Journal of the Operations Research
Society of Japan
Vol. 39, No. 4, December 1996

A NOTE ON THE RESPONSE TIME IN M/G/1 QUEUES WITH SERVICE IN RANDOM ORDER AND BERNOULLI FEEDBACK

Hideaki Takagi
The University of Tsukuba

(Received December 1, 1994; Final August 21, 1995)

Abstract We consider M/G/1 queueing systems with random order of service and Bernoulli feedback of output customers. These systems may model the aggregate queues of packets waiting for transmission in a contention-based multiaccess communication channel. We study the customer's response time defined as the time from its arrival to final departure. The mean response time is equivalent to that in a batch arrival system in which the batch size is geometrically distributed. The second moment of the response time is newly obtained explicitly. Numerical comparison shows that the random order of service sometimes yields smaller values of the second moment of the response time than first-come first-served and last-come first-served disciplines in feedback systems. We deal with a system without server vacations as well as one with multiple server vacations.

1. Introduction

Single server queues with feedback of output customers can model many practical systems. Among them is a system in which each customer may be served repeatedly for a certain reason. When the service is unsuccessful, it may be retried over and again until success. For example, the retransmissions of packets occur when they are not acknowledged by the receiver before timeout in data communication systems. The unsuccessful transmission is inherent in contention-based multiaccess protocols such as ALOHA, carrier sense multiple access (CSMA), and their variations that are used in packet radio networks as well as in local area networks. If we assume that all users in a system with a contention-based multiaccess protocol are statistically identical, randomized chances of transmission are given to each user having a packet with equal probability. Thus the set of all packets waiting for transmission forms a queue that is served in random order with a possibility of feedback.

From this motivation, we consider an M/G/1 queueing system with service in random order and Bernoulli feedback. The rate of a Poisson arrival process is denoted by λ. The LST of the DF, the mean, and the ith moment of the generally distributed service times are denoted by $B^*(s)$, b, and $b^{(i)}$ ($i = 2, 3, \ldots$), respectively. The service discipline is service-in-random-order (SIRO), that is, every customer present in the queue at the end of each service can be selected for the next service with equal probability. The customer whose service is completed immediately joins the queue with probability $1 - \nu$, or departs from the system with probability ν, where $0 < \nu \leq 1$ (Bernoulli feedback). The stability condition for this system is given by $\lambda b/\nu < 1$. This system is different from the M/G/1 queue with first-come first-served (FCFS) discipline and Bernoulli feedback, which was analyzed first by Takács [9] and subsequently by others. The latter may model a system in which a segment of the whole required service is given at each service epoch, such as the time-sliced processing in multitasking computers and the segmented transfer of a large file over a multiplexed communication line.

Clearly the number of customers in the system (called the *queue size* hereafter) does not depend on the service discipline such that the selection of a customer for service is not affected by the service times of the waiting customers. Also, the queue size in a system with Bernoulli feedback is equivalent to the number of batches present in a batch arrival system without feedback in which the number of customers included in each arriving batch has geometric distribution starting at one with mean $1/\nu$. We focus on the *response time* of an arbitrary customer, which is defined as the time from its first arrival to the final departure. The probability distribution of the response times depends on the service discipline. However, from Little's formula the mean response time is independent of the service discipline with the above-mentioned property. It is also identical with the mean response time of each customer in a batch arrival system with geometrically distributed batch sizes. The second moment of the response time in the FCFS system is shown in the Takács paper [9] as a laborious work of W. S. Brown. The primary objective of the present paper is to give the second moment of the response time in the SIRO system. (Treatment of SIRO M/G/1 queues without feedback can be found in Cohen [1, sec. III.3.3], Conolly [2, sec. 5.3.5], Cooper [3, sec. 5.12], Fuhrmann [5], Kingman [6], Takács [8], and Takagi [10, sec. 1.3].) It provides a measure of variability and can be used to obtain the bounds in the distribution of the response time. Note that the distribution of the response time in the last-come first served (LCFS) system with Bernoulli feedback can be obtained immediately by modifying the service time distribution in the LCFS system without feedback. We can deal with a system without server vacations as well as a system with server vacations by the same approach. The vacation of the server usually represents a period during which the server is allocated to some other tasks.

In the rest of this paper, we mainly present the analysis for a system without server vacations. The mean response time is known. In Section 2, we first express the Laplace-Stieltjes transform (LST) of the distribution function (DF) for the response time of an arriving customer in terms of the LST of the DF for the response time of an arbitrary customer in the system at the epoch of service start, conditioned on the number of customers present in the system at that time. We then calculate the second moment of the response time of the arriving customer. In Section 3, we show an outline and results of analysis for the similar system with multiple server vacations. In Section 4, we compare the numerical values of the second moments of the response time for the systems with FCFS, LCFS, and SIRO disciplines without vacations. We also display the effects of vacations on the second moment of the response time for SIRO systems.

2. Second Moment of the Response Time in a System without Server Vacations

Let us consider a system without server vacations. The probability generating function (PGF) $\Pi(z)$ for the queue size L in this system at an arbitrary time is identical with that for the number of batches in a batch arrival system without feedback in which the number of customers included in each arriving batch has geometric distribution starting at one with mean $1/\nu$. Therefore, using

$$B_g^*(s) := \sum_{k=1}^{\infty}(1-\nu)^{k-1}\nu[B^*(s)]^k = \frac{\nu B^*(s)}{1-(1-\nu)B^*(s)} \tag{1}$$

as the LST of the DF for the service time in the Pollaczek-Khinchin transform equation for an M/G/1 system without feedback, we get

$$\Pi(z) = \frac{(\nu - \lambda b)(1-z)B^*(\lambda - \lambda z)}{[\nu + (1-\nu)z]B^*(\lambda - \lambda z) - z} \tag{2}$$

From Little's formula, the mean for the response time T of an arbitrary customer is given by

$$E[T] = \frac{E[L]}{\lambda} = \frac{2(1-\lambda b)b + \lambda b^{(2)}}{2(\nu - \lambda b)} \tag{3}$$

We note that the PGF $\Pi(z)$ in (2) also holds for the queue size immediately after a customer has left the system (a departure epoch). These results for the queue size distribution and the mean response time do not depend on the service discipline that does not use the service time for selecting the next customer to serve, including FCFS, LCFS, and SIRO disciplines.

