Impact of Service Interruptions and the Variability of Service Time in Queueing Systems: Numerical Investigations

Yang Woo Shin[1] and Dug Hee Moon[2]
[1]*Department of Statistics, Changwon National University, Changwon, Gyeongnam 51140, Korea*
[2]*School of Industrial Engineering and Naval Architecture, Changwon National University, Changwon, Gyeongnam 51140, Korea*

Keywords: Service Interruptions, Variance Rate, Departure Process, Finite Buffers.

Abstract: In this paper, we consider the queueing systems with finite buffer and service interruptions. The effects of service interruptions and the variability of service time to measure of departure process such as the asymptotic mean and variance of the number of departures are investigated numerically. We find numerically so called *interruption paradox* or *failure paradox* that the departure rate of the system with service interruptions under preemptive-repeat-different policy can be greater than that of the system with reliable server and it increases as the interruption rate increases for the case of large variability of service time. The results give an insight for the effects of the system and may be helpful to design and control the more complex systems.

1 INTRODUCTION

Queueing networks that consist of nodes with finite buffer and service interruptions have been widely used for modeling and analysis of the system arising from many practical situations such as computer systems, telecommunications and manufacturing systems. The network topology and the characteristics of each node such as buffer size, service time and the interactions between interruptions and service affect not only the performances of each node but also the the performance of whole network.

The mean number of customers, blocking probability of arriving customers and the idle probability of the server are important performance measures of each node. Besides on the internal measures of each node, the measures related with departure process of each node are also important to understand and analyze the whole network performance. In particular, the amount of production in the manufacturing system is presented by the number of departures from a terminal node of a network. The long run average of departures, called departure rate or throughput is an important measure of performance in manufacturing system. The first order measures can be used to get information about the capabilities of a system in the long run. However, there may be tremendous variability of the departures from a time period to period even in a simple queueing network (Gershwin, Section 3.2, 1994). Thus the second order measures such as the variance of the number of departures in a given time period, called variance rate are also very useful to design and control the systems in a more effective way. For a review of recent studies on the variance of the departures for production systems, refer to the paper Tan (2013) and Lagershausen and Tan (2015). Recently, Shin and Moon (2016,2017) present an algorithmic method for asymptotic variance rate of departure process of the system with two-node-one buffer system using the Markovian arrival process.

Interruptions in queueing systems are the elements that prevent the continuous service of customers. Queueing models with service interruptions have been used to model the situations where a service facility is shared by multiple queues, or where the facility is subject to failure. Such interruptions may be caused by breakdowns of the servers, arrival of customers of a higher-priority class or scheduled off-periods by extra jobs. Queueing models with service interruptions and their connection with priority models or machine breakdowns have been studied extensively in the literature, e.g. see White and Christie (1958), Gaver (1962), Nicola (1986), Fiems etc. (2008), Sahba et al. (2015) and refer to the survey paper Krishnamoorthy et al. (2014) for more details. The $M/G/1$ queue with a single type of Poisson interruptions was dealt with extensively by Gaver (1962)

Shin, Y. and Moon, D.
Impact of Service Interruptions and the Variability of Service Time in Queueing Systems: Numerical Investigations.
DOI: 10.5220/0006633202790286
In *Proceedings of the 7th International Conference on Operations Research and Enterprise Systems (ICORES 2018)*, pages 279-286
ISBN: 978-989-758-285-1

for a variety of service-interruption interactions. The analysis was based on the definition of the completion time. He derived the Laplace Stieltjes transform (LST) of the completion time that is the time interval between the instant at which the customer's service begins and the instants at which the service of the next customer (if any exists) may begin and used the method of imbedded Markov chain to obtain the generating function of the distribution of the number of customers in the system. Nicola (1986) derives the LST of completion time for the case with the simultaneous presence of different types of interruptions. The literature cited above deal with the infinite buffer queue and focus on analyzing the stationary distribution of the number of customers in the system, waiting time distribution and related performance measures such as the mean number of customers in the system and blocking probability.

However, the articles reviewed above do not investigate the effects of interactions between interruptions and variability of service time to the system performances. In this paper, we consider the queueing systems with finite buffer and service interruptions and investigate numerically the effects of service interruptions and the variability of service time to measure of departure process such as the asymptotic mean and variance of the number of departures. Numerical results give an insight for the effects of the system and play an important role to prepare the analysis of the extended system of that considered in this present.

This paper is organized as follows. In Section 2, types of interruptions and preliminary results for completion time given by Gaver (1962) are presented. The effects of interruptions and variability of service time to the departure rate and variance rate in the saturated system and $M/PH/1/K$ queue are investigated numerically in In Sections 3 and 4. Concluding remarks are given in Section 5.

2 ASSUMPTIONS AND PRELIMINARY RESULTS

Consider the single server system with service interruptions. In this section, some assumptions and preliminary results to be used later are described.

