[T_{\mathrm{busy}}]}{\mathrm{E}[C]} P_{\mathrm{busy}} + \frac{\mathrm{E}[T_{\mathrm{idle}}]}{\mathrm{E}[C]} P_{\mathrm{idle}} + \sum_{i=0}^{n-1} \left(\frac{\mathrm{E}[T_{\mathrm{sleep}_i}]}{\mathrm{E}[C]} P_i + \frac{\mathrm{E}[T_{\mathrm{setup}_i}]}{\mathrm{E}[C]} P_{\mathrm{setup}} \right),$$

from which (11) easily follows by (5)–(8). □

Remark: Equation (11) applied to the case where $I = 0$ and setup delays D_i are deterministic gives the same result as given in [6, Equation (3)]. In addition, for the special case where $p_{ik} = 1$ for some i and k, we clearly have, by (11),

$$\mathrm{E}[P] = \mathrm{E}[P_{\mathrm{M/G/1}}] + (1-\rho) \frac{\frac{k}{\lambda}(P_i - P_{\mathrm{idle}}) + \mathrm{E}[D_i](P_{\mathrm{setup}} - P_{\mathrm{idle}})}{\mathrm{E}[I] + \frac{k}{\lambda} + \mathrm{E}[D_i]}, \tag{12}$$

which corresponds to the equation of the mean energy consumption given in [14, Theorem 2].

THEOREM 2. *For any policy belonging to* Π_{mixed}*, the mean response time is given by*

$$\mathrm{E}[T] = \mathrm{E}[T_{\mathrm{M/G/1}}] + \frac{1}{(1-\rho)\mathrm{E}[C]} \left(\sum_{k=1}^{\infty} q_k \frac{k(k-1)}{2\lambda^2} + \sum_{i=0}^{n-1} \sum_{k=1}^{\infty} p_{ik} \frac{k}{\lambda} \mathrm{E}[D_i] + \sum_{i=0}^{n-1} p_i \frac{1}{2} \mathrm{E}[D_i^2] \right). \tag{13}$$

Proof: To determine the mean response time, we continue to utilize the regenerative cycle analysis introduced above. More precisely said, we derive the mean response time $\mathrm{E}[T]$ using a similar approach as Medhi for M/G/1-FIFO vacation models in [15].

Let us start with the mean waiting time $\mathrm{E}[W]$. We will derive it in three parts:

$$\mathrm{E}[W] = \mathrm{E}[W_1] + \mathrm{E}[W_2] + \mathrm{E}[W_3].$$

1° First, due to the FIFO service discipline, the arriving job has to wait until the end of the current service, if the system is busy upon the arrival, and until the end of the service of all the jobs already waiting upon the arrival. Thus,

$$\mathrm{E}[W_1] = \pi_{\mathrm{busy}} \mathrm{E}[S^R] + \mathrm{E}[\sum_{j=1}^{N_W} S_j],$$

where π_{busy} denotes the probability that the system is busy upon the arrival, S^R is the remaining service time of the job in service, N_W is the number of waiting jobs upon the arrival, and the S_j refer to their service times. The standard M/G/1 analysis (applying PASTA and Little's result $\mathrm{E}[N_W] = \lambda \mathrm{E}[W]$) gives immediately

$$\mathrm{E}[W_1] = \rho \frac{\mathrm{E}[S^2]}{2\mathrm{E}[S]} + \lambda \mathrm{E}[W]\mathrm{E}[S] = \frac{\lambda}{2}\mathrm{E}[S^2] + \rho \mathrm{E}[W].$$

2° In addition, a job that arrives when the server is in one of the states $sleep_i$ $(i = 0, \ldots, n-1)$ has to wait until the end of the current sleep state, which is controlled by parameter k, as well as the following setup delay. Thus,

$$\mathrm{E}[W_2] = \sum_{i=0}^{n-1} \sum_{k=1}^{\infty} \pi_{ik} \left(\frac{k-1}{2\lambda} + \mathrm{E}[D_i] \right),$$

where π_{ik} denotes the probability that the current cycle is related to state $sleep_i$ and switch-on threshold k and that the arriving job is one of these k jobs. Clearly, we have

$$\pi_{ik} = p_{ik} \frac{k}{\lambda \mathrm{E}[C]},$$

implying that

$$\mathrm{E}[W_2] = \frac{1}{\mathrm{E}[C]} \left(\sum_{k=1}^{\infty} q_k \frac{k(k-1)}{2\lambda^2} + \sum_{i=0}^{n-1} \sum_{k=1}^{\infty} p_{ik} \frac{k}{\lambda} \mathrm{E}[D_i] \right),$$