In order to study the distribution of the response time, we follow an approach by Takács [8], who treated the LST of the DF for the *waiting time* of a customer, which is defined as the time from its arrival to the service start, for an SIRO M/G/1 system without feedback. Instead, we consider the response time that is more appropriate for a system with feedback. Let $T^*(s)$ be the LST of the DF for the response time T of an arriving customer. Since the arriving customer finds the system empty (idle) with probability $1 - \lambda b/\nu$, or finds it nonempty (busy) with probability $\lambda b/\nu$, we have

$$T^*(s) = \left(1 - \frac{\lambda b}{\nu}\right) E[e^{-sT}|\text{idle}] + \frac{\lambda b}{\nu} E[e^{-sT}|\text{busy}] \tag{4}$$

We will express $E[e^{-sT}|\text{idle}]$ and $E[e^{-sT}|\text{busy}]$ in terms of the LST $T_k^*(s)$ of the DF for the time T_k from the instant of service start when there are $k+1$ customers are present in the system to the instant when an arbitrary customer C among them leaves the system. Let us first derive the equations for $\{T_k^*(s); k = 0, 1, 2, \ldots\}$. To do so, we use the joint LST of the DF for the service time and the probability that j customers arrive during the service time, defined by

$$a_j^*(s) := \int_0^\infty e^{-sx} \frac{(\lambda x)^j}{j!} e^{-\lambda x} dB(x) \qquad j = 0, 1, 2, \ldots \tag{5}$$

where $B(x)$ the DF for the service time. Note that

$$\sum_{j=0}^{\infty} a_j^*(s) z^j = B^*(s + \lambda - \lambda z) \tag{6}$$

The set of equations for $\{T_k^*(s); k = 0, 1, 2, \ldots\}$ can be derived from the following recursive arguments. When there are k customers in the system other than C at the time of service start, C is selected for the service with probability $1/(k+1)$. Its response time will then be exactly the service time in the case of no feedback which occurs with probability ν. In the case of feedback which occurs with probability $1-\nu$, the LST of the DF for its response time will be $a_j^*(s)T_{k+j}^*(s)$ if j customers arrive during its service time. When there are k customers in the system other than C at the time of service start, C is not selected for service with probability $k/(k+1)$. If j customers arrive during the next service time, the LST of the DF for C's response time is $a_j^*(s)T_{k+j-1}^*(s)$ if the served customer is not fed back, or it is $a_j^*(s)T_{k+j}^*(s)$ otherwise. Thus we get the relationship:

$$\begin{aligned} T_k^*(s) &= \frac{1}{k+1}\left[\nu B^*(s) + (1-\nu)\sum_{j=0}^{\infty} a_j^*(s)T_{k+j}^*(s)\right] \\ &+\frac{k}{k+1}\sum_{j=0}^{\infty} a_j^*(s)\left[\nu T_{k+j-1}^*(s) + (1-\nu)T_{k+j}^*(s)\right] \end{aligned} \tag{7}$$

which reduces to

$$\begin{aligned} T_k^*(s) &= \nu\left[\frac{1}{k+1}B^*(s) + \frac{k}{k+1}\sum_{j=0}^{\infty} a_j^*(s)T_{k+j-1}^*(s)\right] \\ &+(1-\nu)\sum_{j=0}^{\infty} a_j^*(s)T_{k+j}^*(s) \qquad k=0,1,2,\ldots \end{aligned} \tag{8}$$

From (6) and (8), we can get $E[T_k]$ and $E[(T_k)^2]$. To do so, we derive the recursive relations for the sets $\{E[T_j]; j=0,1,2,\ldots\}$ and $\{E[(T_j)^2]; j=0,1,2,\ldots\}$ by evaluating the derivatives of (8). These relations can be solved by using the sums of the series in the derivatives of $a_j^*(s)$ obtained from (6). See Appendix for the detailed derivation. Thus we get

$$E[T_k] = \frac{kb + (2-\lambda b)b}{2\nu - \lambda b} \tag{9}$$

$$\begin{aligned} E[(T_k)^2] &= \frac{2k(k-1)b^2}{(2\nu-\lambda b)(1+2\nu-2\lambda b)} + \frac{k[2\nu(1+2\nu)+(1-2\nu)\lambda b]b^{(2)}}{(2\nu-\lambda b)^2(1+2\nu-2\lambda b)} \\ &+\frac{2k[3+7\nu-4\nu^2-(5+2\nu)\lambda b+2\lambda^2 b^2]b^2}{(2\nu-\lambda b)^2(1+2\nu-2\lambda b)} \\ &+\frac{[4\nu(1+2\nu)+2(1-5\nu-6\nu^2)\lambda b+(1+10\nu)\lambda^2b^2-2\lambda^3b^3]b^{(2)}}{(2\nu-\lambda b)^2(1+2\nu-2\lambda b)} \\ &+\frac{2(1-\nu)[4\nu(1+2\nu)+(1-7\nu-8\nu^2)\lambda b+4\nu\lambda^2b^2]b^2}{\nu(2\nu-\lambda b)^2(1+2\nu-2\lambda b)} \end{aligned} \tag{10}$$

The LST $E[e^{-sT}|\text{idle}]$ of the DF for the response time of a customer that arrives when the system is empty is clearly given by

$$E[e^{-sT}|\text{idle}] = T_0^*(s) \tag{11}$$

It follows from (9) and (10) that

$$E[T|\text{idle}] = E[T_0] = \frac{(2-\lambda b)b}{2\nu - \lambda b} \tag{12}$$

$$E[T^2|\text{idle}] = E[(T_0)^2]$$
$$= \frac{[4\nu(1+2\nu)+2(1-5\nu-6\nu^2)\lambda b+(1+10\nu)\lambda^2b^2-2\lambda^3b^3]b^{(2)}}{(2\nu-\lambda b)^2(1+2\nu-2\lambda b)} \quad (13)$$
$$+\frac{2(1-\nu)[4\nu(1+2\nu)+(1-7\nu-8\nu^2)\lambda b+4\nu\lambda^2b^2]b^2}{\nu(2\nu-\lambda b)^2(1+2\nu-2\lambda b)}$$

In order to express $E[e^{-sT}|\text{busy}]$ in terms of $\{T_k^*(s); k=0,1,2,\ldots\}$, we note that the PGF $\Pi^o(z)$ for the queue size immediately after a service is completed (an output epoch) is given by

$$\Pi^o(z) = \nu\Pi(z)+(1-\nu)z\Pi(z)$$
$$= \frac{(\nu-\lambda b)[\nu+(1-\nu)z](1-z)B^*(\lambda-\lambda z)}{[\nu+(1-\nu)z]B^*(\lambda-\lambda z)-z} \quad (14)$$

Let X' be the length of the service time during which customer C arrives. By the analogy with (14), the generating function for the probability $\pi_k(X'=x)$ that there are k customers, excluding C, in the system at the end of $X'=x$ is given by

$$\sum_{k=0}^{\infty}\pi_k(X'=x)z^k = \frac{(\nu-\lambda b)[\nu+(1-\nu)z](1-z)e^{-\lambda(1-z)x}}{[\nu+(1-\nu)z]B^*(\lambda-\lambda z)-z} \quad (15)$$

The distribution of X' is given by

$$P\{x<X'\leq x+dx\} = \frac{x\,dB(x)}{b} \quad (16)$$

Given that $X'=x$, the LST of the DF for the remaining service time X_+ is given by

$$E[e^{-sX_+}|X'=x] = \frac{1-e^{-sx}}{sx} \quad (17)$$

The response time of customer C consists of X_+ and T_k if there are k other customers in the system at the end of service. Unconditioning on the length X' of the service time and the number of other customers in the system at the end of X', we obtain

$$E[e^{-sT}|\text{busy}] = \int_0^{\infty}\frac{1-e^{-sx}}{sx}\sum_{k=0}^{\infty}\pi_k(X'=x)T_k^*(s)\frac{xdB(x)}{b} \quad (18)$$