Service time. Service times of successive customers are independently and identically distributed with arbitrary distribution. Denote the generic random variable of service time by B and $B(x) = P(B \le x)$ and $B^*(s) = E[e^{-sB}]$, $s \ge 0$. Let $E[B^k] = b_k$, $k = 1,2$ and denote the squared coefficient of variation (SCV) of B by $c_b^2 = \mathrm{Var}[B]/b_1^2$.

Interruption. Interruptions appear according to a Poisson process with rate ν and each interruption requires random time to clear the effects of this particular interruption to the server. Successive durations are independent random variables, identically distributed with arbitrary distribution function and denote the generic random variable of the duration of interruption by R. Let $R(x) = P(R \le x)$ and $R^*(s) = E[e^{-sR}]$, $s \ge 0$ and $E[R^k] = r_k$, $r = 1,2$. We assume that the interruption process is independent of the arrival process of customers and the number of customers waiting in line, and the elapsed time since the initial instant.

The interruption occurs only when the server is actually working and it does not occurs during the period while the server is idle or it is in state of interrupted (durations of interruption). This type of interruption is called active interruption (AI) or operation dependent interruption (ODI). The AI can be classified into two categories, say postponable interruptions (PI) and preemptive interruption (PR). When a PI appears during a service time, it does not take effect until the end of the service time. All of the interruptions accumulated during that service time must then be cleared before service of next customer maybegin. Under the PR policy, customer's service is preempted immediately upon the arrival of interruption. In this presentation, we consider only the PR.

Completion Time. A completion time is the time period between the instant at which the customer's service begins and the instants at which the service of the next customer (if any exists) may begin. This period is the sum of the customer's service time and the durations of the interruptions occurring in that time. Let C be the completion time, and denote by $C(x)$ and $C^*(s)$ the distribution function of C and its LST, respectively.

The completion time may depend on the ways of occurrence and clearance of interruptions. Gaver (1962) proposed various types of interruptions and derive the LST's, the first and second moments of completion time in each case. Here, some of the results are summarized in the following for later use.

Let

$$\mathcal{E} = \frac{1/\nu}{1/\nu + \mathbb{E}[R]} = \frac{1}{1+\nu r_1}.$$

The quantity $\mathcal{E}$ is sometimes called an efficiency of the server in a manufacturing system, e.g. see Gershawin (1994).

(i) Preemptive-resume (PRS) Interruptions. In a PRS policy, when an interruption is cleared, service is continued from the point at which it was interrupted. The LST and the mean and variance of completion

time are given by

$$C^*(s) = B^*(s+\nu-\nu R^*(s)),$$
$$E[C] = \frac{b_1}{\mathcal{E}},$$
$$\mathrm{Var}[C] = \frac{\mathrm{Var}[B]}{\mathcal{E}^2} + \nu b_1 r_2.$$

(ii) Preemptive-repeat-different (PRT-D) Interruptions. In this case, when an interruption is cleared, service begins again from scratch, but each time another interruption is cleared a new independent (potential) service time whose distribution function is $B(x)$ begins. Service is completed when, for the first time, such a service time elapses without interruption.

$$\begin{aligned} C^*(s) &= \frac{B^*(s+\nu)}{1-R^*(s)\frac{\nu}{s+\nu}(1-B^*(s+\nu))}, \\ E[C] &= \frac{1}{\nu\mathcal{E}}\frac{1-B^*(\nu)}{B^*(\nu)}, \\ \mathrm{Var}[C] &= (E[C])^2 + \left(\nu\mathcal{E}r_2 + \frac{2}{\nu}\right)E[C] \\ &\quad - \frac{2}{\nu\mathcal{E}}\frac{E[Be^{-\nu B}]}{(B^*(\nu))^2}. \end{aligned}$$

(iii) Preemptive-repeat-identical (PRT-I) Interruptions. In this case, when the interruption is cleared, a service period of the same duration as the one interrupted begins again from scratch. Service is completed (completion time terminates) when, for the first time, a (repeated) service period elapses without interruption. The LST and the mean and variance of completion time are given by

$$\begin{aligned} C^*(s) &= \int_0^\infty \frac{e^{-(s+\nu)x}}{1-R^*(s)\frac{\nu}{s+\nu}(1-e^{-(s+\nu)x})}dB(x), \\ E[C] &= \frac{1}{\nu\mathcal{E}}(E[e^{\nu B}]-1), \\ \mathrm{Var}[C] &= \frac{1}{\nu^2\mathcal{E}^2}\left(\mathrm{Var}\left[e^{\nu B}-1\right] + E\left[(e^{\nu B}-1)^2\right]\right) \\ &\quad + \left(\nu\mathcal{E}r_2 + \frac{2}{\nu}\right)E[C] - \frac{2}{\nu\mathcal{E}}E[Be^{\nu B}], \end{aligned}$$

where the expectations may not exist.

3 ASYMPTOTIC RATE FOR THE NUMBER OF DEPARTURES IN A SATURATED SYSTEM

Consider a single server system that is saturated and never blocked. That is, the server always works unless it is down state and the customer leaves the system immediately after the service without blocking.