3° Finally, a job that arrives during one of the setup delays D_i $(i = 0, \ldots, n-1)$ has to wait until the end of the current setup delay. Thus,

$$\mathrm{E}[W_3] = \sum_{i=0}^{n-1} \sum_{k=1}^{\infty} \pi_i \mathrm{E}[D_i^R],$$

where π_i denotes the probability that the current cycle is related to state $sleep_i$ and that the arriving job is one of the jobs arriving during the corresponding setup delay D_i, and D_i^R refers to the remaining part of the setup delay upon the arrival. Clearly, we have

$$\pi_i = \sum_{k=1}^{\infty} p_{ik} \frac{\lambda \mathrm{E}[D_i]}{\lambda \mathrm{E}[C]} = p_i \frac{\mathrm{E}[D_i]}{\mathrm{E}[C]}.$$

In addition, the standard renewal theory says that

$$\mathrm{E}[D_i^R] = \frac{\mathrm{E}[D_i^2]}{2\mathrm{E}[D_i]}.$$

Thus,

$$\mathrm{E}[W_3] = \frac{1}{\mathrm{E}[C]} \left(\sum_{i=0}^{n-1} p_i \frac{1}{2} \mathrm{E}[D_i^2] \right).$$

Formula (13) follows now straightforwardly from 1°–3°, since $\mathrm{E}[T] = \mathrm{E}[S] + \mathrm{E}[W]$. □

Remark: Equation (13) applied to the case where $I = 0$ and setup delays D_i are deterministic gives the same result as given in [6, Equation (1)]. In addition, for the special case where $p_{ik} = 1$ for some i and k, we clearly have, by (13),

$$\mathrm{E}[T] = \mathrm{E}[T_{\mathrm{M/G/1}}] + \frac{\frac{k(k-1)}{2\lambda^2} + \frac{k}{\lambda}\mathrm{E}[D_i] + \frac{1}{2}\mathrm{E}[D_i^2]}{\mathrm{E}[I] + \frac{k}{\lambda} + \mathrm{E}[D_i]}, \qquad (14)$$

which corresponds to the equation of the mean response time given in [14, Theorem 3].[3]

4. OPTIMIZATION

In this section, we prove that, for the cost metrics ERWS (1) and generalized ERP (4), the optimal policy is deterministic (choosing exactly one of states $sleep_i$ and thresholds k) and the optimal idling time is either $I = 0$ and $I = \infty$, which generalizes the earlier results given in [6] and [14] as discussed in Section 1. The result is first split in parts and proved in Propositions 1–4, and finally summarized in Theorem 3.

4.1 Optimal idling time distribution

Let us first consider the optimization of the idling time distribution. It follows from (9), (11), and (13) that both the mean response time $\mathrm{E}[T]$ and the mean power consumption $\mathrm{E}[P]$ are, in fact, insensitive to the shape of the idling time distribution depending just on its mean value $\mathrm{E}[I]$ as follows:

$$\mathrm{E}[T] = A_1 + \frac{B_1}{C + \mathrm{E}[I]}, \quad \mathrm{E}[P] = A_2 + \frac{B_2}{C + \mathrm{E}[I]}, \qquad (15)$$

where constants $A_1, A_2, B_1, C > 0$ but B_2 may be negative.

PROPOSITION 1. *For the ERWS cost metric (1), the optimal policy in* Π_{mixed} *has either* $I = 0$ *or* $I = \infty$.

Proof: By (15), the objective function (1) can be written in the following form:

$$w_1 \mathrm{E}[T] + w_2 \mathrm{E}[P] = w_1 A_1 + w_2 A_2 + \frac{w_1 B_1 + w_2 B_2}{C + \mathrm{E}[I]},$$

which is clearly a monotonic function of $\mathrm{E}[I]$ in the whole interval $[0, \infty)$. □

PROPOSITION 2. *For the generalized ERP cost metric (4), the optimal policy in* Π_{mixed} *has either* $I = 0$ *or* $I = \infty$.