By expanding the integrand in (18) in Taylor series with respect to s, we get

$$E[T|\text{busy}] = \frac{1}{b}\int_0^{\infty}\left[\frac{x^2}{2}+x\sum_{k=0}^{\infty}E[T_k]\pi_k(X'=x)\right]dB(x) \quad (19)$$

$$E[T^2|\text{busy}] = \frac{1}{b}\int_0^{\infty}\left[\frac{x^3}{3}+x^2\sum_{k=0}^{\infty}E[T_k]\pi_k(X'=x)+x\sum_{k=0}^{\infty}E[(T_k)^2]\pi_k(X'=x)\right]dB(x) \quad (20)$$

From the expansion of (15) in power series of $z-1$, we get

$$\sum_{k=0}^{\infty}k\pi_k(X'=x) = \lambda x+\frac{\lambda^2b^{(2)}+2\nu(1-\nu)}{2(\nu-\lambda b)} \quad (21)$$

$$\sum_{k=2}^{\infty} k(k-1)\pi_k(X'=x) = \lambda^2 x^2 + \frac{[\lambda^2 b^{(2)} + 2\nu(1-\nu)]\lambda x}{\nu - \lambda b} + \frac{\lambda^3 b^{(3)}}{3(\nu - \lambda b)} + \frac{2(1-\nu)[\lambda^2 b^{(2)} + (1-\nu)\lambda b]}{\nu - \lambda b} + \frac{[\lambda^2 b^{(2)} + 2(1-\nu)\lambda b]^2}{2(\nu - \lambda b)^2} \tag{22}$$

Using (9) and (21) in (19), we get

$$E[T|\text{busy}] = \frac{\nu b^{(2)}}{2b(\nu - \lambda b)} + \frac{(2-\lambda b)b}{2\nu - \lambda b} + \frac{\nu(1-\nu)b}{(2\nu - \lambda b)(\nu - \lambda b)} \tag{23}$$

Using also (10) and (22) in (20), we get

$$\begin{aligned} E[T^2|\text{busy}] &= \frac{2[\nu^2(1+2\nu) - 2\nu^2\lambda b - (1-\nu)\lambda^2 b^2]}{b(\nu - \lambda b)(2\nu - \lambda b)(1 + 2\nu - 2\lambda b)} \left\{ \frac{b^{(3)}}{3} + \frac{\lambda [b^{(2)}]^2}{2(\nu - \lambda b)} \right\} \\ &+ \frac{\left\{ \begin{array}{l} \nu^2(1+2\nu)(1+9\nu-2\nu^2) - \nu(1+14\nu+35\nu^2+26\nu^3)\lambda b \\ +2\nu^2(18+29\nu)\lambda^2 b^2 + (3-14\nu-46\nu^2)\lambda^3 b^3 \\ +(1+16\nu)\lambda^4 b^4 - 2\lambda^5 b^5 \end{array} \right\} b^{(2)}}{(\nu - \lambda b)^2(2\nu - \lambda b)^2(1+2\nu-2\lambda b)} \\ &+ \frac{2(1-\nu)\left\{ \begin{array}{l} \nu^3(7+15\nu-4\nu^2) - \nu^2(10+31\nu+10\nu^2)\lambda b \\ +\nu(2+25\nu+26\nu^2)\lambda^2 b^2 + (1-7\nu-18\nu^2)\lambda^3 b^3 \\ +4\nu\lambda^4 b^4 \end{array} \right\} b^2}{\nu(\nu - \lambda b)^2(2\nu - \lambda b)^2(1+2\nu-2\lambda b)} \end{aligned} \tag{24}$$

Hence we obtain the unconditional mean response time

$$E[T] = \left(1 - \frac{\lambda b}{\nu}\right) E[T|\text{idle}] + \frac{\lambda b}{\nu} E[T|\text{busy}] = \frac{2(1-\lambda b)b + \lambda b^{(2)}}{2(\nu - \lambda b)} \tag{25}$$

which agrees with (3), and the unconditional second moment of the response time

$$\begin{aligned} E[T^2]_{\text{SIRO}} &= \left(1 - \frac{\lambda b}{\nu}\right) E[T^2|\text{idle}] + \frac{\lambda b}{\nu} E[T^2|\text{busy}] \\ &= \frac{2[\nu^2(1+2\nu) - 2\nu^2\lambda b - (1-\nu)\lambda^2 b^2]}{\nu(\nu - \lambda b)(2\nu - \lambda b)(1+2\nu-2\lambda b)} \left\{ \frac{\lambda b^{(3)}}{3} + \frac{[\lambda b^{(2)}]^2}{2(\nu - \lambda b)} \right\} \\ &+ \frac{\left\{ \begin{array}{l} 4\nu^4(1+2\nu) + \nu^2(1+\nu-18\nu^2-16\nu^3)\lambda b \\ -\nu(1+8\nu-20\nu^2-20\nu^3)\lambda^2 b^2 + \nu(2-5\nu-10\nu^2)\lambda^3 b^3 \\ +(1-\nu+2\nu^2)\lambda^4 b^4 \end{array} \right\} b^{(2)}}{\nu(\nu - \lambda b)^2(2\nu - \lambda b)^2(1+2\nu-2\lambda b)} \\ &+ \frac{2(1-\nu)\left\{ \begin{array}{l} 4\nu^3(1+2\nu) - 4\nu^2(1+4\nu+3\nu^2)\lambda b \\ +\nu(-1+14\nu+18\nu^2)\lambda^2 b^2 + (1-4\nu-10\nu^2)\lambda^3 b^3 \\ +2\nu\lambda^4 b^4 \end{array} \right\} b^2}{\nu(\nu - \lambda b)^2(2\nu - \lambda b)^2(1+2\nu-2\lambda b)} \end{aligned} \tag{26}$$

which is a new result.

3. Second Moment of the Response Time in a System with Multiple Server Vacations

We consider the same SIRO M/G/1 system with Bernoulli feedback as described in Section 2, except that the server now takes vacations if the system is empty at the end of service. If the server returns from a vacation to find the system not empty, it starts to work immediately and continues service until the system becomes empty again (*exhaustive service*). If the server returns from a vacation to find no customers waiting, it begins another vacation immediately, and continues in this manner until it finds at least one customer waiting upon returning from a vacation (*multiple vacations*). The lengths of successive vacations are assumed to be independent and identically distributed, and also independent of the arrival and service processes.

Let $V^*(s)$ be the LST of the DF for the length V of each vacation. We first discuss the queue size. For a system without feedback, the PGF for the queue size L at an arbitrary time is given by

$$\Pi(z)|_{\nu=1} = \frac{(1-\lambda b)[1-V^*(\lambda-\lambda z)]B^*(\lambda-\lambda z)}{\lambda E[V][B^*(\lambda-\lambda z)-z]} \tag{27}$$

For a system with feedback, we replace $B^*(s)$ with $B_g^*(s)$ given in (1) as well as b with b/ν to get

$$\begin{aligned}\Pi(z) &= \frac{(\nu-\lambda b)[1-V^*(\lambda-\lambda z)]B^*(\lambda-\lambda z)}{\lambda E[V]\{[\nu+(1-\nu)z]B^*(\lambda-\lambda z)-z\}} \\ &= \frac{(\nu-\lambda b)(1-z)B^*(\lambda-\lambda z)}{[\nu+(1-\nu)z]B^*(\lambda-\lambda z)-z}\cdot\frac{1-V^*(\lambda-\lambda z)}{\lambda E[V](1-z)}\end{aligned} \tag{28}$$

The last expression exhibits the *stochastic decomposition* property studied by Fuhrmann and Cooper [4]. The mean response time is then given by

$$E[T] = \frac{E[L]}{\lambda} = \frac{2(1-\lambda b)b+\lambda b^{(2)}}{2(\nu-\lambda b)} + \frac{E[V^2]}{2E[V]} \tag{29}$$

As in (4), the LST $T^*(s)$ of the DF for the response time T of an arbitrary customer can be expressed as

$$T^*(s) = \left(1-\frac{\lambda b}{\nu}\right)E[e^{-sT}|\text{vacation}] + \frac{\lambda b}{\nu}E[e^{-sT}|\text{busy}] \tag{30}$$

where $E[e^{-sT}|\text{vacation}]$ is the LST of the DF for the response time of a customer that arrives when the server is on vacation. We can obtain both $E[e^{-sT}|\text{vacation}]$ and $E[e^{-sT}|\text{busy}]$ in terms of $\{T_k^*(s); k=0,1,2,\ldots\}$ given in Section 2.

