Performance Evaluation of Power-aware Communication Network Devices

Hiroyuki Okamura and Tadashi Dohi

Department of Information Engineering
Graduate School of Engineering
Hiroshima University, Higashi-Hiroshima, 739–8527, Japan
{okamu, dohi}@rel.hiroshima-u.ac.jp

Abstract. In this paper, we focus on power-aware network devices, and propose a stochastic model to evaluate their performance quantitatively. More precisely, applying the Markovian additive process (MAP) to the arrival stream for network devices, we develop a stochastic dynamic power management (DPM) model with shutdown policy. Two performance measures for power-saving and processing; steady-state power consumption and throughput, are derived analytically for the DPM model. In numerical examples, we investigate the performance of the power-aware communication device with shutdown in terms of power-saving and processing.

1 Introduction

Energy consumption is a critical issue for ubiquitous computing such that a huge number of computer systems are embedded everywhere. Since most ubiquitous systems have to process the tasks imposed with the limited amount of power supplied by battery packs, both technologies on power reduction and power saving are important to design the ubiquitous systems with embedded, portable and wearable architecture. Thus the power management to reduce energy consumption has received considerable attention for recent years.

Generally speaking, since computer systems, including ubiquitous systems, consist of a number of electric components and devices, the main issues on power management are to reduce energy consumption in each component such as IC chip, microprocessor, CPU, disk drive, display *etc.* To save energy consumption for a lot of devices, the software-based power management technique seems to be more effective than the hardware-based power reduction from the view points of cost and flexibility. In fact, note that the devices installed in most systems are capable to reconfiguration by software control. The simplest but effective idea for saving electricity is to stop the electricity supply for the devices when they become idle. In other words, when the devices terminate their own tasks, the controller of software-based power management, *i.e.* a power manager (PM), should shut down the devices immediately. Such a software-based power management is often called *dynamic power management* (DPM). DPM is a control scheme that reconfigures components dynamically to serve the tasks imposed with a minimum number of active components [1, 3]. In DPM, two approaches on reduction of power consumption are available: *shutdown* and *variable-voltage*. The shutdown approach is the simplest way for the power reduction, and is to shut down a device when it has been idle for long time. On the other hand, the variable-voltage approach is to reduce the supply voltage for processors. That is, adjusting system speed at the minimum level necessary to processing the tasks leads to drastically reducing energy consumption.

Benini *et al.* [1] discuss the system-level DPM. In their modeling, the system can be mainly classified into three states: active, idle and sleep. In this framework, by taking account of trade-off relationship between processing performance and energy consumption, PM can control the system under the DPM. For instance, when applying the shutdown and the variable-voltage, we can carry out the performance evaluation under these basic DPM models. In particular, Benini *et al.* [3] describe that the control with fixed shutdown timing takes an advantage of the implementation, but in the fixed shutdown policy, the inadequate (shorter or longer) timing of shutdown, may cause a failure for reducing energy consumption. That is, it implies that there is an optimal control policy to save energy consumption effectively in DPM. Srivastava *et al.* [12] propose predictive shutdown policies to improve the fixed shutdown timing by making decisions as soon as the device becomes idle, based on the successive observations for active and idle states. The shutdown policy by Hwang and Wu [5] uses the weighted sum of the last idle period and the last prediction as the upcoming prediction of the idle period, and improves ones in [12]. Unlike these models, the stochastic DPM models are considered in [2, 4, 10]. Qiu and Pedram [10] and Benini *et al.* [2] study DPM models with shutdown by applying the Markov decision processes. In their models, the workload is modeled by a simple two-state Markov chain, and the problem is formulated as choosing the optimal strategies on each state. Chung *et al.* [4] point out that the policy optimization in [2, 10] is based on a priori knowledge of the system and its workload statistics, but that the workload is generally much harder to characterize in advance. Since workloads should be often non-stationary in general, they consider a DPM model for non-stationary service requests. Similar shutdown approaches can be applied to the hard disk power management on personal computers. Sandoh *et al.* [11] develop a simple renewal reward model and derive the optimal timing at which the hard disk device should go to the sleep state. Lu and De Micheli [6] and Lu *et al.* [7] present the computation algorithms for controlling the power states of a hard disk device. On the other hand, as the DPM models based on stochastic processes, Okamura *et al.* [8, 9] develop some shutdown models based on the renewal structure and queueing theory to represent the number of arrival requests, *i.e.* arrival stream of requests, more strictly. In their works, under the circumstance that the arrival process of requests is according to the Poisson process, it is optimal to shut down the device just after each task is terminated, or not to shut the device at all, depending on the arrival rate.

