Department of Social Systems and Management

Discussion Paper Series

No. 1188

Analysis of Effective Service Time under MMPP Service Interruptions with State Dependent Interruption Periods

by

Ushio Sumita and Kazuki Takahashi

February 2008

UNIVERSITY OF TSUKUBA
Tsukuba, Ibaraki 305-8573
JAPAN

Analysis of Effective Service Time under MMPP Service Interruptions with State Dependent Interruption Periods

Ushio Sumita and Kazuki Takahashi

Graduate School of Systems and Information Engineering,
University of Tsukuba,
1-1-1, Tennoudai, Tsukuba, Ibaraki, 305-8573, Japan
(e-mail: {sumita,takaha11}@sk.tsukuba.ac.jp)

Abstract. A stochastic model is considered where a service time having a general distribution is delayed by service interruptions generated by an MMPP(Markov Modulated Poisson Process). More specifically, such interruptions occur according to a Poisson process with intensity λ_i whenever the underlying Markov chain is in state i. The distribution of the interruption period depends on the state of the Markov chain at the time of the interruption. The matrix Laplace transform of the effective service time is derived explicitly, where the matrix is formed for the initial state i of the Markov chain and the state j of the Markov chain upon completion of the service time. The mean and variance are obtained explicitly.
Keywords. Effective service time, Markov modulated Poisson process.

1 Introduction

The service process subject to random interruptions due to system failures, arrivals of high priority customers, server vacations and the like has been studied extensively in the literature. The effective service time (or the service completion time) under Poisson interruptions having i.i.d. interruption periods and a general service time distribution has been analyzed by Gaver(1962), Jaiswal(1961,1968) and Keilson(1982). The total time spent in system for an M/G/1 priority queueing system with preempt/resume service discipline has been studied by Keilson and Sumita(1983), yielding both transient and ergodic results. Federgruen and Green(1986) have dealt with an M/G/1 queueing system where service interruptions are captured by an alternating renewal process consisting of on-periods and off-periods, deriving bounds and approximations for important ergodic performance measures. A service system subject to inhomogeneous Poisson interruptions with age dependent interruption periods has been examined by Sumita et al(1989) with application to optimal rollback policy for database management.

In this paper, we study a stochastic model where a service time having a general distribution is delayed by service interruptions generated by an MMPP(Markov Modulated Poisson Process). More specifically, such interruptions occur according to a Poisson process with intensity λ_i whenever the underlying Markov chain is in state i. The distribution of the interruption period depends on the state of the Markov chain at the time of the interruption. The matrix Laplace transform of the effective service time is derived explicitly, where the matrix is formed for the initial state i of the Markov chain and the state j of the Markov chain upon completion of the service time. The mean and the variance are also obtained.

The structure of the paper is as follows. In Section 2, we formally introduce a stochastic model describing the service process with MMPP service interruptions. By examining the probabilistic flow of the underlying multivariate process, the matrix Laplace transform of the effective service time is derived explicitly in Section 3. The mean and the variance are also obtained. Some numerical examples are presented in Section 4.