We first consider $E[e^{-sT}|\text{vacation}]$. Let V' be the length of the vacation during which customer C arrives, and let $q_k(V'=x)$ be the probability that there are k customers, excluding C, in the system at the end of $V'=x$. The generating function for $\{q_k(V'=x); k=0,1,2.\ldots\}$ is simply given by

$$\sum_{k=0}^{\infty} q_k(V'=x)z^k = e^{-\lambda(1-z)x} \tag{31}$$

By the same arguments that let to (18), we obtain

$$E[e^{-sT}|\text{vacation}] = \int_0^{\infty}\frac{1-e^{-sx}}{sx}\sum_{k=0}^{\infty} q_k(V'=x)T_k^*(s)\frac{xdV(x)}{E[V]} \tag{32}$$

where $V(x)$ is the DF for V. From (32), we get

$$E[T|\text{vacation}] = E[T_0] + \frac{(2\nu + \lambda b)E[V^2]}{2(2\nu - \lambda b)E[V]} \tag{33}$$

$$\begin{aligned} E[T^2|\text{vacation}] &= E[(T_0)^2] + \frac{2[\nu(1+2\nu) + \lambda b + \lambda^2 b^2]E[V^3]}{3(2\nu - \lambda b)(1 + 2\nu - 2\lambda b)E[V]} \\ &+ \frac{[2\nu(1+2\nu) + (1-2\nu)\lambda b]\lambda b^{(2)} E[V^2]}{(2\nu - \lambda b)^2(1 + 2\nu - 2\lambda b)E[V]} \\ &+ \frac{[4\nu(1+2\nu) + 4(1-3\nu^2)\lambda b + (-5+2\nu)\lambda^2 b^2 + 2\lambda^3 b^3] b E[V^2]}{(2\nu - \lambda b)^2(1 + 2\nu - 2\lambda b)E[V]} \end{aligned} \tag{34}$$

where $E[T_0]$ and $E[(T_0)^2]$ are given in (12) and (13), respectively.

We can express $E[e^{-sT}|\text{busy}]$ exactly as in (18) with $\{\pi_k(X' = x); k = 0, 1, 2, \ldots\}$ now given by

$$\sum_{k=0}^{\infty} \pi_k(X' = x) z^k = \frac{(\nu - \lambda b)[\nu + (1-\nu)z][1 - V^*(\lambda - \lambda z)]e^{-\lambda(1-z)x}}{\lambda E[V]\{[\nu + (1-\nu)z]B^*(\lambda - \lambda z) - z\}} \tag{35}$$

Thus we get

$$E[T|\text{busy}] = E[T|\text{busy}]|_{V=0} + \frac{\lambda b E[V^2]}{2(2\nu - \lambda b)E[V]} \tag{36}$$

$$\begin{aligned} E[T^2|\text{busy}] &= E[T^2|\text{busy}]|_{V=0} + \frac{2\lambda^2 b^2 E[V^3]}{3(2\nu - \lambda b)(1 + 2\nu - 2\lambda b)E[V]} \\ &+ \frac{\nu[2\nu(1+2\nu) - 2(1+2\nu)\lambda b + \lambda^2 b^2]\lambda b^{(2)} E[V^2]}{(\nu - \lambda b)(2\nu - \lambda b)^2(1 + 2\nu - 2\lambda b)E[V]} \\ &+ \frac{[\nu(3 + 11\nu - 8\nu^2) - (3 + 14\nu - 4\nu^2)\lambda b + (5 + 4\nu)\lambda^2 b^2 - 2\lambda^3 b^3]\lambda b^2 E[V^2]}{(\nu - \lambda b)(2\nu - \lambda b)^2(1 + 2\nu - 2\lambda b)E[V]} \end{aligned} \tag{37}$$

where $E[T|\text{busy}]|_{V=0}$ and $E[T^2|\text{busy}]|_{V=0}$ are those given in (23) and (24), respectively, for the corresponding system without server vacations.

From (30), we finally get the unconditional mean response time

$$E[T] = \left(1 - \frac{\lambda b}{\nu}\right) E[T|\text{vacation}] + \frac{\lambda b}{\nu} E[T|\text{busy}] = \frac{2(1-\lambda b)b + \lambda b^{(2)}}{2(\nu - \lambda b)} + \frac{E[V^2]}{2E[V]} \tag{38}$$

which agrees with (29), and the unconditional second moment of the response time

$$\begin{aligned} E[T^2] &= \left(1 - \frac{\lambda b}{\nu}\right) E[T^2|\text{vacation}] + \frac{\lambda b}{\nu} E[T^2|\text{busy}] \\ &= \frac{2[\nu^2(1+2\nu) - 2\nu^2\lambda b - (1-\nu)\lambda^2 b^2]}{\nu(2\nu - \lambda b)(1 + 2\nu - 2\lambda b)} \left\{ \frac{\lambda b^{(3)}}{3(\nu - \lambda b)} + \frac{[\lambda b^{(2)}]^2}{2(\nu - \lambda b)^2} + \frac{E[V^3]}{3E[V]} \right\} \\ &+ \frac{[2\nu^3(1+2\nu) - \nu^2(1+6\nu)\lambda b - 2\nu(1-2\nu)\lambda^2 b^2 + (1-\nu)\lambda^3 b^3]\lambda b^{(2)} E[V^2]}{\nu(\nu - \lambda b)(2\nu - \lambda b)^2(1 + 2\nu - 2\lambda b)E[V]} \end{aligned}$$

$$
+\frac{\left\{\begin{array}{l}4\nu^4(1+2\nu)+\nu^2(1+\nu-18\nu^2-16\nu^3)\lambda b \\ -\nu(1+8\nu-20\nu^2-20\nu^3)\lambda^2 b^2+\nu(2-5\nu-10\nu^2)\lambda^3 b^3 \\ +(1-\nu+2\nu^2)\lambda^4 b^4\end{array}\right\} b^{(2)}}{\nu(\nu-\lambda b)^2(2\nu-\lambda b)^2(1+2\nu-2\lambda b)}
$$

$$
+\frac{\left\{\begin{array}{l}4\nu^3(1+2\nu)-4\nu^2(1+4\nu+3\nu^2)\lambda b \\ +\nu(-1+14\nu+18\nu^2)\lambda^2 b^2+(1-4\nu-10\nu^2)\lambda^3 b^3 \\ +2\nu\lambda^4 b^4\end{array}\right\} b}{\nu(\nu-\lambda b)(2\nu-\lambda b)^2(1+2\nu-2\lambda b)}\left\{\frac{2(1-\nu)b}{\nu-\lambda b}+\frac{E[V^2]}{E[V]}\right\} \tag{39}
$$

which is also a new result.