This paper focuses on power-aware communication network devices such as a wireless LAN. In such devices, the problem on power reduction is critical, because the energy consumption of network devices cannot be ignored for the total consuming electricity for mobile system. After developing a stochastic DPM model for the network device and the mobile communication environment, we derive analytically some performance measures on both power-saving and processing. As mentioned before, if the arrival stream is given by the Poisson process, it is proved that the optimal shutdown timing becomes zero or infinity. However, in the practical communication network, the arrival stream for devices might not obey a simple Poisson process, because there is the correlation of requests. This implies that there are the states of high and low arrival rates in the communication network. Thus it is worth noting to consider the modeling with the correlation of requests. For the modeling with correlation, the Markov modulated Poisson process (MMPP) is often used. The MMPP is the stochastic process in which the arrival rate changes according to a continuous-time Markov chain (CTMC), and is one of the simplest stochastic processes to represent correlation of accesses. Since the Markovian additive process (MAP) includes the MMPP as a special case and allows the arrival rate to change at the occurrence of an arrival, it can model the more complex correlation of accesses.

2 Model Description

2.1 Modeling of Arrival Stream

First we model an arrival stream of requests in the underlying network device. Suppose that an arrival stream for the underlying device has a number of internal states which are called phases, and that the transition of phases occurs according to a CTMC. Let $\{J(t); t \geq 0\}$ be the CTMC which indicates the phase at time t, and the corresponding infinitesimal generator is given by

$$\boldsymbol{M} = \begin{pmatrix} -\sum_j \mu_{1,j} & \mu_{1,2} & \cdots & \mu_{1,m} \\ \mu_{2,1} & -\sum_j \mu_{2,j} & \cdots & \mu_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{m,1} & \mu_{m,2} & \cdots & -\sum_j \mu_{m,j} \end{pmatrix}. \tag{1}$$

For sake of simplification, the number of phases, m (≥ 1), is to be finite. The arrival rate of requests depends on the phase, and is governed by a CTMC of phases. That is, from the Markovian property, the arrival rate can be determined only by the current phase. Define the following matrix whose element indicates the arrival rate of request on the corresponding phase:

$$\boldsymbol{D} = \begin{pmatrix} \lambda_{1,1} & \lambda_{1,2} & \cdots & \lambda_{1,m} \\ \lambda_{2,1} & \lambda_{2,2} & \cdots & \lambda_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m,1} & \lambda_{m,2} & \cdots & \lambda_{m,m} \end{pmatrix}. \tag{2}$$

The diagonal elements of the matrix $\boldsymbol{D}$ indicate the occurrence rates of requests corresponding to each phase. This rate does not include the occurrence rate of phase transition, *i.e.* the corresponding event is that one request occurs but the phase does not change. On the other hand, non-diagonal elements represent the rates of the simultaneous events that a request occurs and the phase changes.

Let $\{N(t); t \geq 0\}$ denote the cumulative number of requests before time t and $\boldsymbol{P}(t)$ be the matrix whose (i, j)-element is given by

$$p_{i,j}(t) = \Pr\{N(t) = n, J(t) = j \mid N(0) = 0, J(0) = i\} \quad i, j = 1, 2, \ldots, m. \tag{3}$$