Let $N(t)$ be the number of service completions during an interval $(0,t]$. Then $\boldsymbol{N} = \{N(t), t \geq 0\}$ is a renewal process whose inter-renewal distribution is the same as the completion time C. It follows from the well known results of the renewal theory (e.g. see Cox (page 58, 1962)) that the long run average number of departures and the variance rate of $\boldsymbol{N}$ are given by

$$\begin{aligned} \mu &= \lim_{t\to\infty}\frac{\mathbb{E}[N(t)]}{t} = \frac{1}{E[C]}, \\ V &= \lim_{t\to\infty}\frac{\mathrm{Var}[N(t)]}{t} = \frac{\mathrm{Var}[C]}{(E[C])^3}. \end{aligned}$$

Indeed, the distribution of $N(t)$ is asymptotically normal with mean μt and variance Vt, $i = 1,2$.

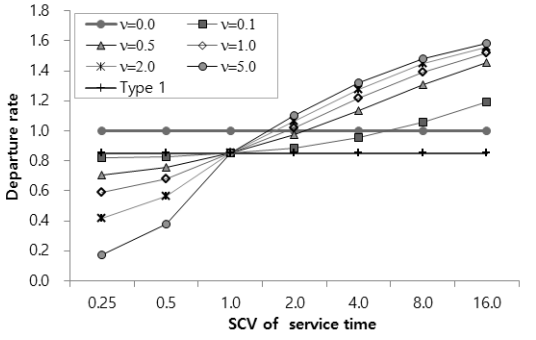
Now, we investigate the effects of the interactions between interruptions and service time, interruption rate ν and the variability of service time to the departure rate μ and variance rate V. The PH-distributiion (PH) and lognormal distribution (LN) of service times and exponential distribution of duration R of an interruption with rate η are considered. The mean service time and the efficiency of the server is fixed by $b_1 = 1.0$ and $\mathcal{E} = 0.85$ and the repair rate is determined by $\eta = \frac{\nu\mathcal{E}}{1-\mathcal{E}} = \frac{17}{3}\nu$ for interruption rate $\nu > 0$. For PH-distribution, we use the Erlang distribution of order k (E_k) for $C_b^2 = \frac{1}{k} < 1$, exponential distribution (Exp) for $C_b^2 = 1$ and hyperexponential distribution of order 2 with balanced mean for $C_b^2 > 1$ denoted by $H_2(p_1,\lambda_1,\lambda_2)$ whose probability density function is

$$f(t) = p_1\lambda_1 e^{-\lambda_1 t} + p_2\lambda_2 e^{-\lambda_2 t},\ t \geq 0,$$

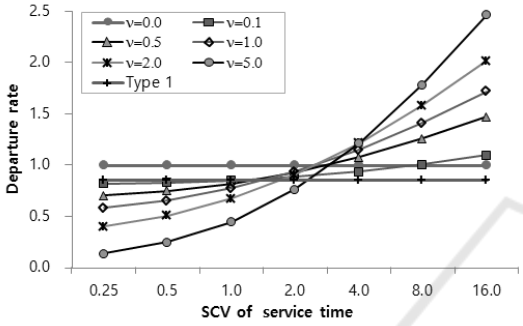
with $\lambda_1 = 2p_1\mu$, $\lambda_2 = 2p_2\mu$ and

$$p_1 = \frac{1}{2}\left(1 + \sqrt{\frac{c_b^2-1}{c_b^2+1}}\right),\ p_2 = 1-p_1.$$

Preemptive-resume versus preemptive-repeat-different. The departure rate μ as a function of c_b^2 for the various interruption rate ν are depicted in Figure 1. The figures show that the departure rates of the systems with reliable server $(\nu = 0.0)$ and the server with PRS policy (denoted by Type 1) do not depend on the SCV of service time. However, the departure rate μ_2 of the system with PRT-D policy increases as c_b^2 increases for each ν and it can be greater than the service rate $\mu = 1.0$ of reliable server $(\nu = 0.0)$. Furthermore, the departure rate increases as the interruption rate increases for large c_b^2. These seems surprising and we shall phrase it the *interruption paradox* or *failure paradox*. We have found that these results holds for Weibul distribution and gamma distribution of service time although the results are not presented in this paper.



(a) PH-service time



(c) lognormal-service time

Figure 1: Departure rate as a function of SCV C_s^2.

The reason of failure paradox can be explained as follows. The parameters of hyperexponential distribution $H_2(p_1,\lambda_1,\lambda_2)$ with mean 1.0 are listed in Table 1.

Table 1: Parameters for $H_2(p_1,\lambda_1,\lambda_2)$

C_s^2	p_1	p_2	λ_1	λ_2
1.0	0.5000	0.5000	1.0000	1.0000
2.0	0.7887	0.2113	1.5774	0.4227
4.0	0.8873	0.1127	1.7746	0.2254
8.0	0.9410	0.0590	1.8819	0.1181
16.0	0.9697	0.0303	1.9393	0.0607

It can be seen from Table 1 that p_1 approach to 1.0, and λ_1 increases and is greater than the service rate 1.0 and λ_2 decreases as SCV increases. When a service time of a customer is assigned to long service time corresponding to the rate λ_2, the service time can be interrupted by a failure and the server starts a new service with short service time corresponding to λ_1 with high probability p_1. Thus a failure can make the service time be shorter than that of the system with reliable server.