2 Model description

A stochastic model is considered where the service to a customer commences at time $t = 0$. The customer has a random service requirement S that is absolutely continuous with $\bar{A}(x) \stackrel{\text{def}}{=} \mathrm{P}[S > x], a(x) \stackrel{\text{def}}{=} -\frac{d}{dx}\bar{A}(x), \eta_A(x) \stackrel{\text{def}}{=} \frac{a(x)}{\bar{A}(x)}$ and $\alpha(w) \stackrel{\text{def}}{=} \int_0^\infty e^{-wx} a(x) dx$. This service time is subject to interruptions of preemptive-resume type, where such interruptions occur according to a MMPP characterized by $[\underline{\underline{\nu}}, \underline{\underline{\Lambda}}_D]$ with state dependent interruption periods. Here, $\underline{\underline{\Lambda}}_D \stackrel{\text{def}}{=} [\delta_{\{i=j\}}\lambda_i]_{i,j\in\mathcal{J}}$ with $\mathcal{J} \stackrel{\text{def}}{=} \{1, 2, \cdots, J\}$ and $\underline{\underline{\nu}} \stackrel{\text{def}}{=} [\nu_{ij}]$ is a hazard rate matrix. The infinitesimal generator $\underline{\underline{Q}}$ is then given by $\underline{\underline{Q}} \stackrel{\text{def}}{=} -\underline{\underline{\nu}}_D + \underline{\underline{\nu}}$, where $\underline{\underline{\nu}}_D \stackrel{\text{def}}{=} [\delta_{\{i=j\}}\nu_i]$ and $\nu_i \stackrel{\text{def}}{=} \sum_{j\in\mathcal{J}} \nu_{ij}$. More formally, let $\{J(t) : t \geqslant 0\}$ be a Markov chain in continuous time on $\mathcal{J}$ governed by $\underline{\underline{\nu}}$. Then service interruptions occur according to a Poisson process with intensity λ_i whenever $J(t)$ is in state i. The interruption period denoted by T_{INT} is a nonnegative random variable with $B_{ij}(y) \stackrel{\text{def}}{=} \mathrm{P}[T_{\text{INT}} \leqslant y, J(T_{\text{INT}}) = j | J(0) = i], \bar{B}_{ij}(y) \stackrel{\text{def}}{=} 1 - B_{ij}(y), b_{ij}(y) \stackrel{\text{def}}{=} \frac{\partial}{\partial y} B_{ij}(y), \bar{B}_i(y) \stackrel{\text{def}}{=} \sum_{j\in\mathcal{J}} \bar{B}_{ij}(y), \eta_{B:ij}(y) \stackrel{\text{def}}{=} \frac{b_{ij}(y)}{\bar{B}_i(y)}$ and $\beta_{ij}(v) \stackrel{\text{def}}{=} \int_0^\infty e^{-vy} b_{ij}(y) dy$. The matrices for $b_{ij}(y)$ and $\beta_{ij}(v)$ are defined as $\underline{\underline{b}}(y)$ and $\underline{\underline{\beta}}(v)$ respectively. During an interruption period with $J(t)$ in state j, the interruption period has the competing hazard rates $\eta_{B:jk}(y)$ for completion with the Poisson interruption intensity switched to λ_k upon completion accordingly. The service to the customer resumes from the point of the previous interruption and this process continues until the service to the customer is completed. Of interest, then, is the effective service time S_{eff} representing the time interval between the service commencement and the service completion.

We now introduce a stochastic process $\{I(t) : t > 0\}$ defined by

$$I(t) = \begin{cases} 0 & \text{if the service to the customer has been completed by time } t \ , \\ 1 & \text{if the server is in service at time } t \ , \\ 2 & \text{if the server is in an interruption period at time } t \ . \end{cases}$$

Since the service to the customer commences at time $t = 0$, one has $I(0) = 1$. The cumulative service time given to the customer until time t is denoted by $X(t)$ provided that $I(t) \in \{1, 2\}$. When $I(t) = 2$, $Y(t)$ is defined as the elapsed time since the last service interruption and $Y(t) = 0$ if $I(t) \neq 2$. The state 0 of $I(t)$ is absorbing and both $J(t)$ and $X(t)$ freeze as soon as $I(t)$ enters this absorbing state.

In the next section, we analyze the multivariate stochastic process $[X(t), Y(t), I(t), J(t)]$, which is Markov, by examining its probabilistic flow in the state space $\mathcal{R}^+ \times \mathcal{R}^+ \times \{0, 1, 2\} \times \mathcal{J}$. This in turn enables one to derive the matrix transform involving S_{eff}, $J(0)$ and $J(S_{\text{eff}})$. A typical sample path of $[X(t), Y(t), I(t), J(t)]$ is depicted in Figure 1.