From the matrix-based Markovian analysis, we obtain the following differential-difference equations:

$$\frac{d}{dt}\boldsymbol{P}_0(t) = \boldsymbol{P}_0(t)\boldsymbol{C}, \tag{4}$$

$$\frac{d}{dt}\boldsymbol{P}_k(t) = \boldsymbol{P}_k(t)\boldsymbol{C} + \boldsymbol{P}_{k-1}(t)\boldsymbol{D}, \quad k = 1, 2, \ldots \tag{5}$$

and

$$\boldsymbol{C} = \boldsymbol{M} - \mathrm{diag}(\boldsymbol{D}\boldsymbol{e}), \tag{6}$$

where $\boldsymbol{e}$ is the column vector whose elements are 1 and $\mathrm{diag}(\boldsymbol{a})$ gives a matrix with the elements of $\boldsymbol{a}$ on the leading diagonal. Then the stochastic process $N(t)$ is called the Markovian additive process (MAP) with parameters $\boldsymbol{C}$ and $\boldsymbol{D}$.

For example, letting $\boldsymbol{C} = -\lambda$ and $\boldsymbol{D} = \lambda$, the corresponding MAP is reduced to the homogeneous Poisson process with the arrival rate λ. Also, we can represent the switched Poisson process (SPP), interrupted Poisson process (IPP) and the Markov modulated Poisson process (MMPP) with the same framework. Further the MAP models an arrival stream in the practical situation. Actually, if we set the parameters as,

$$\boldsymbol{C} = \begin{pmatrix} 0 & 0 \\ \mu & -(\lambda_a + \mu) \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} 0 & \lambda_s \\ 0 & \lambda_a \end{pmatrix}, \tag{7}$$

then the associated MAP can represent the situation where a communicative session starts and terminates at the rates λ_s and μ, respectively, and the requests arrive at the underlying network device with the rate λ_a while the session is alive.

An advantage of the MAP is to represent convolutional of arrival streams easily. Let $\boldsymbol{C}_k$ and $\boldsymbol{D}_k$, $k = 1, \ldots, K$ denote the parameters of MAPs representing the arrival streams by respective peripheral devices. Then the total number of requests for the underlying devices is given by the convolutional of these MAPs with the parameters

$$\boldsymbol{C} = \boldsymbol{C}_1 \oplus \cdots \oplus \boldsymbol{C}_K, \quad \boldsymbol{D} = \boldsymbol{D}_1 \oplus \cdots \oplus \boldsymbol{D}_K, \tag{8}$$

where $\oplus$ is the Kronecker sum.

2.2 Modeling of the network device

Consider a network device which has the three energy consumption states: active, idle and inactive (sleep). The total number of requests from other devices follows

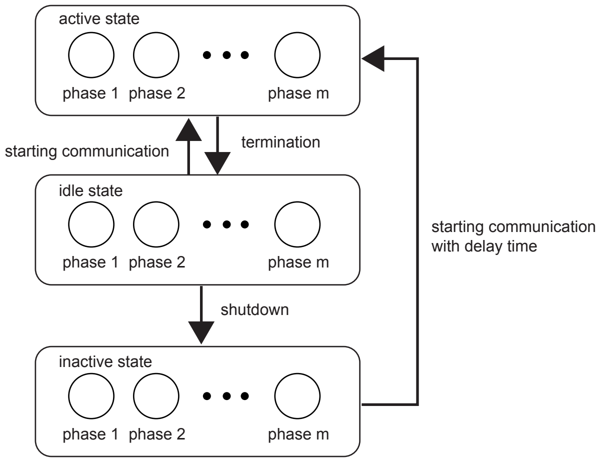

Fig. 1. Configuration of state transition in DPM model.

the MAP with parameters $\boldsymbol{M}$ and $\boldsymbol{D}$. To simplify the discussion, the arrival stream has m (> 0) different states of phases. When a request arrives at the underlying network device, the device communicates with the corresponding device. The time required to communicate the other is given by a random variable S having the probability distribution function $F(t)$. During the communication, the other succeeding requests are completely refused. After the communication is terminated, the device waits for next request in the idle state. If no access occurs for a limited time period t_0, PM shuts down the device to save the electricity, *i.e.* the state of device becomes inactive. In the inactive state, when a request occurs, the device starts to communicate with the other device, where it takes a random time delay D to set up for the communication, having the probability distribution function $H(t)$. On energy consumption, it is assumed that the amounts of energy consumption per unit time in active, idle and inactive states are given by P_1 (> 0), P_2 $(0 < P_2 < P_1)$ and 0, respectively. Figure 1 shows the state transition diagram of DPM model.