The ratios $\frac{V_1}{V_0}$ between the variance rate V_0 of the noninterrupted system and V_1 of the system with PRS policy and the variance rate V_2 of the system with PRT-D policy for the system with PH-service time are depicted in Figure 2. The ratios V_1/V_0 tends to $\mathcal{E} = 0.85$ as C_b^2 increases which can be expected from the formula V_1/V_0. The variance rate V_2 of the system with PRT-D policy increases as SCV c_b^2 of service time increases, but V_2 is less than V of the reliable system for $C_b^2 > 1$. The ratio $\frac{V_2}{V_1}$ are depicted in Figure 3. It can be seen from the figures 3 that the variance rate V_2 depends severely on the distribution of service time. We have seen from extensive numerical experiments that the variance rate V_2 depends severely on the distribution of service time. The variance ratio $\frac{V_2}{V_1}$ decreases and becomes less than 1.0 for the system with H_2 and Weibul distribution of service time, however, it increases and becomes greater than 1.0 for the system with gamma distribution of service time as CSV of the service time increases.

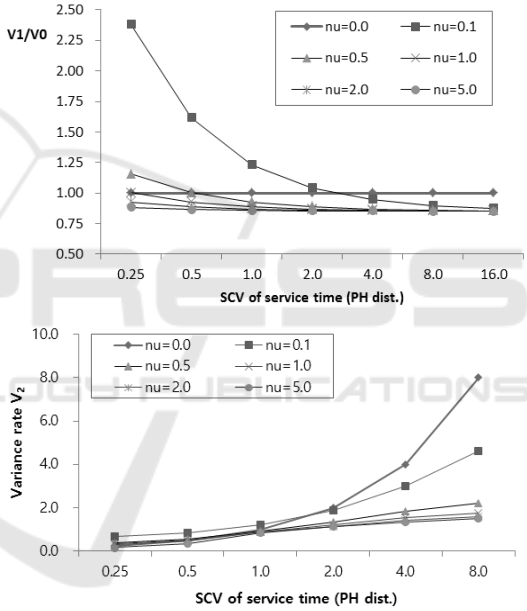


Figure 2: The ratios $\frac{V_1}{V_0}$ for the system with PH service time.

Preemptive-repeat-identical versus preemptive-repeat-different. In case of preemptive-repeat-identical policy, the expectation $E[e^{\nu B}]$ for $E[C]$ exists if $1 - B(x) = o(e^{-cx})$ with $t \to \infty$ for $0 < \nu < c$. For example, if the distribution of service time is gamma distribution, then the expectations exist for $\nu < \frac{1}{b_1 c_b^2}$. If the distribution of B is Weibul, the the expectation exists only for $c_b^2 < 1$. If B has lognormal distribution, then the expectation does not exist. Let $E[C_I]$ and $E[C_D]$ denote the expectations of C for the cases of PRT-I and PRT-D, respectively. It can be easily seen from the formulae $E[C_I]$ and $E[C_D]$ that $E[C_I] > E[C_D]$ is equivalent to

$$B^*(\nu)B^*(-\nu) > 1. \quad (1)$$

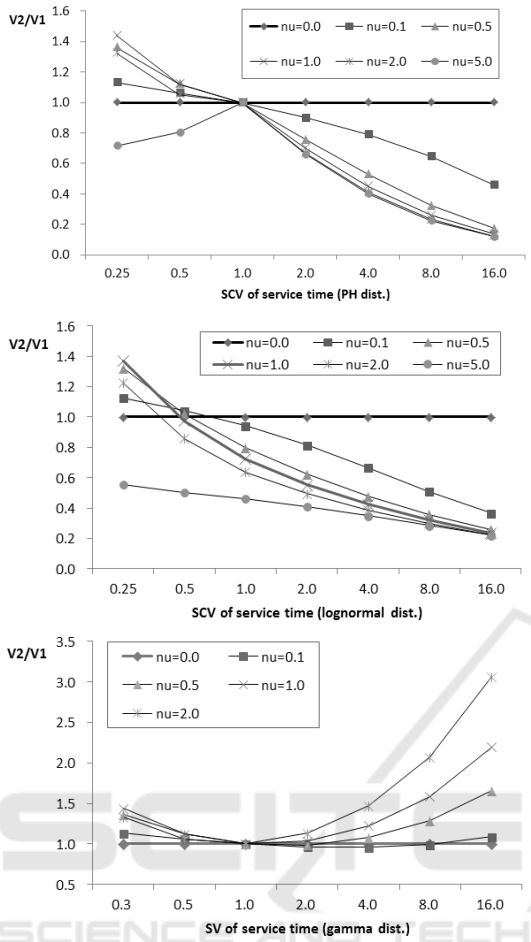


Figure 3: The ratio $\frac{V_2}{V_1}$.