If we let $\nu = 1$ in the above, we get the results for the SIRO M/G/1 system with multiple server vacations without feedback, which was studied by Scholl [7, Appendix B].

4. Numerical Comparison

As noted in Section 2, the mean response time $E[T]$ is identical for the FCFS, LCFS, and SIRO systems. We first compare the second moment $E[T^2]$ of the response time for these systems without server vacations.

The second moment $E[T^2]_{\text{FCFS}}$ of the response time in the FCFS system is given in the appendix of Takács [9]:

$$
\begin{aligned}
E[T^2]_{\text{FCFS}} &= \frac{\nu^2-2\nu}{6(\nu-\lambda b)^2[\nu^2-\nu(2+\lambda b)+\lambda b]} \\
&\quad \times \left\{2\nu[6\lambda b^3-6b^2-6\lambda b b^{(2)}+3b^{(2)}+\lambda b^{(3)}]\right. \\
&\quad \left. -12\lambda b^3+12b^2+6\lambda b b^{(2)}-2\lambda^2 b b^{(3)}+3[\lambda b^{(2)}]^2\right\}
\end{aligned} \tag{40}
$$

From the analysis of the LCFS system (see, e.g., Cohen [1, sec. III.3.2], Cooper [3, prob. 5.20], Takács [8], and Takagi [10, sec. 1.3]) with the LST of the DF for the service time given by $B_g^*(s)$ in (1), we have

$$
\begin{aligned}
E[T^2]_{\text{LCFS}} &= \frac{\nu\lambda b^{(3)}}{3(\nu-\lambda b)^2}+\frac{\nu[\lambda b^{(2)}]^2}{2(\nu-\lambda b)^3}+\frac{[\nu^2(1-2\lambda b)+\lambda^2 b^2]b^{(2)}}{(\nu-\lambda b)^3} \\
&\quad +\frac{2(1-\nu)[\nu^2-\nu(1+\nu)\lambda b+\lambda^2 b^2]b^2}{\nu(\nu-\lambda b)^3}
\end{aligned} \tag{41}
$$

For systems with no feedback ($\nu = 1$), we know [8] that the second moment $E[W^2]$ of the waiting time satisfies the relation

$$
E[W^2]_{\text{FCFS}} = (1-\lambda b)E[W^2]_{\text{LCFS}} = \left(1-\frac{\lambda b}{2}\right)E[W^2]_{\text{SIRO}} \tag{42}
$$

Since

$$
E[T^2] = E[W^2]+2E[W]b+b^2 \tag{43}
$$

and $E[W]$ is the same for the FCFS, LCFS, and SIRO systems, it follows that we have uniformly

$$
E[T^2]_{\text{FCFS}} \le E[T^2]_{\text{SIRO}} \le E[T^2]_{\text{LCFS}} \qquad \text{for } \nu = 1 \text{ (no feedback)} \tag{44}
$$

as shown in Figure 1(a) for the systems with constant service time $b = 1$. However, the inequalities in (44) do not always hold for systems with feedback ($\nu < 1$). In fact, Figure 1(b) for $\nu = 0.5$ displays the case in which

$$E[T^2]_{\mathrm{SIRO}} \le E[T^2]_{\mathrm{FCFS}} \le E[T^2]_{\mathrm{LCFS}} \tag{45}$$

Figure 1(c) for $\nu = 0.2$ shows that there is a wide range of λb in which

$$E[T^2]_{\mathrm{SIRO}} \le E[T^2]_{\mathrm{LCFS}} \le E[T^2]_{\mathrm{FCFS}} \tag{46}$$

and a narrow range of λb in which

$$E[T^2]_{\mathrm{LCFS}} \le E[T^2]_{\mathrm{SIRO}} \le E[T^2]_{\mathrm{FCFS}} \tag{47}$$

These disorders result from the various degree of variability in the time from the start of service to the departure in these systems. However it is noteworthy that SIRO discipline sometimes yields smaller values of the second moment of the response time than FCFS (and LCFS) disciplines.

We also show the second moment $E[T^2]$ of the response time in SIRO systems with multiple server vacations in Figures 2(a)–(c) for different values of constant vacation length V (the service time is again assumed to be a constant $b = 1$). The effects of vacations vanish at high loads generally, and appear more evidently for systems with large values of ν (low probability of feedback).

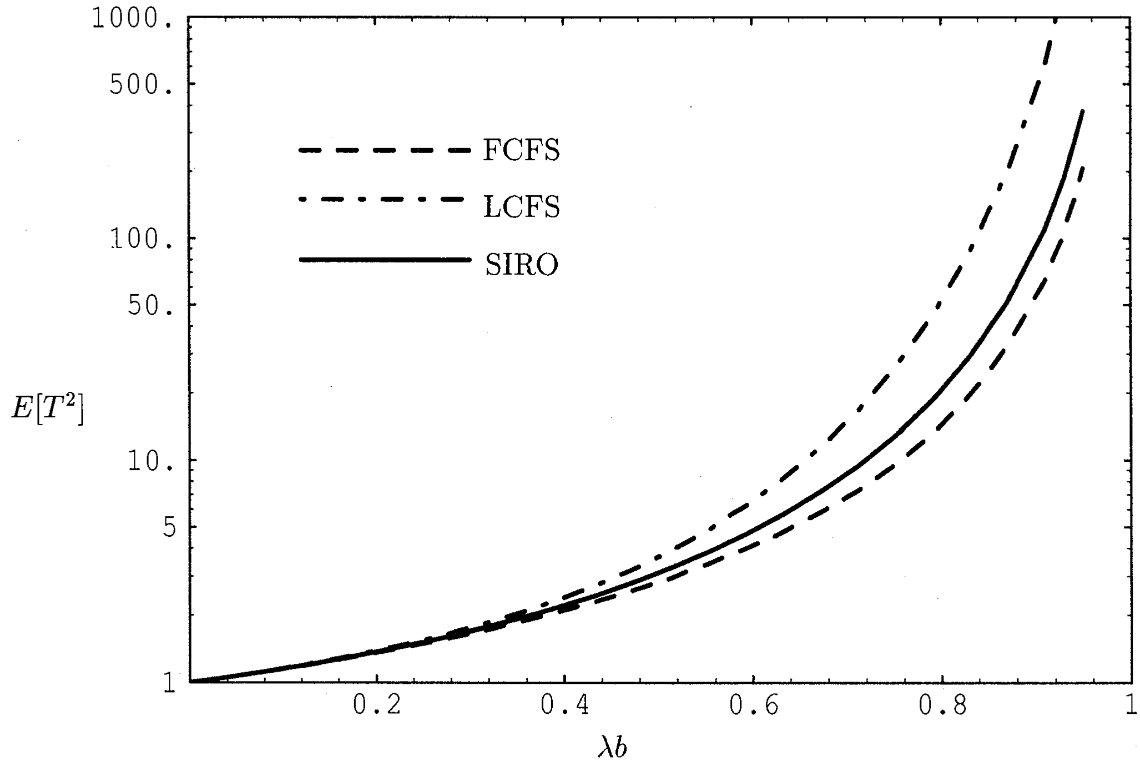

Figure 1(a). Effects of service discipline ($\nu = 1, b = 1, V = 0$).