3 Performance Measures

Here two kinds of performance measures will be considered in terms of power saving and processing: (i) Steady-state power consumption and (ii) Throughput. To derive these performance measures, we apply the hidden Markovian method. Define the time period between two successive terminations of communication as one cycle. Taking account of the state transition of phase at each termination, the temporal behavior of phase transition is described by a discrete-time Markov chain (DTMC). Let $\boldsymbol{G}_0(t_0)$ denote the transition probability matrix for phases at time instance of communicative termination, provided that the shutdown time is given by t_0. From the matrix-based Markovian analysis [13], the transition probability matrix is given by

$$\begin{aligned}\boldsymbol{G}_0(t_0) &= \int_0^{t_0} \exp(\boldsymbol{C}t)dt\boldsymbol{D}\boldsymbol{F} + \int_{t_0}^{\infty} \exp(\boldsymbol{C}t)dt\boldsymbol{D}\boldsymbol{H}\boldsymbol{F} \\ &= \{\boldsymbol{I} - \exp(\boldsymbol{C}t_0)\}(-\boldsymbol{C})^{-1}\boldsymbol{D}\boldsymbol{F} + \exp(\boldsymbol{C}t_0)(-\boldsymbol{C})^{-1}\boldsymbol{D}\boldsymbol{H}\boldsymbol{F} \\ &= (-\boldsymbol{C})^{-1}\boldsymbol{D}\boldsymbol{F} - \exp(\boldsymbol{C}t_0)(-\boldsymbol{C})^{-1}\boldsymbol{D}\{\boldsymbol{I} - \boldsymbol{H}\}\boldsymbol{F}. \qquad (9)\end{aligned}$$

In this equation, the matrices $\boldsymbol{H}$ and $\boldsymbol{F}$ denote the transition probability matrices for the delay time and the communication time, respectively, where

$$\boldsymbol{H} = \int_0^{\infty} \exp(\boldsymbol{Q}t)dH(t), \quad \boldsymbol{F} = \int_0^{\infty} \exp(\boldsymbol{Q}t)dF(t), \qquad (10)$$

and $\boldsymbol{Q} = \boldsymbol{C} + \boldsymbol{D}$. Let us define the steady-state probabilities for phases on the hidden Markov chain, provided that the shutdown timing is t_0 as

$$\boldsymbol{\pi}(t_0) = [\pi_1(t_0), \pi_2(t_0), \ldots, \pi_m(t_0)]. \qquad (11)$$

Then it is evident that

$$\boldsymbol{\pi}(t_0) = \boldsymbol{\pi}(t_0)\boldsymbol{G}_0(t_0), \quad \boldsymbol{\pi}(t_0)\boldsymbol{e} = 1. \qquad (12)$$

In fact, the computation in Eq. (12) can be reduced into the problem on eigenvalue and eigenvectors for matrices.

(i) Steady-state power consumption

First, we consider the expected time length of one cycle. Since the expected time length depends on the state of phase at the beginning of cycle, *i.e.*, the termination of communication, the expected time length of one cycle is represented by a column vector. When the shutdown timing is given by t_0, the expected time length of one cycle becomes

$$\begin{aligned}\boldsymbol{\tau}(t_0) &= \int_0^{t_0} \{t + \mathrm{E}[S]\} \exp(\boldsymbol{C}t)\boldsymbol{D}\boldsymbol{e}dt \\ &\quad + \int_{t_0}^{\infty} \{t + \mathrm{E}[D] + \mathrm{E}[S]\} \exp(\boldsymbol{C}t)\boldsymbol{D}\boldsymbol{e}dt \\ &= (-\boldsymbol{C})^{-2}\boldsymbol{D}\boldsymbol{e} + \mathrm{E}[S]\boldsymbol{D}\boldsymbol{e} + \mathrm{E}[D]\exp(\boldsymbol{C}t_0)(-\boldsymbol{C})^{-1}\boldsymbol{D}\boldsymbol{e}. \qquad (13)\end{aligned}$$