It can be seen from the formula $B^*(s) = (1 - b_1 c_b^2 s)^{-1/c_b^2}$ of LST that gamma distribution and the condition of existence of $E[C_I] < \infty$ that gamma distribution satisfies (1). We showed numerically that the Weibul distribution with $c_b^2 < 1$ satisfies (1) for any $\nu > 0$, but we omit the details here.

4 $M/PH/1/K$ QUEUE WITH SERVICE INTERRUPTIONS

4.1 Model

We consider a $M/PH/1/K$ queue with service interruption and a buffer of finite capacity K in which customers arrive according to a Poisson process with rate λ. Interruptions occurs only while the server is working. The inter occurrence time of interruption is assumed to be exponential distribution with rate ν. The service time and duration of interruption are assumed to be of phase type distributions $PH(\boldsymbol{\alpha}, S)$ and $PH(\boldsymbol{\gamma}, G)$, respectively. Let $\mathbf{s}^0 = -S\mathbf{e}$ and $\mathbf{g}^0 = -G\mathbf{e}$. Let w and r be the number of phases of the distributions of service time and duration of interruption, respectively.

Let $X(t)$ be the number of customers in the system at time t. The state space of $X(t)$ is $\{0, 1, \cdots, K\}$. By $J^w(t)$ and $J^r(t))$ denote the phases of $PH(\boldsymbol{\alpha}, S)$ and $PH(\boldsymbol{\gamma}, G)$, respectively at time t. The state $M(t)$ of the server M at time t is

$$M(t) = \begin{cases} J^w(t), & \text{the server is up at time } t \\ (J^w(t), J^r(t)), & \text{the server is down at time } t \end{cases}$$

Let $[D_0]_{ij}$ ($[D_0^*]_{ij}$) be the rate that a transition of $M(t)$ occurs from i to j and no service is completed given $X(t) \geq 1$ ($X(t) = 0$, respectively) and $[D_1]_{ij}$ ($[D_1^*]_{ij}$) be the rate that a transition of $M(t)$ occurs from i to j and a service is completed given $X(t) \geq 2$ ($X(t) = 1$, respectively). Let $\mathcal{M}$ ($\mathcal{M}^*$) be the state space of $M(t)$ for $X(t) \geq 1$ ($X(t) = 0$, respectively) and m and m^* be the number of elements of $\mathcal{M}$ and $\mathcal{M}^*$, respectively. Let P_1 be the $m^* \times m$ matrix whose (i, j)-component $[P_1]_{ij}$ is the probability that the phase of M is j immediately after an arrival occurs given that $X(t) = 0$ and $M(t) = i$. The matrices D_0 and D_1 depends on the service initiation policies after clearance of interruption.

We consider the following three policies of initiation of service when an interruption is cleared.

S_1 policy. When an interruption is cleared, service resume at the last phase in which a failure occurs. In this case, $m = w(r+1)$ and $m^* = 1$ and

$$D_0 = \begin{pmatrix} -\nu\mathbf{I}_w + S & \nu\mathbf{I}_w \otimes \boldsymbol{\gamma} \\ \mathbf{I}_w \otimes \mathbf{g}^0 & \mathbf{I}_w \otimes G \end{pmatrix},$$
$$D_1 = \begin{pmatrix} \mathbf{s}^0\boldsymbol{\alpha} & \mathbf{O} \\ \mathbf{O} & \mathbf{O}_{wr \times wr} \end{pmatrix}, D_1^* = \begin{pmatrix} \mathbf{s}^0 \\ \mathbf{O}_{wr \times 1} \end{pmatrix}$$

and $D_0^* = 0$, $P_1 = (\boldsymbol{\alpha}\ O_{1 \times wr})$, where $\mathbf{O}_{k \times n}$ is the zero matrix of size $k \times n$ and $\mathbf{I}_n$ is the identity matrix of size n.

S_2 policy. When an interruption is cleared, new service starts anew according to a PH-distribution $PH(\boldsymbol{\alpha}, S)$. In this case, $m = w + r$ and $m_1^* = 1$ and

$$D_0 = \begin{pmatrix} -\nu\mathbf{I}_w + S & \nu\mathbf{e}_w\boldsymbol{\gamma} \\ \mathbf{g}^0\boldsymbol{\alpha} & G \end{pmatrix},$$
$$D_1 = \begin{pmatrix} \mathbf{s}^0\boldsymbol{\alpha} & \mathbf{O} \\ \mathbf{O} & \mathbf{O}_{r \times r} \end{pmatrix}, D_1^* = \begin{pmatrix} \mathbf{s}^0 \\ \mathbf{O}_{r \times 1} \end{pmatrix},$$

and $D_0^* = 0$, $P_1 = (\boldsymbol{\alpha}\ O_{1 \times r})$.