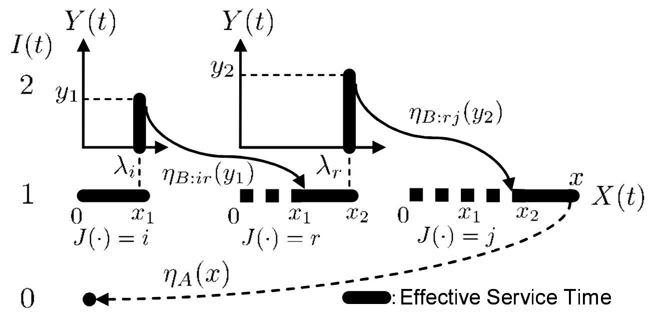

Fig. 1. Typical sample path of $[X(t), Y(t), I(t), J(t)]$

3 Analysis of effective service time with MMPP service interruptions

The key entities of interest for analyzing S_{eff} are $\underline{\underline{S}}_{\text{eff}}(t) = [S_{\text{eff}:ij}(t)]$; $S_{\text{eff}:ij}(t) \overset{\text{def}}{=} \text{P}[S_{\text{eff}} \leqslant t, J(S_{\text{eff}}) = j | I(0) = 1, J(0) = i]$ and

$$\underline{\underline{\sigma}}_{\text{eff}}(s) \overset{\text{def}}{=} \int_0^\infty e^{-st} \underline{\underline{s}}_{\text{eff}}(t) dt \ ; \ \underline{\underline{s}}_{\text{eff}}(t) \overset{\text{def}}{=} \frac{d}{dt} \underline{\underline{S}}_{\text{eff}}(t) \ . \tag{1}$$

For the multivariate process $[X(t), Y(t), I(t), J(t)]$, let $\underline{\underline{F}}_0(t)$ be defined by $F_{0:ij}(t) \overset{\text{def}}{=} \text{P}[I(t) = 0, J(t) = j | I(0) = 1, J(0) = i]$ and $\underline{\underline{F}}_0(t) \overset{\text{def}}{=} [F_{0:ij}(t)]$. We also define

$$\underline{\underline{\hat{f}}}_0(s) \overset{\text{def}}{=} \int_0^\infty e^{-st} \underline{\underline{f}}_0(t) dt \ ; \ \underline{\underline{f}}_0(t) \overset{\text{def}}{=} \frac{d}{dt} \underline{\underline{F}}_0(t) \ . \tag{2}$$

Since state 0 is the unique absorbing state of $I(t)$ and $J(t)$ freezes as soon as $I(t)$ enters state 0, one sees that $S_{\text{eff}:ij}(t) = \text{P}[S_{\text{eff}} \leqslant t, J(S_{\text{eff}}) = j | I(0) = 1, J(0) = i] = \text{P}[I(t) = 0, J(t) = j | I(0) = 1, J(0) = i] = F_{0:ij}(t)$. From (1) and (2), this equation leads to $\underline{\underline{\sigma}}_{\text{eff}}(s) = \underline{\underline{\hat{f}}}_0(s)$. In order to find $\underline{\underline{\sigma}}_{\text{eff}}(s)$, we define $F_{1:ij}(x, t) \overset{\text{def}}{=} \text{P}[X(t) \leqslant x, I(t) = 1, J(t) = j | I(0) = 1, J(0) = i]$,