Proof: By (15), the objective function (4) can be written in the following form:

$$w \mathrm{E}[T]^a \mathrm{E}[P]^b = w \left(A_1 + \frac{B_1}{C + \mathrm{E}[I]} \right)^a \left(A_2 + \frac{B_2}{C + \mathrm{E}[I]} \right)^b.$$

If $B_2 \geq 0$, then the objective function is clearly strictly decreasing so that the optimal idling time is $I = \infty$. Thus, from this on, we assume that $B_2 < 0$. Let us now consider the function $f(x)$ defined for all real values of x as follows:

$$f(x) = \left(A_1 + \frac{B_1}{C + x} \right)^a \left(A_2 + \frac{B_2}{C + x} \right)^b.$$

1° First we show that the first derivative $f'(x)$ has at most one root. By taking the first derivative of $f(x)$ and rearranging the terms, we have

$$f'(x) = \left(A_1 + \frac{B_1}{C + x} \right)^{a-1} \left(A_2 + \frac{B_2}{C + x} \right)^{b-1} \frac{1}{(C+x)^3} \times \left(-aB_1(A_2(C + x) + B_2) - bB_2(A_1(C + x) + B_1) \right).$$

[3] The formula for the mean response time given in [14, Theorem 3] is more complicated but gives (14) after some manipulations when interpreting the symbol σ^2_{setup} in [14] as the second moment $\mathrm{E}[D_i^2]$ (instead of the variance $\mathrm{V}[D_i]$) of the setup delay.

Clearly, the only possible root comes from the last part of this equation, which is a linear function of x. If such a root x_0 exists, it satisfies

$$x_0 = -\frac{B_1 B_2 (a+b)}{a A_2 B_1 + b A_1 B_2} - C.$$

2° Now we show that any root of $f'(x)$ is a local maximum point. Let us assume that such a root, x_0, exists. The second derivative test can be applied to prove this claim since $f(x)$ is twice differentiable at x_0. By taking the second derivative of $f(x)$ and applying it at x_0, we get

$$f''(x_0) = \left(A_1 + \tfrac{B_1}{C+x_0}\right)^{a-1} \left(A_2 + \tfrac{B_2}{C+x_0}\right)^{b-1} \tfrac{(a+b)B_1 B_2}{(C+x_0)^4}.$$

Since $B_2 < 0$, we have $f''(x_0) < 0$, which justifies the claim.

From 1° and 2°, we deduce that $f(x)$, when restricted to the interval $x \in [0, \infty)$, has its minimum value at $x = 0$ or when $x \to \infty$, which completes the proof. □

4.2 Optimal choice of sleep state and switch-on threshold

Now we consider, for the same cost metrics as above, optimization of the distribution $\mathbf{p} = (p_{ik}; i \in \{0, \ldots, n-1\}, k \in \{1, 2, \ldots\})$, which is used to select the sleep state i and the switch-on threshold k every time when the idling time timer expires. From Propositions 1 and 2, we know that $I = 0$ or $I = \infty$ for the optimal policy. On the other hand, if the optimal idling time is $I = \infty$, the distribution $\mathbf{p}$ does not have any role. Thus, we may assume below that $I = 0$.

Consider first any distribution $\mathbf{p} = (p_{ik}; i \in \{0, \ldots, n-1\}, k \in \{1, 2, \ldots\})$. For a while, fix i and k, and define the following conditional probabilities:

$$q_{j\ell} = \frac{p_{j\ell}}{1 - p_{ik}} \quad \text{for any } (j, \ell) \neq (i, k).$$

Utilizing these conditional probabilities, we define a related distribution $\mathbf{p}^0 = (p^0_{j\ell}; j \in \{0, \ldots, n-1\}, \ell \in \{1, 2, \ldots\})$ by

$$p^0_{ik} = 0, \quad p^0_{j\ell} = q_{j\ell} \quad \text{for any } (j, \ell) \neq (i, k),$$

and still another related distribution $\mathbf{p}^1 = (p^1_{j\ell}; j \in \{0, \ldots, n-1\}, \ell \in \{1, 2, \ldots\})$ by

$$p^1_{ik} = 1, \quad p^1_{j\ell} = 0 \quad \text{for any } (j, \ell) \neq (i, k).$$

We note that, with $I = 0$, these three distributions $(\mathbf{p}, \mathbf{p}^0, \mathbf{p}^1)$ define three different control policies that all belong to Π_{mixed}.

It follows from (9), (11), and (13) that the mean response time $\mathrm{E}[T]$ and the mean power consumption $\mathrm{E}[P]$ for these three policies can be written as follows:

$$\begin{aligned}
&\mathrm{E}[T|\mathbf{p}] = A_1 + \tfrac{B_1 p_{ik} + C_1}{D p_{ik} + E}, && \mathrm{E}[P|\mathbf{p}] = A_2 + \tfrac{B_2 p_{ik} + C_2}{D p_{ik} + E}, \\
&\mathrm{E}[T|\mathbf{p}^0] = A_1 + \tfrac{C_1}{E}, && \mathrm{E}[P|\mathbf{p}^0] = A_2 + \tfrac{C_2}{E}, \\
&\mathrm{E}[T|\mathbf{p}^1] = A_1 + \tfrac{B_1 + C_1}{D + E}, && \mathrm{E}[P|\mathbf{p}^1] = A_2 + \tfrac{B_2 + C_2}{D + E},
\end{aligned} \tag{16}$$

where $A_1, A_2, C_1, C_2, E > 0$ and $B_1 + C_1, B_2 + C_2, D + E > 0$ but B_1, B_2, and D may be negative.[4]

[4] While we use partly the same symbols, do not confuse these constants with those given in (15).