Similarly, the expected power consumption during one cycle also depends on the state of phases at the beginning of cycle, and is given by

$$\begin{aligned}\boldsymbol{\xi}(t_0) &= \int_0^{t_0} \{P_1 t + P_2\mathrm{E}[S]\} \exp(\boldsymbol{C}t)\boldsymbol{D}\boldsymbol{e}dt \\ &\quad + \int_{t_0}^{\infty} \{P_1 t_0 + P_2(\mathrm{E}[D] + \mathrm{E}[S])\} \exp(\boldsymbol{C}t)\boldsymbol{D}\boldsymbol{e}dt\end{aligned}$$

$$= P_1(-\boldsymbol{C})^{-2}\boldsymbol{De} + P_2\mathrm{E}[S](-\boldsymbol{C})^{-1}\boldsymbol{De}$$

$$+ \{P_2\mathrm{E}[D]\boldsymbol{I} - P_1(-\boldsymbol{C})^{-1}\}\exp(\boldsymbol{C}t_0)(-\boldsymbol{C})^{-1}\boldsymbol{De}. \quad (14)$$

From the renewal reward argument on the Markov renewal process [14], the steady-state power consumption is given by

$$C_{ss}(t_0) = \lim_{t\to\infty} \frac{\mathrm{E}[\text{the amount of consuming electricity during } [0,t)]}{t}$$

$$= \frac{\boldsymbol{\pi}(t_0)\boldsymbol{\xi}(t_0)}{\boldsymbol{\pi}(t_0)\boldsymbol{\tau}(t_0)}. \quad (15)$$

(ii) Throughput

The throughput for the network device is defined as

$$T_{ss}(t_0) = \lim_{t\to\infty} \frac{\mathrm{E}[\text{the number of achieved tasks during } [0,t)]}{t}. \quad (16)$$

Similar to the steady-state power consumption, since the network device always processes only one task requested during one cycle, the throughput can be derived as

$$T_{ss}(t_0) = \frac{1}{\boldsymbol{\pi}(t_0)\boldsymbol{\tau}(t_0)}. \quad (17)$$

4 Performance Evaluation

In this section, we investigate the effectiveness of shutdown policy on the network device under the performance measures. In particular, we consider the following four kinds of arrival streams:

Case 1: Poisson process with the parameters:

$$\boldsymbol{C} = -\lambda, \quad \boldsymbol{D} = \lambda. \quad (18)$$

Case 2: IPP with 2-phases

$$\boldsymbol{C} = \begin{pmatrix} -\mu_1 & \mu_1 \\ \mu_2 & -(\mu_2+\lambda) \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix}. \quad (19)$$

Case 3: IPP with 2-phases and the first arrival causes succeeding requests

$$\boldsymbol{C} = \begin{pmatrix} -\mu_1 & 0 \\ \mu_2 & -(\mu_2+\lambda) \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} 0 & \mu_1 \\ 0 & \lambda \end{pmatrix}. \quad (20)$$

Case 4: Convolutional of two IPPs in Case 3

$$\boldsymbol{C} = \begin{pmatrix} -2\mu_1 & 0 & 0 & 0 \\ \mu_2 & -(\mu_1+\mu_2+\lambda) & 0 & 0 \\ \mu_2 & 0 & -(\mu_1+\mu_2+\lambda) & 0 \\ 0 & \mu_2 & \mu_2 & -2(\mu_2+\lambda) \end{pmatrix}, \quad (21)$$

$$\boldsymbol{D} = \begin{pmatrix} 0 & \mu_1 & \mu_1 & 0 \\ 0 & \lambda & 0 & \mu_1 \\ 0 & 0 & \lambda & \mu_1 \\ 0 & 0 & 0 & 2\lambda \end{pmatrix}. \tag{22}$$