S_3 policy. When an interruption occurs, the customer being served is scrapped, and the server begins new service of length whose distribution is of

$PH(\boldsymbol{\alpha}, S)$ after clearance of interruption. In this case, $m = w + r$ and $m_1^* = r + 1$ and

$$D_0 = \begin{pmatrix} -\nu \mathbf{I}_w + S & \mathbf{O}_{w\times r} \\ \mathbf{g}^0\boldsymbol{\alpha} & G \end{pmatrix}, D_0^* = \begin{pmatrix} 0 & \mathbf{O} \\ \mathbf{g}^0 & G \end{pmatrix},$$
$$D_1 = \begin{pmatrix} \mathbf{s}^0\boldsymbol{\alpha} & \nu \mathbf{e}_w\boldsymbol{\gamma} \\ \mathbf{O} & \mathbf{O}_{r\times r} \end{pmatrix}, \quad D_1^* = \begin{pmatrix} \mathbf{s}^0 & \nu \mathbf{e}_w\boldsymbol{\gamma} \\ \mathbf{O} & \mathbf{O}_{r\times r} \end{pmatrix},$$
$$P_1 = \begin{pmatrix} \boldsymbol{\alpha} & O \\ O & \mathbf{I}_r \end{pmatrix}.$$

Let $Z(t) = (X(t), M(t))$. The stochastic process $\boldsymbol{Z} = \{Z(t), t \geq 0\}$ forms a Markov chain with generator of the form

$$Q = \begin{pmatrix} B_0 & A_0 & & & \\ C_1 & B_1 & A_1 & & \\ & \ddots & \ddots & \ddots & \\ & & C_{K-1} & B_{K-1} & A_{K-1} \\ & & & C_K & B_K \end{pmatrix},$$

Then the block matrix components are as follows:

$$B_0 = -\lambda \mathbf{I}_{m^*} + D_0^*,\ B_K = D_0,\ A_0 = \lambda P_1,\ C_1 = D_1^*,$$
$$B_n = -\lambda \mathbf{I}_m + D_0,\ A_n = \lambda \mathbf{I}_m,\ 1 \leq n \leq K-1,$$
$$C_n = D_1,\ 2 \leq n \leq K.$$

4.2 Departure Process

Let $N(t)$ be the number of service completions in the system. Then $\boldsymbol{N} = \{N(t), t \geq 0\}$ is a Markovian arrival process (MAP) with representation MAP$(\mathcal{D}_0, \mathcal{D}_1)$, where the matrices $\mathcal{D}_0$ and $\mathcal{D}_1$ are as follows:

(1) S_1 and S_2 policies.

$$\mathcal{D}_0 = \begin{pmatrix} B_0 & A_0 & & & \\ & B_1 & A_1 & & \\ & & \ddots & \ddots & \\ & & & B_{K-1} & A_{K-1} \\ & & & & B_K \end{pmatrix},$$
$$\mathcal{D}_1 = \begin{pmatrix} O & & & & \\ C_1 & O & & & \\ & C_2 & O & & \\ & & \ddots & \ddots & \\ & & & C_K & O \end{pmatrix}$$

and O is the zero matrix of appropriate size.

(2) S_3 policy. For the system with scrapping,

$$\mathcal{D}_0 = \begin{pmatrix} B_0 & A_0 & & & & \\ C_{1,0} & B_1 & A_1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & C_{K-1,0} & B_{K-1} & A_{K-1} \\ & & & C_{K,0} & B_K \end{pmatrix},$$
$$\mathcal{D}_1 = \begin{pmatrix} O & & & & \\ C_{1,1} & O & & & \\ & C_{2,1} & O & & \\ & & \ddots & \ddots & \\ & & & C_{K,1} & O \end{pmatrix}$$

where

$$C_{1,j} = D_{1,j}^*,\ C_{n,j} = D_{1,j},\ 2 \leq n \leq K,\ j = 0, 1,$$

and

$$D_{1,0} = \begin{pmatrix} \mathbf{O}_{w\times w} & \nu \mathbf{e}_w\boldsymbol{\gamma} \\ \mathbf{O} & \mathbf{O}_{r\times r} \end{pmatrix},$$
$$D_{1,1} = \begin{pmatrix} \mathbf{s}^0\boldsymbol{\alpha} & O \\ \mathbf{O} & \mathbf{O}_{r\times r} \end{pmatrix}$$
$$D_{1,0}^* = \begin{pmatrix} \mathbf{O}_{w\times 1} & \nu \mathbf{e}_w\boldsymbol{\gamma} \\ \mathbf{O} & \mathbf{O}_{r\times r} \end{pmatrix},$$
$$D_{1,1}^* = \begin{pmatrix} \mathbf{s}^0 & O \\ \mathbf{O} & \mathbf{O}_{r\times r} \end{pmatrix}.$$

A Markovian arrival process (MAP) with representation MAP$(\mathcal{D}_0, \mathcal{D}_1)$ is a counting process whose inter event time distribution is of the form

$$F(t) = \int_0^t \exp(\mathcal{D}_0 u)\, du \mathcal{D}_1,\ t \geq 0.$$

For details of MAP, see Lucantoni (1991) or Artalejo et al. (2010).