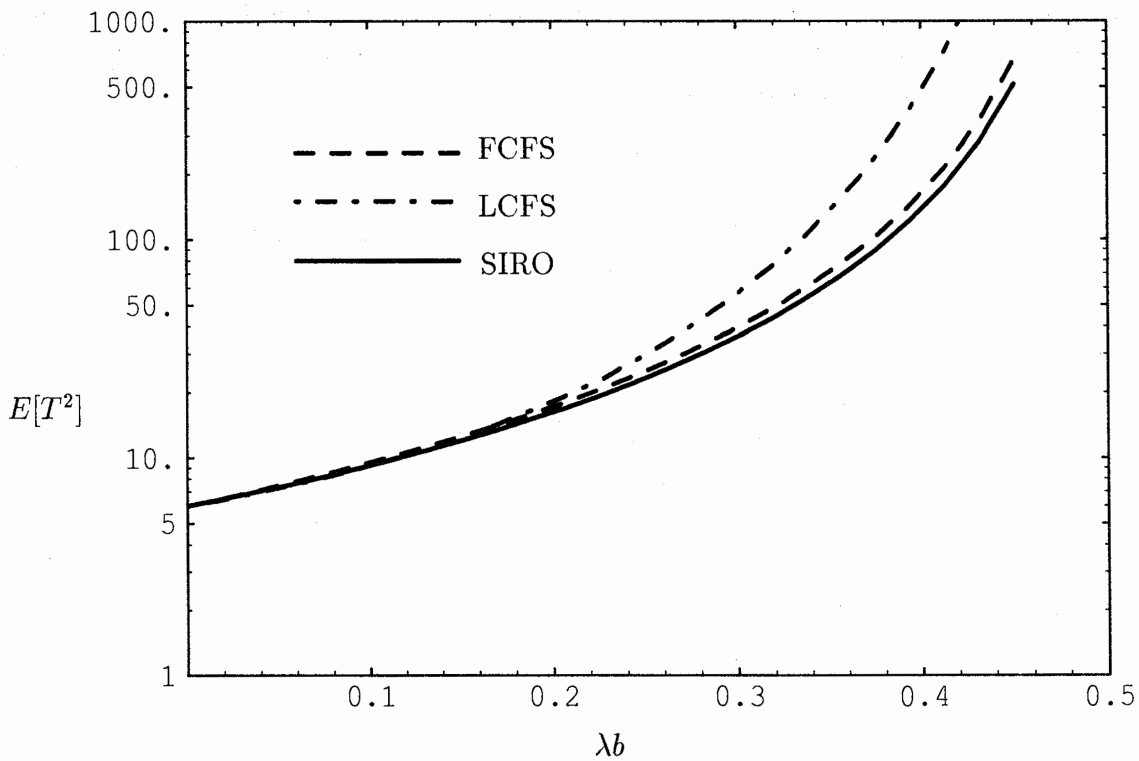

Figure 1(b). Effects of service discipline ($\nu = 0.5, b = 1, V = 0$).

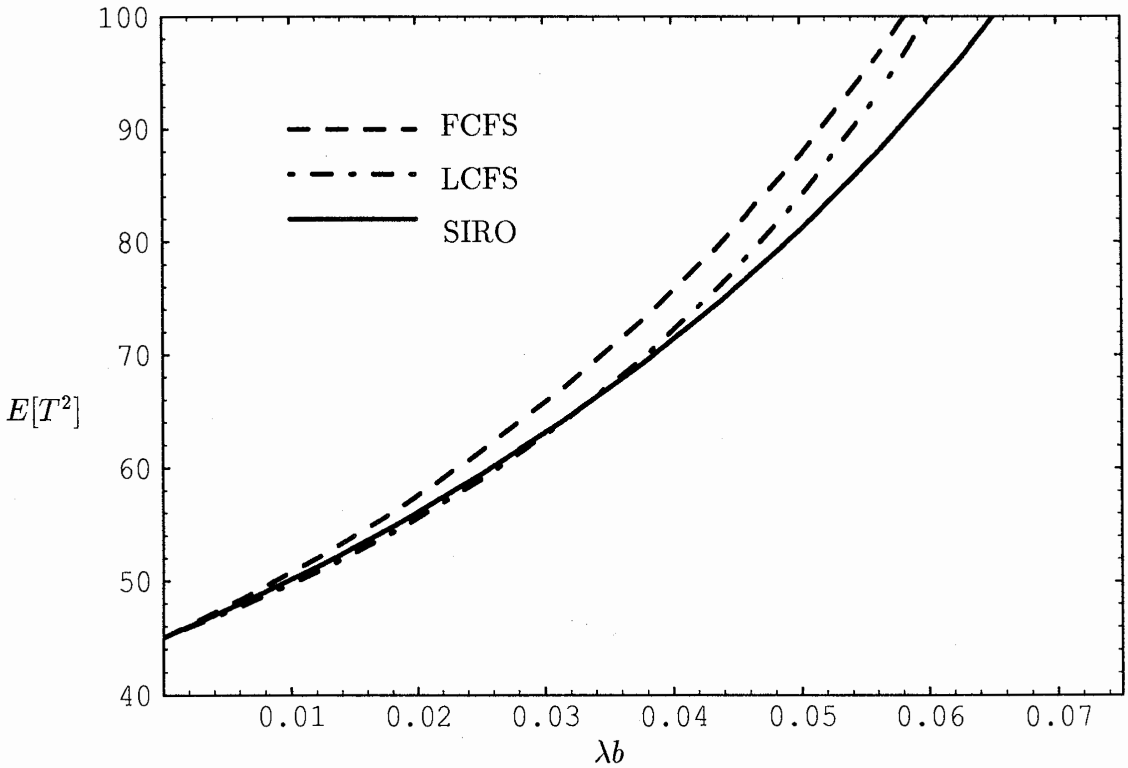

Figure 1(c). Effects of service discipline ($\nu = 0.2, b = 1, V = 0$).