$\underline{\underline{F}}_1(x,t) \stackrel{\text{def}}{=} [F_{1:ij}(x,t)]$ and $\underline{\underline{f}}_1(x,t) \stackrel{\text{def}}{=} \frac{\partial}{\partial x}\underline{\underline{F}}_1(x,t)$. The matrix Laplace transform of $\underline{\underline{f}}_1(x,t)$ is defined by $\hat{\underline{\underline{f}}}_1(x,s) \stackrel{\text{def}}{=} \int_0^\infty e^{-st}\underline{\underline{f}}_1(x,t)dt$. When $I(t)=2$, let $\tau_{1\to 2}$ be the time at which the current interruption occurred, i.e., $\tau_{1\to 2} \stackrel{\text{def}}{=} \sup\{\tau : Y(\tau-)=0,\ Y(\tau+)>0,\ 0 \leqslant \tau \leqslant t\}$. By using $\tau_{1\to 2}$, the joint distribution function and the joint density function of $[X(t),Y(t),I(t),J(t)]$ when $J(t)$ is in state r at the time of occurrence of the current interruption are denoted by $F_{2:ir}(x,y,t) \stackrel{\text{def}}{=} \mathrm{P}[X(t)\leqslant x, Y(t)\leqslant y, I(t)=2, J(\tau_{1\to 2})=r | I(0)=1, J(0)=i]$ and $f_{2:ir}(x,y,t) \stackrel{\text{def}}{=} \frac{\partial^2}{\partial x \partial y}F_{2:ir}(x,y,t)$.

We begin our analysis by investigating the probabilistic flow of the multivariate process $[X(t),Y(t),I(t),J(t)]$ centering on the state of $I(\cdot)$ at 1 given that $I(0)=1$ and $J(0)=i$. For the multivariate process to be at $(x+\Delta,0,1,j)$ for $x>0$ at time $t+\Delta$, one of the following three cases should have happened.

- The multivariate process was already at $(x,0,1,j)$ at time t and no state change has occurred in $[t,t+\Delta)$ except $X(t)=x$ increased to $X(t+\Delta)=x+\Delta$.
- The multivariate process was at $(x,y,2,j)$ at time t with $J(t-y)=r$ for some $r\in\mathcal{J}$, where $\tau_{1\to 2}=t-y$ and y is the length of the current interruption. Then the interruption period has been completed in $[t,t+\Delta)$ and $J(t)$ has moved from r to j at the competing hazard rate of $\eta_{B:rj}(y)$.
- The multivariate process was at $(x,0,1,m)$ at time t for some $m\in\mathcal{J}$, and $J(t)$ has moved from m to j in $[t,t+\Delta)$ with $I(t)$ in state 1.

Combining these three cases, one observes that $f_{1:ij}(x+\Delta,t+\Delta)=[1-\{\lambda_j+\nu_j+\eta_A(x)\}\Delta]f_{1:ij}(x,t)+\int_0^\infty\sum_{r\in\mathcal{J}}f_{2:ir}(x,y,t)\eta_{B:rj}(y)\Delta dy+\sum_{m\in\mathcal{J}}f_{1:im}(x,t)\nu_{mj}\Delta+o(\Delta)$ for $x,y>0$. By dividing both sides of this equation by Δ and letting $\Delta\to 0$, one has

$$\Big(\frac{\partial}{\partial x}+\frac{\partial}{\partial t}\Big)f_{1:ij}(x,t) = -\big\{\lambda_j+\nu_j+\eta_A(x)\big\}f_{1:ij}(x,t) + \int_0^\infty \sum_{r\in\mathcal{J}} f_{2:ir}(x,y,t)\eta_{B:rj}(y)dy + \sum_{m\in\mathcal{J}} f_{1:im}(x,t)\nu_{mj}\ . \quad (3)$$