If $D \neq 0$, then an elementary polynomial division results in the following formulas:

$$\begin{aligned}
&\mathrm{E}[T|\mathbf{p}] = A'_1 + \tfrac{C'_1}{D p_{ik} + E}, && \mathrm{E}[P|\mathbf{p}] = A'_2 + \tfrac{C'_2}{D p_{ik} + E} \\
&\mathrm{E}[T|\mathbf{p}^0] = A'_1 + \tfrac{C'_1}{E}, && \mathrm{E}[P|\mathbf{p}^0] = A'_2 + \tfrac{C'_2}{E}, \\
&\mathrm{E}[T|\mathbf{p}^1] = A'_1 + \tfrac{C'_1}{D + E}, && \mathrm{E}[P|\mathbf{p}^1] = A'_2 + \tfrac{C'_2}{D + E},
\end{aligned} \tag{17}$$

where $A'_j = A_j + \frac{B_j}{D}$ and $C'_j = C_j - \frac{B_j E}{D}$ for $j = 1, 2$.

PROPOSITION 3. *For the ERWS cost metric (1), the optimal policy in Π_{mixed} is deterministic, i.e., there are i^* and k^* such that $p_{i^*,k^*} = 1$.*

Proof: 1° Assume first that $D = 0$. By (16), the objective function (1) can be written in the following form:

$$\begin{aligned}
&w_1 \mathrm{E}[T|\mathbf{p}] + w_2 \mathrm{E}[P|\mathbf{p}] = \\
&\quad w_1 (A_1 + \tfrac{C_1}{E}) + w_2 (A_2 + \tfrac{C_2}{E}) + \tfrac{w_1 B_1 + w_2 B_2}{E} p_{ik},
\end{aligned}$$

which is a linear and, thus, monotonic function of p_{ik} in the interval $[0, 1]$. Thus, any distribution $\mathbf{p}$ can be improved by one of the related distributions $\mathbf{p}^0$ or $\mathbf{p}^1$ for any i and k, which proves the claim in this case.

2° Assume now that $D \neq 0$. By (17), the objective function (1) can be written in the following form:

$$w_1 \mathrm{E}[T|\mathbf{p}] + w_2 \mathrm{E}[P|\mathbf{p}] = w_1 A'_1 + w_2 A'_2 + \tfrac{w_1 C'_1 + w_2 C'_2}{D p_{ik} + E},$$

which is clearly a monotonic function of p_{ik} in the interval $[0, 1]$. Thus, also in this case, any distribution $\mathbf{p}$ can be improved by one of the related distributions $\mathbf{p}^0$ or $\mathbf{p}^1$ for any i and k, which completes the proof. □

PROPOSITION 4. *For the generalized ERP cost metric (4), the optimal policy in Π_{mixed} is deterministic, i.e., there are i^* and k^* such that $p_{i^*,k^*} = 1$.*

Proof: 1° Assume first that $D = 0$. By (16), the objective function (1) can be written in the following form:

$$\begin{aligned}
&w \mathrm{E}[T|\mathbf{p}]^a w_2 \mathrm{E}[P|\mathbf{p}]^b = \\
&\quad w \left(A_1 + \tfrac{C_1}{E} + \tfrac{B_1}{E} p_{ik}\right)^a \left(A_2 + \tfrac{C_2}{E} + \tfrac{B_2}{E} p_{ik}\right)^b.
\end{aligned}$$

If B_1 and B_2 have the same sign, then the objective function is clearly a monotonic function of p_{ik} in the interval $[0, 1]$, implying that any distribution $\mathbf{p}$ can be improved by one of the related distributions $\mathbf{p}^0$ or $\mathbf{p}^1$ for any i and k.

From this on (until the end of 1°), we assume that $B_1 B_2 < 0$. Consider now the function $f(x)$ defined for $x \in [0, 1]$ as follows:

$$f(x) = \left(A_1 + \tfrac{C_1}{E} + \tfrac{B_1}{E} x\right)^a \left(A_2 + \tfrac{C_2}{E} + \tfrac{B_2}{E} x\right)^b.$$

By inspecting the first and the second derivatives of $f(x)$, it is a straightforward task to prove that $f(x)$ has its minimum value at $x = 0$ or $x = 1$ (cf. the proof of Proposition 2). It follows that again in this case any distribution $\mathbf{p}$ can be

improved by one of the related distributions $\mathbf{p}^0$ or $\mathbf{p}^1$ for any i and k.