As shown in Section 3, the arrival stream in Case 3 represents the situation where a communicative session starts and terminates at the rates μ_1 and μ_2, respectively, and where the requests occur at the rate λ while the session is alive. In Case 4, we consider the convolutional of such two arrival streams. Also, it is assumed that both communication time and delay time obey the Gamma distributions with respective shape and scale parameters; (m_s, η_s) and (m_d, η_d), *i.e.*,

$$F(t) = \int_0^t \frac{\eta_s^{m_s} u^{m_s-1} e^{-\eta_s u}}{\Gamma(m_s)} du, \quad H(t) = \int_0^t \frac{\eta_d^{m_d} u^{m_d-1} e^{-\eta_d u}}{\Gamma(m_d)} du. \tag{23}$$

Under these assumptions, the mean times of communication and delay are given by

$$\mathrm{E}[S] = \frac{m_s}{\eta_s}, \quad \mathrm{E}[D] = \frac{m_d}{\eta_d}. \tag{24}$$

If PM does not perform the shutdown, then the traffic intensity is given by

$$\rho = \boldsymbol{\pi}_0 \boldsymbol{D}\boldsymbol{e}\mathrm{E}[S]. \tag{25}$$

Note that this is not an actual traffic intensity, because the additional arrivals are canceled while the network device communicates with one device, but is a appropriate measure to evaluate the busyness of network device. In this paper, we set the above traffic intensity as $\rho = 0.3, 0.5$ by adjusting the parameter λ in each case.

The basic parameters are set as follows.

$\rho = 0.3, 0.5,\quad \mu_1 = 0.05,\quad \mu_2 = 0.05,\quad m_s = 2.0,\quad \mathrm{E}[S] = 1.0,$
$m_d = 2.0,\quad \mathrm{E}[D] = 1.0,\quad P_1 = 1.0,\quad P_2 = 10.0$

Figure 2 shows the behavior of steady-state power consumption and throughput. Especially, we illustrate the dependence of each measure on varying shutdown timing. From Fig. 2, except for Case 1, *i.e.*, Poisson arrival stream case, there is a unique optimal shutdown timing which minimizes the steady-state power consumption. Also, the minimum expected power consumptions in Case 2 and Case 3 are lower than those in the other cases, so that the effectiveness of power-saving in Case 2 and Case 3 is relatively higher. Next, we focus on the throughput. In this result, the processing performance of network device becomes higher as the shutdown timing is longer. Although the traffic intensities are fixed as 0.3 in all the cases, the throughput in Case 1 is better than the other cases. In other words, the evaluation of performance is considerably optimistic. This is because the Poisson arrival stream has no correlation of arrivals. Figure 3 shows the results in the case of $\rho = 0.5$. For these results, we can make similar observations to the case of $\rho = 0.3$. Thus the optimal shutdown timing strongly

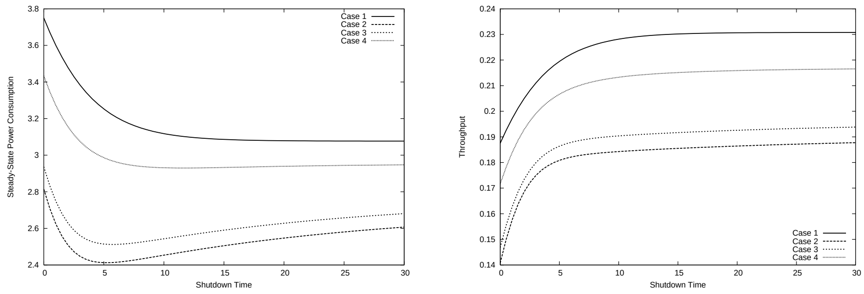

Fig. 2. Performance measures ($\rho = 0.3$).

