Journal of the Operations Research
Society of Japan
Vol. 43, No. 4, December 2000

WAITING TIME ANALYSIS OF $M^X/G/1$ QUEUES WITH/WITHOUT VACATIONS UNDER RANDOM ORDER OF SERVICE DISCIPLINE

Norikazu Kawasaki
Sumitomo Electric Industries, Ltd.

Hideaki Takagi
University of Tsukuba

Yutaka Takahashi
Kyoto University

Sung-Jo Hong
Dongguk University

Toshiharu Hasegawa
Nanzan University

(Received September 10, 1999; Final June 2, 2000)

Abstract We study (batch arrival) $M^X/G/1$ queues with/without vacations under random order of service (ROS) discipline. By considering the conditional waiting times given the states of the system when an arbitrary message arrives, we derive the Laplace-Stieltjes transforms of the waiting time distributions and explicitly obtain their first two moments. The relationship for the second mements under ROS and first-come first-served disciplines is shown to be precisely the same as that found by Takács and Fuhrmann for (single arrival) M/G/1 queues.

1. Introduction

We consider (batch Poisson arrival) $M^X/G/1$ queues with/without vacations under *random order of service* (ROS) discipline. Messages arrive in batches at a buffer of infinite capacity and are served for generally distributed service times. A single server works continuously until the system becomes empty. When the server finds the system empty, he waits for the first batch to arrive at the system in non-vacation models, or he takes a *vacation* in vacation models. We assume that the lengths of vacations are independent and identically distributed.

The $M^X/G/1$ queues under *first-come first-served* (FCFS) discipline have been studied in the literature. For example, Burke [3], Cooper (sec. 5.10 in [5]), Kleinrock (prob. 5.11 and 5.12 in [10]), and Takagi (sec. 1.4 in [17]) studied those without vacations, so did Baba [1] with vacations. Baba [2] also studied $M^X/G/1$ queues under *last-come first-served* (LCFS) discipline with and without vacations.

Under the ROS discipline, the next message for service is selected at random among messages waiting in the queue. The ROS discipline is one of the three basic queueing disciplines (FCFS, LCFS, and ROS) whereby a message is selected for the next service. Kingman [9], Takács [16], Conolly [4], and Takagi and Kudoh [18] studied (single arrival) M/G/1 queues without vacations. Scholl and Kleinrock [15] studied an M/G/1 queue with multiple vacations. The results in this paper for $M^X/G/1$ queues with and without vacations under ROS discipline are new, and include all the above as special cases.

As for vacations, we consider two cases (Doshi [6], Levy and Yechiali [13]). If the server returns from a vacation to find no messages waiting, in the *multiple vacation* case, he begins another vacation immediately; in the *single vacation* case, he waits for the first batch to arrive while keeping the system idle.

In this paper we study the following three models:

NV $M^X/G/1$ without vacations,

MV $M^X/G/1$ with multiple vacations,

SV $M^X/G/1$ with single vacations.

Our objective is the derivation of the first two moments of the waiting time distribution for the above three cases. First, in Section 2 we derive the queue size distribution of messages in each model at the beginning of service to a message. Next, we derive the waiting time moments for the NV model in Section 3, for the MV model in Section 4, and for the SV model in Section 5. We then make some comparisons with FCFS systems through numerical examples in Section 6.

We use the following notation:

λ	arrival rate of batches,
$B(x)$	cumulative distribution function (CDF) for service time of a message,
$B^*(s)$	Laplace-Stieltjes transform (LST) of $B(x)$,
b	mean service time,
$b^{(i)}$	ith moment of the service time,
V	vacation time,
$V(x)$	CDF for V,
$V^*(s)$	LST of $V(x)$,
g_k	probability that the batch size (number of messages in a batch) is k,
$G(z)$	generating function (GF) for g_k,
$G^{(1)}(z)$	first derivative of $G(z)$,
g	mean batch size,
$g^{(i)}$	ith factorial moment of the batch size,
ρ	traffic intensity ($\rho = \lambda g b$),
$W^*(s)$	LST of the CDF for the waiting time of an arbitrary message,
$E[W^i]$	ith moment of the waiting time,
$E[\cdot]$	expected value of a random variable.

We assume the existence of the steady state in the system, namely, $\rho < 1$. We also assume that the moments for the batch size, the service time, and the vacation time exist to the degree that appears in $E[W]$ and $E[W^2]$, and that $V^*(s)$ exists for MV as well as for SV models.