Let $\boldsymbol{\pi} = (\pi(x), x \in \mathcal{S})$ be the stationary distribution of $\mathcal{D} = \mathcal{D}_0 + \mathcal{D}_1$ and

$$\mathbf{c} = \boldsymbol{\pi}\mathcal{D}_1(\mathbf{e}\boldsymbol{\pi} - \mathcal{D})^{-1},\ \mathbf{d} = (\mathbf{e}\boldsymbol{\pi} - \mathcal{D})^{-1}\mathcal{D}_1\mathbf{e}.$$

It is known that (e.g. Neuts,Theorem 5.4.1, 5.4.2, 1989; Artalejo et al., 2010) in stationary state, the mean and the variance of $N(t)$ are given as follows:

$$\mathbb{E}[N(t)] = \mu t,$$
$$\mathrm{Var}[N(t)] = Vt + 2(\mu^2 - \mathbf{cd}) + o(1).$$

where $\mu = \boldsymbol{\pi} D_1 \mathbf{e}$ and

$$V = \mu - 2\mu^2 + 2\mathbf{c}D_1\mathbf{e}.$$

One can refer to Shin and Moon (2016, 2017) for an algorithm for computing μ and V, the mean and variance of inter departure time, and the covariance between inter departure times.

4.3 Numerical Results

In this subsection, we investigate the effects of the interactions between the interruptions and service time and the variability of service time to the departure rate and variance rate of departure process in $M/PH/1/5$ queue with arrival rate $\lambda = 1.0$, mean service time $b_1 = 1.0$ and $\mathcal{E}_1 = 0.85$. We use the Erlang distribution of order k (E_k) for $c_b^2 = \frac{1}{k} < 1$, exponential distribution (Exp) for $c_b^2 = 1$ and hyperexponential distribution of order 2 with balanced mean for $c_b^2 > 1$.

The comparisons of departure rates for the service policies S_1, S_2, S_3 and interruption free system $(\nu = 0.0)$ are presented in Figure 4. It can be seen from the figure that the departure rate for the system with S_2 policy increases and can be greater than the isolated efficiency while the departure rate decreases in the system with S_1 policy and reliable system as SCV c_b^2 of service time increases. It can be also seen from the figure that the departure rate of the system with scrap can be greater than that of reliable system as SCV of service time increases.

Denote the variance rates for S_i by V_i, $i = 1,2,3$ and let V_0 be the variance rate for interruption free system $(\nu = 0)$. Here, the variance rates V_1 and V_2 are depicted in Figure 5. We can see from the figure that the variance rates increase in both types of service initiation policies as c_b^2 increases. It also can be seen that $V_1 > V_0 > V_2$ for c_b^2 sufficiently greater than 1. Furthermore, V_2 decreases as ν increases for $c_b^2 > 1$. We also can see from the figure that the behaviors V_3 are similar to those of V_2.

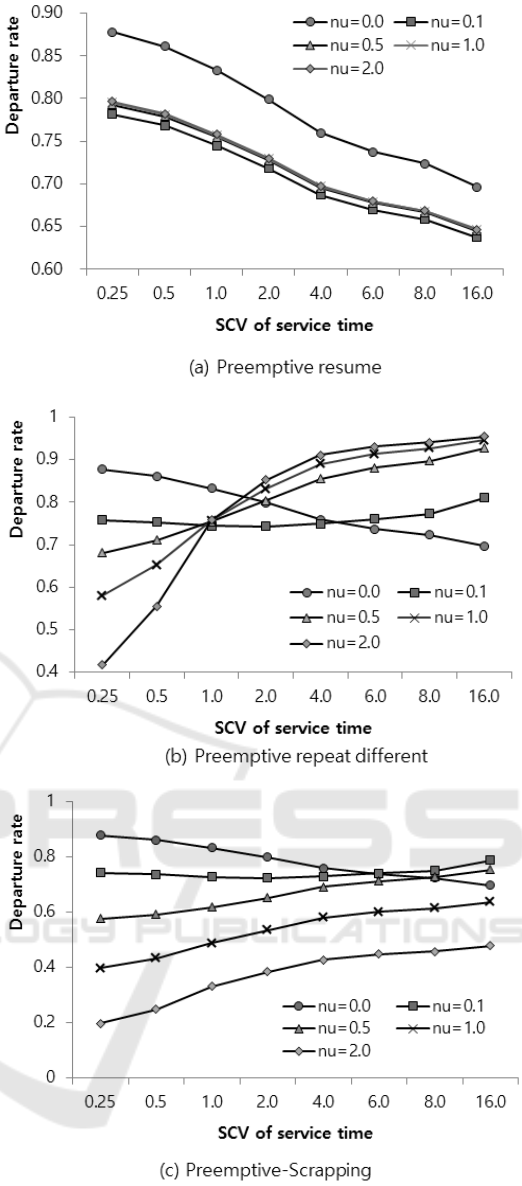


Figure 4: Departure rate as a function of SCV c_b^2.