2° Assume now that $D \neq 0$. By (17), the objective function (4) can be written in the following form:

$$w\mathrm{E}[T|\mathbf{p}]^a\mathrm{E}[P|\mathbf{p}]^b = w\left(A_1' + \frac{C_1'}{Dp_{ik}+E}\right)^a\left(A_2' + \frac{C_2'}{Dp_{ik}+E}\right)^b.$$

If C_1' and C_2' have the same sign, then the objective function is clearly a monotonic function of p_{ik} in the interval $[0,1]$, implying that any distribution $\mathbf{p}$ can be improved by one of the related distributions $\mathbf{p}^0$ or $\mathbf{p}^1$ for any i and k.

Thus, from this on, we assume that $C_1'C_2' < 0$. Let us now consider the function $f(x)$ defined for $x \in [0,1]$ as follows:

$$f(x) = \left(A_1' + \frac{C_1'}{Dx+E}\right)^a\left(A_2' + \frac{C_2'}{Dx+E}\right)^b.$$

By again inspecting the first and the second derivatives of $f(x)$, it is easy to prove that $f(x)$ has its minimum value at $x = 0$ or $x = 1$ (cf. the proof of Proposition 2). It follows that again in this case any distribution $\mathbf{p}$ can be improved by one of the related distributions $\mathbf{p}^0$ or $\mathbf{p}^1$ for any i and k, which completes the proof. □

Remark: In [6, Theorem (1)] the optimal switch-on threshold for the ERP cost metric is proven to be $k^* = 1$ for exponentially distributed service times. However, we show, via counter-example, that this optimality result does not hold for generally distributed service times by considering job sizes with high variability.

Mean service time of $\mathrm{E}[S] = 1.0$ and arrival rate of $\lambda = 0.5$ are assumed giving a load value of $\rho = 0.5$. The second moment of service time is assumed to be $\mathrm{E}[S^2] = 30.0$. Furthermore, we assume that the system is equipped with a single sleep state having the same parameter values as the *suspend* state given in Section 6 with a deterministic setup delay of $\mathrm{E}[D] = \mathrm{E}[S] = 1.0$. With these assumptions, the ERP metric of the standard $M/G/1$ system, $I = \infty$, will read as 2560 while that of the sleeping policy with $I = 0$ is 2329 for $k = 1$ and 2230 for $k = 2$. Therefore, the optimal policy, in this case, has a threshold value different from 1. Intuitively, the reason why the optimal k may be greater than 1 is that with highly variable job sizes it does not hurt the overall delay so much to have $k > 1$ and wait for some additional jobs to save more energy, because the delays are anyway dominated by the long busy period delays caused by the highly variable job sizes.

All the optimality results proved above are summarized in the following theorem.

THEOREM 3. *For the cost metrics ERWS (1) and generalized ERP (4), the optimal policy in* Π_{mixed} *has either* $I = 0$ *or* $I = \infty$ *and there are* i^* *and* k^* *such that* $p_{i^*,k^*} = 1$.

Remark: Below we demonstrate, by constructing a counter-example, that the optimal idling time may be different from $I = 0$ and $I = \infty$ if there are multiple terms in the general cost metric (3), unlike as argued in [14].

Assume that new jobs arrive with rate $\lambda = 0.1$ and service times are exponential with mean $\mathrm{E}[S] = 1.0$ so that $\rho = 0.1$. In addition, assume that the only sleep state is state *off*. The corresponding setup delay D_0 is assumed to be deterministic with $D_0 = 1.0$. The power consumption in different states is as follows: $P_{\mathrm{busy}} = P_{\mathrm{setup}} = 1.0$, $P_{\mathrm{idle}} = 0.6$, and $P_{\mathrm{off}} = 0.0$. With these parameters, we have, for the ordinary M/G/1 queue (i.e., $I = \infty$),

$$\mathrm{E}[T_{\mathrm{M/G/1}}] = 1.111, \quad \mathrm{E}[P_{\mathrm{M/G/1}}] = 0.640.$$

Consider now the general cost metric (3) with parameters $w_1 = 1, a_1 = 1, b_1 = 0, w_2 = 3, a_2 = 0, b_2 = 3$ so that the objective function is

$$\mathrm{E}[T] + 3\mathrm{E}[P]^3.$$

For the ordinary M/G/1 queue with $I = \infty$, the objective function takes value 1.898, while for the policy with $I = 0$ and $k = 1$, it is 2.084. With $I = 0$ and $k > 1$, the objective function takes even higher values. On the other hand, it is easy to check that a strictly lower value, 1.776, is achieved when the mean idling time is $\mathrm{E}[I] = 20.7$ and the switch-on threshold $k = 1$. Thus, the optimal idling time is, indeed, different from $I = 0$ and $I = \infty$ in this case.