In order to understand the probabilistic flow of the multivariate process centering on the state of $I(\cdot)$ at 2, we suppose that $[X(t),Y(t),I(t),J(t)]=(x,y,2,j)$ with $J(\tau_{1\to 2})=r$. This means that $\tau_{1\to 2}=t-y$ and the current interruption lasts with probability $\bar{B}_r(y)$. Consequently, one sees that $f_{2:ir}(x,y,t)=\lambda_r f_{1:ir}(x,t-y)\bar{B}_r(y)$. Substitution of this expression into (3) yields that $(\frac{\partial}{\partial x}+\frac{\partial}{\partial t})f_{1:ij}(x,t)=-\{\lambda_j+\nu_j+\eta_A(x)\}f_{1:ij}(x,t)+\int_0^\infty\sum_{r\in\mathcal{J}}\lambda_r f_{1:ir}(x,t-y)b_{rj}(y)dy+\sum_{m\in\mathcal{J}}f_{1:im}(x,t)\nu_{mj}$. This differential equation can be rewritten in matrix form as

$$\Big(\frac{\partial}{\partial x}+\frac{\partial}{\partial t}\Big)\underline{\underline{f}}_1(x,t) = -\underline{\underline{f}}_1(x,t)\big\{\underline{\underline{\Lambda}}_D - \underline{\underline{Q}} + \eta_A(x)\underline{\underline{I}}\big\} + \int_0^\infty \underline{\underline{f}}_1(x,t-y)\underline{\underline{\Lambda}}_D\underline{\underline{b}}(y)dy\ , \quad (4)$$

where $\underline{\underline{I}}$ is the J-dimensional identity matrix. We are now in a position to prove the key theorem of this paper.

Theorem 1 *For $\underline{\underline{\sigma}}_{\mathit{eff}}(s)$ defined in (1), one has $\underline{\underline{\sigma}}_{\mathit{eff}}(s) = \alpha(\underline{\underline{\zeta}}(s))$, where $\underline{\underline{\zeta}}(s) \stackrel{def}{=} s\underline{\underline{I}} - \underline{\underline{Q}} + \underline{\underline{\Lambda}}_D - \underline{\underline{\Lambda}}_D \underline{\underline{\beta}}(s)$.*

Proof. By taking Laplace transforms of (4) with respect to t, one finds that $\frac{\partial}{\partial x}\hat{\underline{\underline{f}}}_1(x,s) + \hat{\underline{\underline{f}}}_1(x,s)\underline{\underline{H}}(x,s) = \underline{\underline{0}}$ since $\underline{\underline{f}}_1(x,0+) = \underline{\underline{0}}$, where $\underline{\underline{H}}(x,s) \stackrel{\text{def}}{=} \eta_A(x)\underline{\underline{I}} + \underline{\underline{\zeta}}(s)$. In this matrix differential equation, multiplying $e^{\int_0^x \underline{\underline{H}}(z,s)dz}$ from the right and integrating with respect to x lead to

$$\hat{\underline{\underline{f}}}_1(x,s)e^{\int_0^x \underline{\underline{H}}(z,s)dz} = \hat{\underline{\underline{f}}}_1(0+,s) \ . \tag{5}$$

One has $\underline{\underline{f}}_1(0+,t) = \delta(t)\underline{\underline{I}}$ so that the corresponding Laplace transform with respect to t is given by $\hat{\underline{\underline{f}}}_1(0+,s) = \underline{\underline{I}}$. Substituting this expression into (5), one has $\hat{\underline{\underline{f}}}_1(x,s)e^{\int_0^x \underline{\underline{H}}(z,s)dz} = \underline{\underline{I}}$. Since $\bar{A}(x) = \exp[-\int_0^x \eta_A(\tau)d\tau]$, multiplying $e^{-\int_0^x \underline{\underline{H}}(z,s)dz}$ from the right then yields

$$\hat{\underline{\underline{f}}}_1(x,s) = \bar{A}(x)e^{-\underline{\underline{\zeta}}(s)x} \ . \tag{6}$$

We now turn our attention to analysis of $\hat{\underline{\underline{f}}}_0(s)$. One sees that $\underline{\underline{f}}_0(t) = \int_0^\infty \underline{\underline{f}}_1(x,t)\eta_A(x)dx$. Taking the Laplace transform with respect to t in this expression, it can be seen that $\underline{\underline{\sigma}}_{\text{eff}}(s) = \hat{\underline{\underline{f}}}_0(s) = \int_0^\infty \hat{\underline{\underline{f}}}_1(x,s)\eta_A(x)dx$. By substituting (6) into this equation, the theorem follows.