5 CONCLUSIONS

In this study, the effects of structural parameters such as the variabilities of service and the interactions between interruptions and service time to the departure rates and variance rates have been investigated numerically. We have observed from numerical experiments that the variance rate of departures increases as the SCV of the service time increases. However, the departure rate in the system with interruptions can be greater than that of the interruption free system and it can increase as the interruption rate increase for large SCV of service time. This result is different from the case of the system with reliable servers in which the departure rate decreases as SCV's of service times increase. We have coined this surprising results the *interruption paradox* or *failure paradox*. The effects of interruption rate to the departure rate and variance rate are dependent of the SCV's of service time.

It remains to analyze the more complicated system such as the queueing system with more general arrival and/or service time and queueing networks with interruptions as further research area. Numerical results give some insights for the more complicated systems. So, our experiments may be helpful to design and control the system with interruptions and may play a useful role to prepare the analysis of the extended systems.

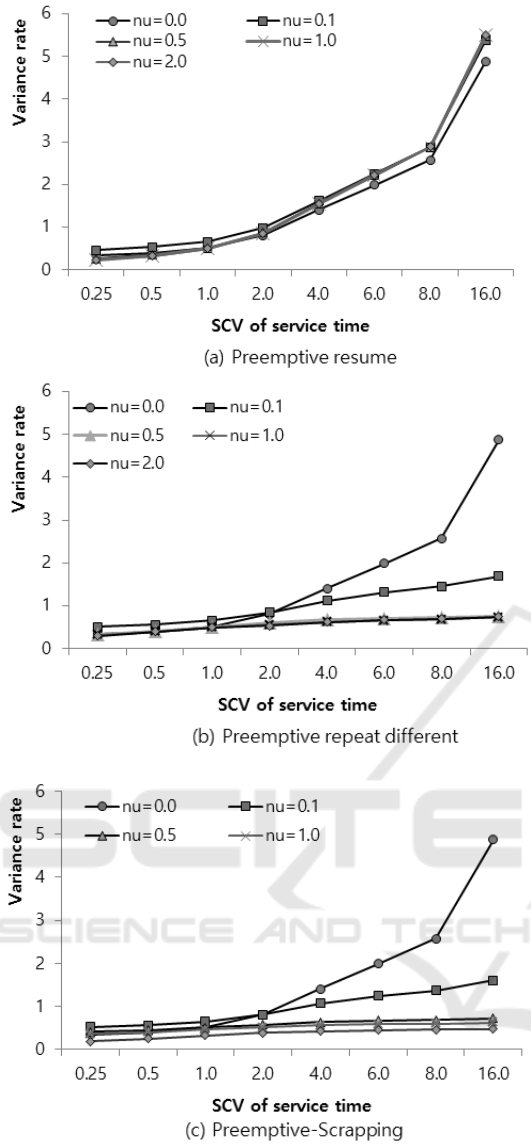


Figure 5: Variance rate as a function of c_b^2.

REFERENCES

Artalejo, J. R., Gómez-Corral, A., He, Q. M. (2010). Markovian arrivals in stochastic modelling: a survey and some new results, SORT 34 (2), 101-144.

Gaver, D. Jr. (1962). A waiting line with interrupted service, including priorities, Journal of the Royal Statistical Society, Series B24, 73-90.

Gershwin, S. B. (1994). *Manufacturing systems engineering*. Prentice-Hall, Englewood Cliffs.

Krishnamoorthy, A., Pramod, P. K., Chakravarthy, S. R. (2014) Queues with interruptions: a survey, TOP 22(1), 290-320.

Lagershausena, S., Tan, B. (2015). On the exact inter-departure and inter-start time distribution of closed queueing networks subject to blocking, IIE Transactions 47, 673-692

Lucantoni, D. M., Meier-Hellstern, K. S., Neuts, M. F. (1990). A single server queue with server vacations and a class of non-renewal arrival processes, Advances in Applied Probability 22, 676-705.

Narayana, S., Neuts, M. F. (1992). The first two moment matrices of the counts for the Markovian arrival process, Stochastic Models 8(3), 459-477.

Neuts, M. F. (1989). *Structured Stochastic Matrices of $M/G/1$ Type and Their Applications*. Marcel Dekker, New York.

Nicola, V. F. (1986). A single server queue with mixed types of interruptions, Acta Informatica 23, 465-486.

Sahba, P., Balcoğlu, B. Banjevic, D. (2015). The impact of disruption characteristics on the performance of a server, Annals of Operations Research, https://doi.org/10.1007/s10479-015-2075-2, pp1-14.

Shin, Y. W., Moon, D. H. (2016). Variability of output in two-node tandem production line. *Proceedings of QTNA 2016*, December 2016, Wellington, New Zealand, pp. 13-15.

Shin, Y. W., Moon, D. H. (2017). Variance of departure process in two-node tandem queue with unreliable server and blcking. *Proceedings of ICORES 2017*, February 2017, Porto, Portugal, pp. 258-264.

Tan, B. (2013). Modeling and analysis of output variability in discrete material flow production systems. In *Handbook of Stochastic Models and Analysis of Manufacturing System Operations*. Tan,B. and Smith, J. M. (eds), Springer, New York, pp. 287-311.

White, H., Christie, L. (1958). Queuing with preemptive priorities or with breakdown, Operation Research 6(1), 79-95.