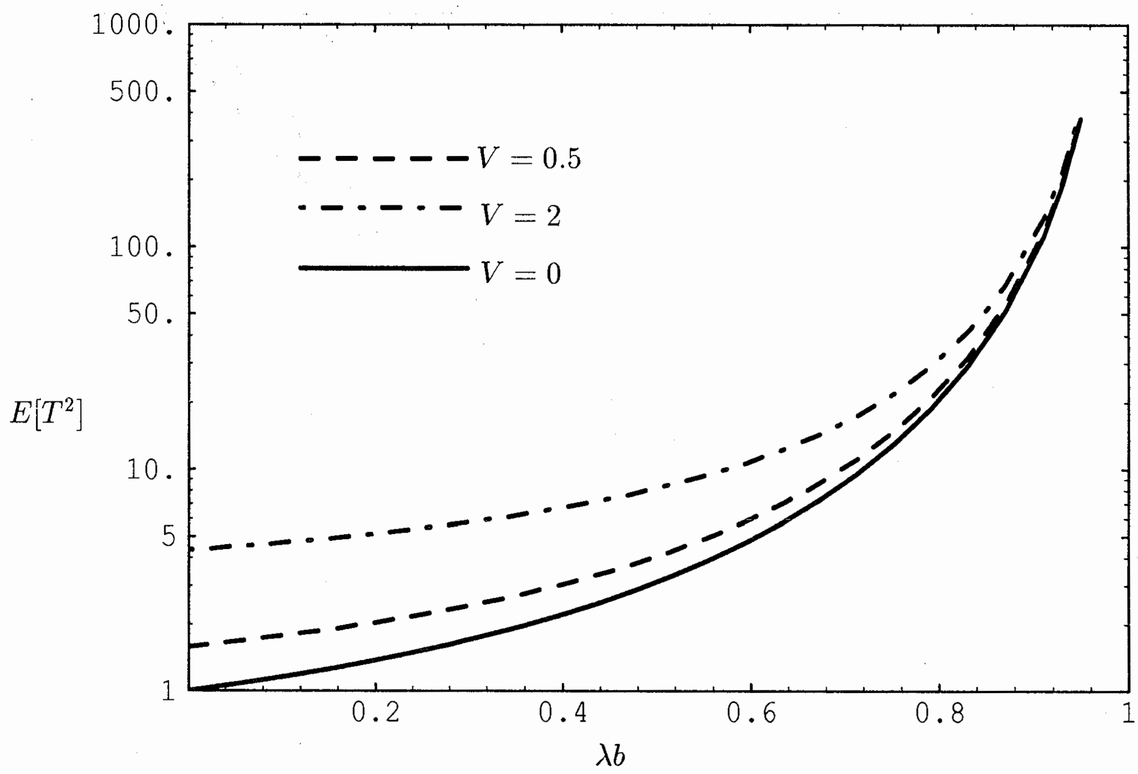

Figure 2(a). Effects of vacations (SIRO, $\nu = 1, b = 1$).

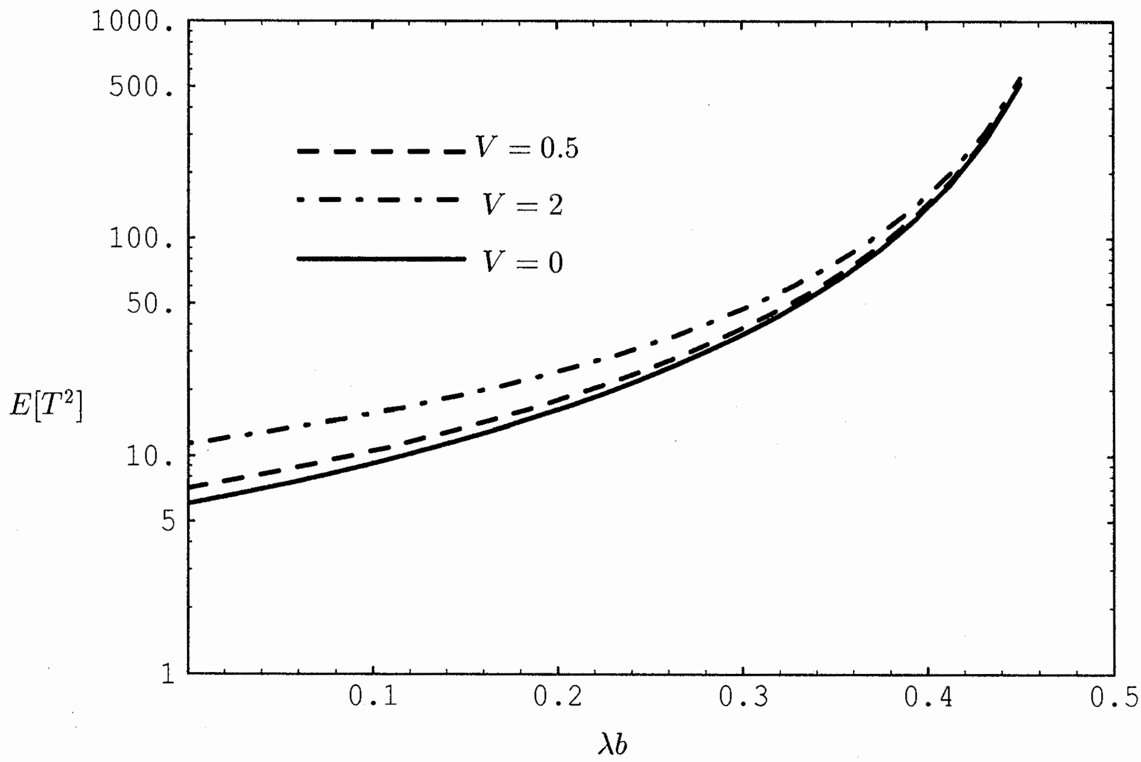

Figure 2(b). Effects of vacations (SIRO, $\nu = 0.5, b = 1$).

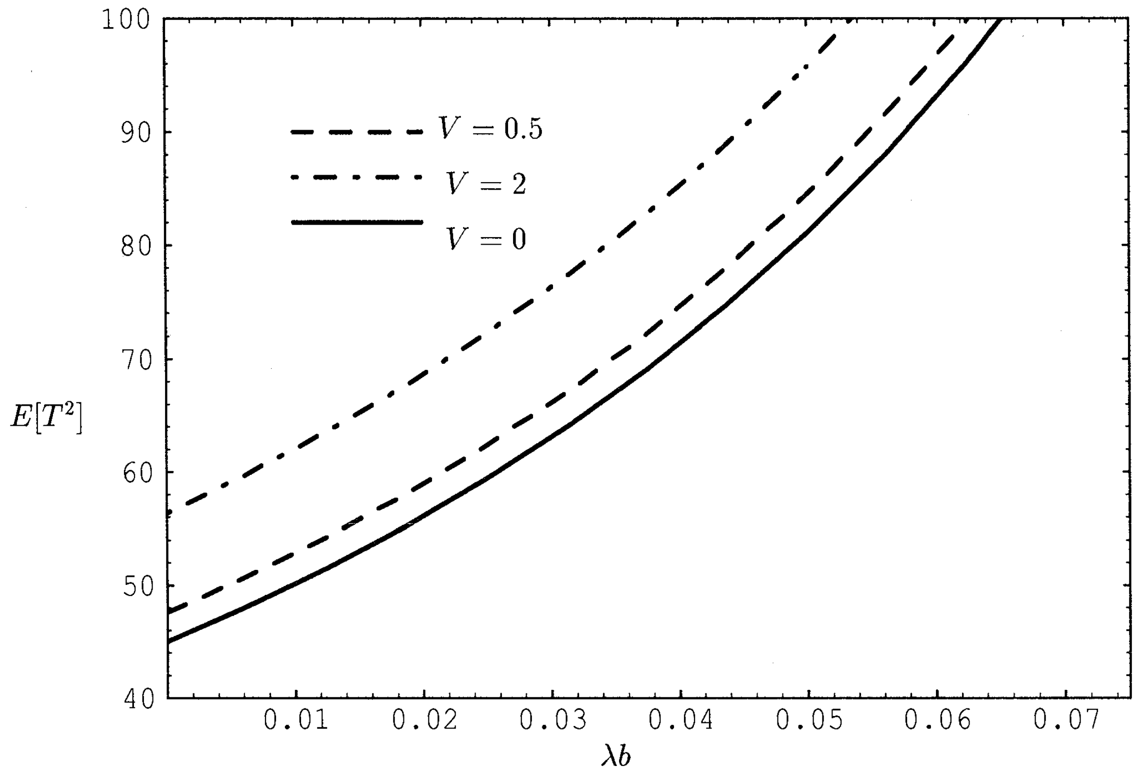

Figure 2(c). Effects of vacations (SIRO, $\nu = 0.2, b = 1$).