5. RESETTING THE IDLE TIME TIMER

Thus far, we have assumed, as in [14], that the idling time timer is reset (to 0) only when it expires. Let us now consider a modified system where the timer is reset after each idle period (whether the timer expired or not). We still assume that I can have a general distribution.

Recall from Section 3 that T_{idle} refers to the *aggregate* time during one cycle that the system is in state *idle*. Now, as the timer is reset after each idle period, we clearly have

$$\mathrm{E}[T_{\mathrm{idle}}] = \sum_{j=1}^{\infty} j\,\mathrm{E}[\min\{I,A\}]\,\mathrm{P}\{I \geq A\}^{j-1}\mathrm{P}\{I < A\} = \frac{\mathrm{E}[\min\{I,A\}]}{\mathrm{P}\{I < A\}}, \tag{18}$$

where I denotes the idling time (as before) and A is an independent exponentially distributed random variable with mean $1/\lambda$.

In the analysis part, the only modification needed is to replace $\mathrm{E}[I]$ with $\mathrm{E}[T_{\mathrm{idle}}]$ in Equation (9) for the mean cycle length $\mathrm{E}[C]$ so that

$$\mathrm{E}[C] = \frac{1}{1-\rho}\left(\mathrm{E}[T_{\mathrm{idle}}] + \sum_{k=1}^{\infty} q_k \frac{k}{\lambda} + \sum_{i=0}^{n-1} p_i \mathrm{E}[D_i]\right). \tag{19}$$

With this expression for $\mathrm{E}[C]$, the formulas (11) and (13) for $\mathrm{E}[P]$ and $\mathrm{E}[T]$, respectively, remain still valid. In addition, note that conditions $I = 0$ and $I = \infty$ are equivalent with conditions $T_{\mathrm{idle}} = 0$ and $T_{\mathrm{idle}} = \infty$, respectively. Thus, using exactly the same arguments as in the optimization part (Section 4), we can conclude that the optimality results in Theorem 3 remain valid even when the timer is reset after each idle period.

6. NUMERICAL RESULTS

From the analytical results we already know that there is no need to randomize between sleep states. However, the

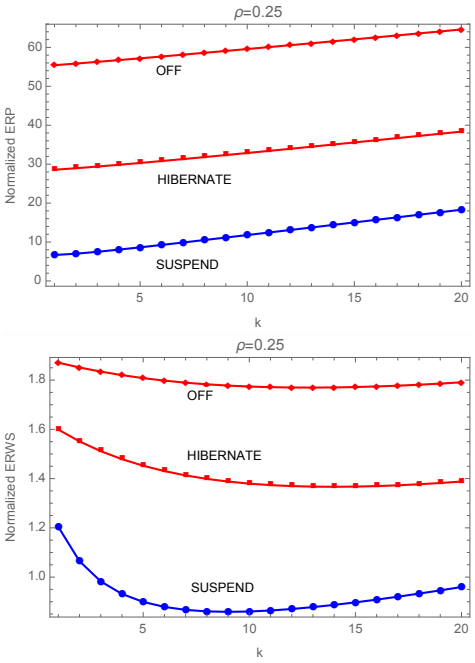

Figure 1: ERP and ERWS of the sleeping policies as a function of the switch-on threshold, normalized with respect to the non energy-aware $M/G/1$ system.

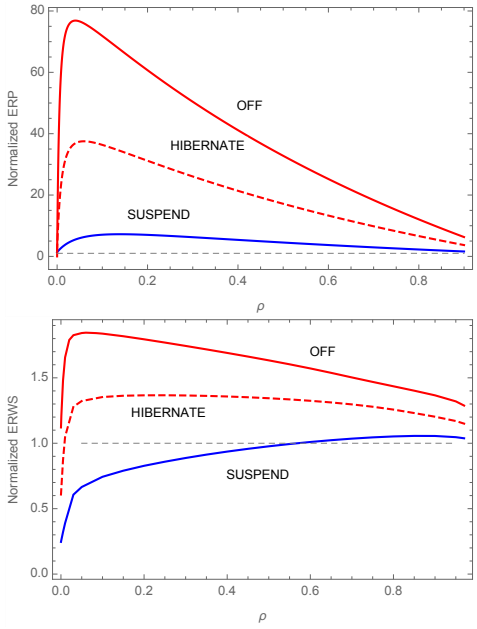

Figure 2: ERP and ERWS of the sleeping policies as a function of system load, normalized with respect to the non energy-aware $M/G/1$ system. At a given load value, each policy is optimized with respect to the switch-on threshold to produce these plots.