Let $\mu_{\text{eff}:i:n}$ be the n-th moment of S_{eff} and let $\underline{\mu}_{\text{eff}:n}$ be the corresponding vector. The following theorem then holds from Theorem 1.

Theorem 2 *Let $\hat{\underline{\underline{\mu}}}_{A:k} \stackrel{def}{=} \int_0^\infty e^{\underline{\underline{Q}}x} x^k a(x)dx$, $\underline{\underline{\mu}}_{B:k} \stackrel{def}{=} [E[T_{INT}^k | J(T_{INT}) = j, J(0) = i] = (-1)^n(\frac{d}{ds})^n \underline{\underline{\beta}}(s)|_{s=0}$ and $\underline{1} \stackrel{def}{=} [1 \ \cdots \ 1]^\top$.*

a) $\underline{\mu}_{\mathit{eff}:1} = (\underline{\underline{I}} + \underline{\underline{\Lambda}}_D \underline{\underline{\mu}}_{B:1})\hat{\underline{\underline{\mu}}}_{A:1}\underline{1}$

b) $\underline{\mu}_{\mathit{eff}:2} = \{(\underline{\underline{I}} + \underline{\underline{\Lambda}}_D \underline{\underline{\mu}}_{B:1})^2 \hat{\underline{\underline{\mu}}}_{A:2} + \underline{\underline{\Lambda}}_D \underline{\underline{\mu}}_{B:2} \hat{\underline{\underline{\mu}}}_{A:1}\}\underline{1}$

4 Numerical examples

In this section, numerical examples are provided by using Theorem 2. For the underlying Markov chain, the uniformization procedure of Keilson(1979) is employed, where $\nu \geqslant \max\{\nu_i\}_{i\in\mathcal{J}}$ and $\underline{\underline{a}}_\nu = [a_{\nu:ij}] \stackrel{\text{def}}{=} \underline{\underline{I}} + \frac{1}{\nu}\underline{\underline{Q}}$. We consider $\alpha(s)$ in the following three cases: $\alpha_{\text{IFR}}(s) = \frac{\theta_{\alpha:1}}{s+\theta_{\alpha:1}} \cdot \frac{\theta_{\alpha:2}}{s+\theta_{\alpha:2}}$, $\alpha_{\text{DFR}}(s) = q_\alpha \frac{\theta_{\alpha:1}}{s+\theta_{\alpha:1}} + (1-q_\alpha)\frac{\theta_{\alpha:2}}{s+\theta_{\alpha:2}}$ and $\alpha_{\text{CFR}}(s) = \frac{\nu}{s+\nu}$. $\beta_{\text{IFR}:ij}(s)$, $\beta_{\text{DFR}:ij}(s)$ and $\beta_{\text{CFR}:ij}(s)$ are defined similarly. The parameters are set in such a way that the underlying distributions for $\alpha(s)$ (or $\beta_{ij}(s)$) share the same mean, where

$$\underline{\underline{\nu}} = \begin{bmatrix} 0.1 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.1 & 0.1 & 0.3 & 0.1 & 0.1 \\ 0.2 & 0.1 & 0.1 & 0.1 & 0.2 \\ 0.3 & 0.2 & 0.1 & 0.1 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.1 \end{bmatrix}, \ \underline{\underline{\Lambda}}_D = \begin{bmatrix} 2 & & & & \underline{0} \\ & 1 & & & \\ & & 2 & & \\ & & & 4 & \\ \underline{0} & & & & 3 \end{bmatrix},$$