Appendix: Derivation of $E[T_k]$ and $E[(T_k)^2]$

We derive the expressions for $E[T_k]$ and $E[(T_k)^2]$ in (9) and (10), respectively. Taking the first derivative of (8) at $s = 0$, we get

$$\begin{aligned} E[T_k] &= \nu \left[\frac{b}{k+1} + \frac{k}{k+1} \sum_{j=0}^{\infty} \left\{ -a_j^{*(1)}(0) + a_j^*(0) E[T_{k+j-1}] \right\} \right] \\ &+ (1-\nu) \sum_{j=0}^{\infty} \left\{ -a_j^{*(1)}(0) + a_j^*(0) E[T_{k+j}] \right\} \end{aligned} \quad \text{(A.1)}$$

where $a_j^{*(i)}(s) := d^i a_j^*(s)/ds^i$ $(i = 1, 2, \ldots)$. By differentiating (6) with respect to s and then letting $s = 0, z = 1$, we have

$$\sum_{j=0}^{\infty} a_j^{*(1)}(0) = -b \quad ; \quad \sum_{j=0}^{\infty} a_j^{*(i)}(0) = (-1)^i b^{(i)} \qquad i = 2, 3, \ldots \quad \text{(A.2)}$$

Thus, (A.1) can be written as

$$E[T_k] = \frac{\nu}{k+1} \left\{ (k+1)b + k \sum_{j=0}^{\infty} a_j^*(0) E[T_{k+j-1}] \right\} + (1-\nu) \left\{ b + \sum_{j=0}^{\infty} a_j^*(0) E[T_{k+j}] \right\} \quad \text{(A.3)}$$

This recursive relation for $\{E[T_k]; k = 0, 1, 2, \ldots\}$ can be solved by assuming the form

$$E[T_k] = ck + d \quad \text{(A.4)}$$

where c and d are constants. We substitute (A.4) into (A.3), and use

$$\sum_{j=0}^{\infty} a_j(0) = 1 \quad ; \quad \sum_{j=1}^{\infty} j a_j^*(0) = \lambda b \tag{A.5}$$

which can also be obtained from (6). The result is given by

$$(k+1)(ck+d) = \nu\{(k+1)b + k[c(k-1)+d] + ck\lambda b\} + (1-\nu)(k+1)(b+ck+d+c\lambda b) \tag{A.6}$$

Comparing the coefficients of k and constant terms, we determine that

$$c = \frac{b}{2\nu - \lambda b} \quad ; \quad d = \frac{(2-\lambda b)b}{2\nu - \lambda b} \tag{A.7}$$

Substituting (A.7) into (A.4) leads to (9).

Similarly, taking the second derivative of (8) at $s = 0$ and using (A.2), we get the following recursive relation for $\{E[(T_k)^2]; k = 0, 1, 2, \ldots\}$:

$$\begin{aligned} E[(T_k)^2] &= \frac{\nu}{k+1}\left\{(k+1)b^{(2)} - 2k\sum_{j=0}^{\infty} a_j^{*(1)}(0)E[T_{k+j-1}] + k\sum_{j=0}^{\infty} a_j^*(0)E[(T_{k+j-1})^2]\right\} \\ &+(1-\nu)\left\{b^{*(2)} - 2\sum_{j=0}^{\infty} a_j^{*(1)}(0)E[T_{k+j}] + \sum_{j=0}^{\infty} a_j^*(0)E[(T_{k+j})^2]\right\} \end{aligned} \tag{A.8}$$

We now assume the form

$$E[(T_k)^2] = ek^2 + fk + g \tag{A.9}$$

where e, f, and g are constants. Substituting (A.4) and (A.9) into (A.8), and using

$$\sum_{j=1}^{\infty} j a_j^{*(1)}(0) = -\lambda b^{(2)} \quad ; \quad \sum_{j=2}^{\infty} j(j-1)a_j^*(0) = \lambda^2 b^{(2)} \tag{A.10}$$

we get

$$\begin{aligned} (k+1)(ek^2+fk+g) &= \nu\left\{(k+1)b^{(2)} + 2k\frac{(k+1-\lambda b)b^2 + \lambda b b^{(2)}}{2\nu - \lambda b}\right. \\ &\left. +k[e\{(k-1)^2 + (2k-1)\lambda b + \lambda^2 b^{(2)}\} + f(k-1+\lambda b) + g]\right\} \\ +(1-\nu)(k+1)&\left\{b^{(2)} + 2\frac{(k+2-\lambda b)b^2 + \lambda b b^{(2)}}{2\nu - \lambda b} + e\{k^2 + (2k+1)\lambda b + \lambda^2 b^{(2)}\} + f(k+\lambda b) + g\right\} \end{aligned} \tag{A.11}$$

Comparing the coefficients of k^0, k^1, and k^2, we determine e, f, and g, which are substituted into (A.9) to derive (10).

Acknowledgment

This work is supported in part by the Grant No. 94-15 of the Okawa Institute of Information and Telecommunication. The author is grateful to anonymous referees whose comments were valuable in revising the paper.

References

[1] J. W. Cohen, *The Single Server Queue.* Revised edition, North-Holland Publishing Company, Amsterdam, 1982.

[2] B. Conolly, *Lecture Notes on Queueing Systems.* Ellis Horwood Limited, Sussex, England, 1975.

[3] R. B. Cooper, *Introduction to Queueing Theory.* Second edition, North-Holland Publishing Company, New York, 1981.

[4] S. W. Fuhrmann and R. B. Cooper, Stochastic decompositions in the M/G/1 queue with generalized vacations. *Operations Research,* Vol.33, No.5, pp.1117–1129, September–October, 1985.

[5] S. W. Fuhrmann, Second moment relationships for waiting times in queueing systems with Poisson input. *Queueing Systems,* Vol.8, No.4, pp.397–406, June 1991.

[6] J. F. C. Kingman, On queues in which customers are served in random order. *Proceedings of the Cambridge Philosophical Society,* Vol.58, Part 1, pp.79–91, January 1962.

[7] M. O. Scholl, Multiplexing techniques for data transmission over packet switched radio systems. UCLA-ENG-76123, Computer Science Department, University of California, Los Angeles, 1976.

[8] L. Takács, Delay distributions for one line with Poisson input, general holding times, and various orders of service. *The Bell System Technical Journal,* Vol.42, No.2, pp.487–503, March 1963.

[9] L. Takács, A single-server queue with feedback. *The Bell System Technical Journal,* Vol.42, No.2, pp.505–519, March 1963.

[10] H. Takagi, *Queueing Analysis : A Foundation of Performance Evaluation, Volume 1: Vacation and Priority Systems, Part 1.* Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1991.

Hideaki Takagi
Institute of Policy and Planning Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba-shi
Ibaraki 305, Japan
e-mail: takagi@shako.sk.tsukuba.ac.jp