optimal sleep state under specific system parameters can only be determined numerically. To address this question we study three energy states that are available in modern servers. These are the common sleep states *suspend* and *hibernate* along with the *off* state, which shuts down the server completely. In the *suspend* state, operational state of the server is saved in the RAM so as to take shorter time to set it up. The *hibernate* state writes the system state in a hard disk and turns all devices off consuming less power than the *suspend* state at a price of higher setup delay.

Power consumption values of $P_{\text{busy}} = P_{\text{setup}} = 200W$, $P_{\text{idle}} = 120W$ $P_{\text{s}} = 15W$, $P_{\text{h}} = 5W$ and $P_{\text{off}} = 0$ will be used in this demonstration with subscripts s and h representing the *suspend* and *hibernate* states, respectively. Moreover, mean setup delays of 100 s, 50 s and 10 s will be used for the *off, hibernate* and *suspend* states, respectively. All values are taken from experimental measurements performed in [7, 10] while the service times obey an exponential distribution, where the mean is exaggerated to $\text{E}[S] = 1$ s for illustration. Weighting factors of $w_1 = 1$ and $w_2 = 0.75$ are used for all the ERWS plots. Using these parameters, the system cost of the sleeping policies relative to the non energy-aware $M/G/1$ system is illustrated in Figures 1-3. Thus, whenever a policy has an ERP or ERWS score of less than 1, it has a lower cost than the standard $M/G/1$ system.

The normalized ERP and ERWS metrics under light load are depicted in Figure 1 as a function of the switch-on threshold. As already proven in [6, Theorem (1)] for exponentially distributed service times, the upper panel shows that $k = 1$ is the optimal choice that minimizes ERP. In the case of ERWS, the lower panel of Figure 1 clearly shows that the optimal value of k may lie elsewhere. In line with the findings in [10] *suspend* gives the best ERP and ERWS amongst the sleep options. This is because setup delay from the *suspend* sleep state is significantly lower while the power consumption is still kept reasonably low. This affects ERP the most since it gives equal weight to both performance and power reductions.

Figure 2 shows the normalized ERP and ERWS metrics as a function of system load. The ERWS curves in the lower panel are produced by first optimizing with respect to the switch-on threshold at a given load value. With the *suspend* state giving the lowest cost, all the sleeping policies are found to have a worse ERP as compared to the standard $M/G/1$ system. However, *suspend* is found to have significant improvement in ERWS for up to moderate system load.

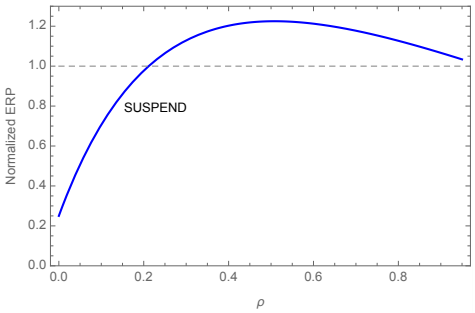

Figure 3: Normalized ERP metric with ideally low setup delay ($\text{E}[T_{\text{setup}}] = \text{E}[S]$).

Motivated by this observation, we studied the ERP of *suspend* further. If technological advancements were to reduce the setup delay to the level of the mean service time, the system would have had the ERP curve given in Figure 3. In this ideal case, a significant reduction in system cost can be achieved by suspending the server whenever the load is low.

7. CONCLUSIONS

We analyzed the energy-performance tradeoff in the M/G/1 queue, where the server can utilize sleep states to reduce the energy consumption. Specifically, we considered policies, where first, after becoming idle, the server waits if the idling time timer expires. If it does, the server goes into sleep state, which may be selected randomly as well as the number of jobs needed in the queue before restarting the server again. However, the cost of reduced energy consumption is the extra delay, the setup delay, needed to get the server operational when waking up from the sleep state. In our model, the service time, the setup delay and the idling time timer may have a general distribution. This model includes as special cases certain previously studied models.