$\nu = 1$ and $[p_i(j)] = 0.2\underline{\underline{I}}$. We also set $(\theta_{\alpha:1}, \theta_{\alpha:2}, q_\alpha) = (1.000102, 100, 0.99)$,

$$[\theta_{\beta:ij1}] = \begin{bmatrix} 1 & 100 & 100 & 100 & 100 \\ 2.02 & 100 & 100 & 100 & 100 \\ 100 & 100 & 100 & 100 & 100 \\ 100 & 100 & 100 & 100 & 100 \\ 100 & 100 & 100 & 100 & 100 \end{bmatrix}, \ [\theta_{\beta:ij2}] = \begin{bmatrix} 100 & 1.005 & 1.005 & 1.005 & 1.005 \\ 100 & 0.5013 & 0.6689 & 2.0202 & 2.0202 \\ 1.005 & 2.0202 & 0.6689 & 2.0202 & 0.6689 \\ 0.669 & 1.005 & 2.0202 & 1.005 & 1.005 \\ 1.005 & 1.005 & 1.005 & 1.005 & 1.005 \end{bmatrix},$$

and $q_\beta = 0.99$. The mean and variance of the effective service time are plotted in Figure 2. When $\beta_{ij}(s) = \beta_{\mathrm{IFR}:ij}(s)$, the changes of the mean and the variance are consistent with those of λ_j. When $\beta_{ij}(s) = \beta_{\mathrm{DFR}:ij}(s)$, the mean and the variance are extremely small compared with those for the case of $\beta_{ij}(s) = \beta_{\mathrm{IFR}:ij}(s), \beta_{\mathrm{CFR}:ij}(s)$.

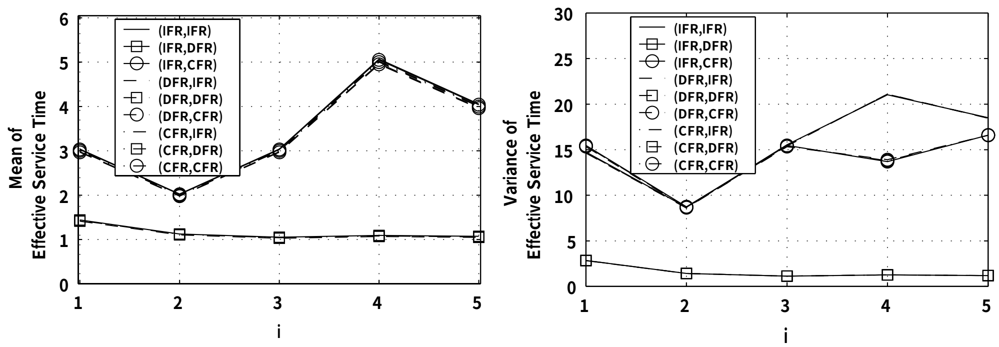

Fig. 2. Mean and variance of effective service time

References

Federgruen, A. and Green, L. (1986): Queueing Systems with Service Interruptions. *Operations Research*, *34*, 752–768.

Gaver, D.P.Jr. (1962): A Waiting Line with Interruption Service, Including Priorities. *Journal of the Royal Statistical Society. Series B*, *24*, 73–90.

Jaiswal, N.K. (1961): Preemptive Resume Queue. *Operations Research*, *9*, 732–770.

Jaiswal, N.K. (1968): *Priority Queues*. Academic Press, New York.

Keilson, J. (1962): Queues Subject to Service Interruption. *The Annals of Mathematical Statistics*, *33*, 1314–1322.

Keilson, J. (1979): *Markov Chain Models - Rarity and Exponentiality*. Springer, New York.

Keilson, J. and Sumita, U. (1983): Evaluation of the Total Time in System in a Preempt/Resume Priority Queue via Modified Lindley Process. *Advances in Applied Probability*, *15*, 840–856.

Sumita, U., Kaio, N. and Goes, P.B. (1989): Analysis of Effective Service Time with Age Dependent Interruptions and Its Application to Optimal Rollback Policy for Database Management. *Queueing Systems*, *4*, 193–212.