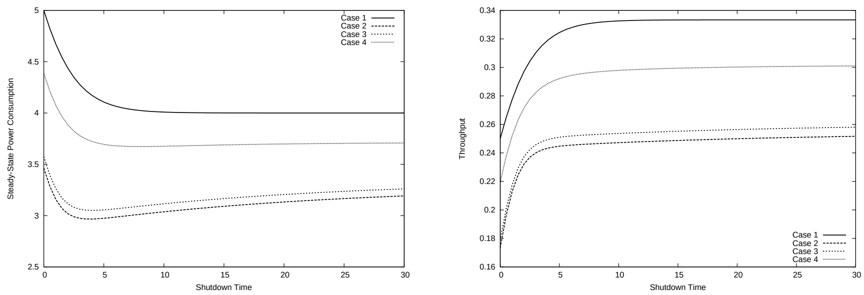

Fig. 3. Performance measures ($\rho = 0.5$).

depends on the correlation of arrivals, rather than the quantities of network traffic. In the situation where the arrival streams are IPPs, the optimal shutdown timing leads to higher power-saving performance in the network device.

5 Conclusion

In this paper, we have considered the power-saving for network devices. Applying the MAP to the arrival stream, we have proposed the DPM model with shutdown policy. Furthermore, based on the DPM model, two performance measures for power-saving and processing have been delivered from the renewal reward theory. In the numerical examples, for four kinds of arrival streams, we have examined the difference on the steady-state power consumption and throughput. As a result, the power-saving performance strongly depends on the correlation of arrival requests. In future, we will develop the DPM model for the network device with communication buffer and the computation algorithm for the optimal shutdown timing.

References

1. L. Benini and G. De Micheli, *Dynamic Power Management: Design Techniques and CAD Tools*, Kluwer Academic Publishers, New York, 1997.
2. L. Benini, A. Bogliolo, G. A. Paloologo, and G. De Micheli, Policy optimization for dynamic power management, *IEEE Trans. on Computer-Aided Design of Circuits Systems*, vol. 18, no. 6, pp. 813–833, 1999.
3. L. Benini and G. De Micheli, System-level power optimization: techniques and tools, *ACM Trans. on Design Automation of Electronic Systems*, vol. 5, no. 2, pp. 115–192, 2000.
4. E. Chung, L. Benini, A. Bogliolo, and G. De Micheli, Dynamic power management for non-stationary service requests, *Proc. 1999 Design Automation and Test in Europe*, pp. 77–81, IEEE Computer Society Press, Los Alamitos, 1999.
5. C. Hwang and A. C. Wu, Predictive system shutdown method for energy saving of event-driven computation, *ACM Trans. on Design Automation of Electronic Systems*, vol. 5, no. 2, pp. 226–241, 2000.
6. Y. Lu and G. De Micheli, Adoptive hard disk power management on personal computers, *Proc. 9th Great Lakes Symp. on VLSI*, pp. 50–53, IEEE Computer Society Press, Los Alamitos, 1999.
7. Y. Lu, T. Šimunić, and G. De Micheli, Software controlled power management, *Proc. 7th Int'l Workshop on Hardware/Software Codesign*, pp. 157–161, ACM Press, New York, 1999.
8. H. Okamura, T. Dohi, and S. Osaki. The phase type approximation for the optimal auto-sleep scheduling. *Mathematical and Computer Modelling*, vol. 38, nos. 11/13, pp. 1391–1398, 2003.
9. H. Okamura, T. Dohi, and S. Osaki. A structural approximation method to generate the optimal auto-sleep schedule for a computer system. *Computers and Mathematics with Applications*, vol. 46, no. 7, pp. 1103–1110, 2003.
10. Q. Qiu and M. Pedram, Dynamic power management based on continuous-time Markov decision processes, *Proc. 36th Design Automation Conference*, pp. 555–561, IEEE Computer Society Press, Los Alamitos, 1999.
11. H. Sandoh, H. Hirakoshi, and H. Kawai, An optimal time to sleep for an auto-sleep system, *Computers & Operations Research*, vol. 23, pp. 221–227, 1996.
12. M. Srivastava, A. Chandrakasan, and B. Brodersen, Predictive system shutdown and other architectural techniques for energy efficient programmable computation, *IEEE Trans. on Very Large Scale Integrated Systems*, vol. 4, no. 1, pp. 42–55, 1996.
13. M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Johns Hopkins University Press, Baltimore, 1981.
14. D. R. Cox, *Renewal theory*, John Wiley & Sons Inc., London, 1962.