The mean performance and power metrics were derived using standard regenerative techniques. Our main result for ERWS and ERP cost metrics provided the characterization of the optimal policy: either the server never uses any of the sleep states, i.e., only uses the idle state, or the server directly enters some specific sleep state without any idling time. Moreover, the optimal threshold for the the number of jobs in the sleep state was shown to be deterministic. However, we also demonstrated that these results may not hold for more general metrics. Our numerical results highlighted that with realistic values for the power consumption and setup delays, the potential gain from using sleep states may be limited, except at light loads.

8. ACKNOWLEDGEMENTS

This research was partially supported by the TOP-Energy project funded by Academy of Finland (grant no. 268992).

9. REFERENCES

[1] S. Albers and H. Fujiwara. Energy-efficient algorithms for flow time minimization. *ACM Trans. Algorithms*, 3(4), Nov. 2007.

[2] L. L. Andrew, M. Lin, and A. Wierman. Optimality, fairness, and robustness in speed scaling designs. *SIGMETRICS Perform. Eval. Rev.*, 38(1):37–48, Jun. 2010.

[3] N. Bansal, H.-L. Chan, and K. Pruhs. Speed scaling with an arbitrary power function. In *Proc. of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '09)*, pages 693–701, Jan. 2009.

[4] N. Bansal, K. Pruhs, and C. Stein. Speed scaling for weighted flow time. In *Proc. of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '07)*, pages 805–813, Jan. 2007.

[5] O. Boxma, S. Schlegel, and U. Yechiali. A note on the M/G/1 queue with waiting server, timer and vacations. *American Mathematical Society Translations*, 207:25–35, 2002.

[6] A. Gandhi, V. Gupta, M. Harchol-Balter, and M. A. Kozuch. Optimality analysis of energy-performance trade-off for server farm management. *Perform. Eval.*, 67(11):1155–1171, Nov. 2010.

[7] A. Gandhi, M. Harchol-Balter, and M. A. Kozuch. Are sleep states effective in data centers? In *Proc. of the 2012 International Green Computing Conference (IGCC)*, pages 1–10, Jun. 2012.

[8] R. Gonzales and M. Horowitz. Energy dissipation in general purpose microprocessors. *IEEE Journal of Solid-State Circuits*, 31(9):1277–1284, Sep. 1996.

[9] E. Hyytiä, R. Righter, and S. Aalto. Task assignment in a heterogeneous server farm with switching delays and general energy-aware cost structure. *Performance Evaluation*, 75-76:17 – 35, 2014.

[10] C. Isci, S. McIntosh, J. Kephart, R. Das, J. Hanson, S. Piper, R. Wolford, T. Brey, R. Kantner, A. Ng, J. Norris, A. Traore, and M. Frissora. Agile, efficient virtualization power management with low-latency server power states. In *Proc. of the 40th Annual International Symposium on Computer Architecture (ISCA '13)*, pages 96–107, Jun. 2013.

[11] P. Juang, Q. Wu, L.-S. Peh, M. Martonosi, and D. W. Clark. Coordinated, distributed, formal energy management of chip multiprocessors. In *Proc. of the 2005 international symposium on Low power electronics and design (ISLPED Š05)*, pages 127–130, Aug. 2005.

[12] C. W. Kang, S. Abbaspour, and M. Pedram. Buffer sizing for minimum energy-delay product by using an approximating polynomial. In *Proc. of the 13th ACM Great Lakes Symposium on VLSI (GLSVLSI '03)*, pages 112–115, Apr. 2003.

[13] S. Kaxiras and M. Martonosi. *Computer Architecture Techniques for Power-Efficiency*. Morgan and Claypool Publishers, 1st edition, 2008.

[14] V. J. Maccio and D. G. Down. On optimal policies for energy-aware servers. In *Proc. of IEEE 21st International Symposium on Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS 2013)*, pages 31–39, Aug. 2013.

[15] J. Medhi. *Stochastic Models in Queueing Theory*. Academic Press, 2nd edition, 2013.

[16] A. Penttinen, E. Hyytia, and S. Aalto. Energy-aware dispatching in parallel queues with on-off energy consumption. In *Proc. of IEEE 30th International Performance Computing and Communications Conference (IPCCC 2011)*, pages 1–8, Nov. 2011.

[17] M. R. Stan. Optimal voltages and sizing for low power. In *Proc. of the 12th International Conference on VLSI Design - 'VLSI for the Information Appliance' (VLSID '99)*, pages 428–433, Jan. 1999.

[18] H. Takagi. *Queueing Analysis: A Foundation of Performance Evaluation, Vol. 1 : Vacation and Priority Systems*. North-Holland, 1991.

[19] A. Wierman, L. L. H. Andrew, and A. Tang. Power-aware speed scaling in processor sharing systems: Optimality and robustness. *Perform. Eval.*, 69(12):601–622, Dec. 2012.