2. Queue Size at a Service Start Point

In this section, we derive the probability generating function (PGF) for the number of messages waiting for service in the queue at the beginning of service to a message in the steady state, denoted by $\Phi(z)$. We can apply an identical approach to all the above models. Note that the queue size distribution is invariant to the order of service as long as the service discipline selects customers in a way that is independent of their service time (sec. 3.4 in Kleinrock [11]).

First, we derive the PGF $\Pi(z)$ for the queue size at the departure point of an arbitrary message in the steady state, by using the method of the *embedded Markov chain*. By adopting each departure point as a Markov point in a manner which is standard in the analysis of M/G/1 type queues (sec. 5.8 in Cooper [5], sec. 5.3 in Kleinrock [10], and sec. 1.1 in Takagi [17]), we have the following equations for each model.

NV

$$\Pi(z) = \pi_0 G(z) \frac{B^*[\lambda - \lambda G(z)]}{z} + [\Pi(z) - \pi_0] \frac{B^*[\lambda - \lambda G(z)]}{z}, \tag{1a}$$

MV

$$\Pi(z) = \pi_0 \frac{V^*[\lambda - \lambda G(z)] - V^*(\lambda)}{1 - V^*(\lambda)} \cdot \frac{B^*[\lambda - \lambda G(z)]}{z} + [\Pi(z) - \pi_0] \frac{B^*[\lambda - \lambda G(z)]}{z}, \tag{1b}$$

SV

$$\begin{aligned} \Pi(z) &= \pi_0 V^*(\lambda) G(z) \frac{B^*[\lambda - \lambda G(z)]}{z} + \pi_0 [V^*[\lambda - \lambda G(z)] - V^*(\lambda)] \frac{B^*[\lambda - \lambda G(z)]}{z} \\ &\quad + [\Pi(z) - \pi_0] \frac{B^*[\lambda - \lambda G(z)]}{z}, \end{aligned} \tag{1c}$$

where π_0 denotes the probability that there are no messages in the system at the departure point. Solving (1) and determining π_0 by normalization condition $\Pi(1) = 1$, we have

NV

$$\Pi(z) = \frac{1-\rho}{g} \cdot \frac{[1 - G(z)] B^*[\lambda - \lambda G(z)]}{B^*[\lambda - \lambda G(z)] - z}, \tag{2a}$$

MV

$$\Pi(z) = \frac{1-\rho}{\lambda g E[V]} \cdot \frac{[1 - V^*[\lambda - \lambda G(z)]] B^*[\lambda - \lambda G(z)]}{B^*[\lambda - \lambda G(z)] - z}, \tag{2b}$$

SV

$$\Pi(z) = \frac{1-\rho}{g(V^*(\lambda) + \lambda E[V])} \cdot \frac{1 - V^*[\lambda - \lambda G(z)] + [1 - G(z)] V^*(\lambda)}{B^*[\lambda - \lambda G(z)] - z} \cdot B^*[\lambda - \lambda G(z)]. \tag{2c}$$

We note that the expression in (2a) appears in Kleinrock and Gail (prob. 5.12 in [12]) and Takagi (exercise 1.5 in [17]) as a correction to Kleinrock (prob. 5.12 in [10]). From (2) and by noting that the PGF of the number of messages arriving in a service time is given by $B^*[\lambda - \lambda G(z)]$, we can obtain $\Phi(z)$ as follows.

NV

$$\Phi(z) = \frac{1-\rho}{g} \cdot \frac{1 - G(z)}{B^*[\lambda - \lambda G(z)] - z}, \tag{3a}$$

MV

$$\Phi(z) = \frac{1-\rho}{\lambda g E[V]} \cdot \frac{1 - V^*[\lambda - \lambda G(z)]}{B^*[\lambda - \lambda G(z)] - z}, \tag{3b}$$

SV

$$\Phi(z) = \frac{1-\rho}{g(V^*(\lambda) + \lambda E[V])} \cdot \frac{1 - V^*[\lambda - \lambda G(z)] + [1 - G(z)] V^*(\lambda)}{B^*[\lambda - \lambda G(z)] - z}. \tag{3c}$$

3. Waiting Time for the NV Model

The waiting time W of an arbitrary message is defined as the time interval from its arrival to the service start. Consider an arbitrary message, denoted by M, in a system without server vacations. First, we derive the conditional waiting time distribution of M when it arrives during an idle period in Section 3.1, and that during a busy period in Section 3.2. Because Poisson arrivals see time averages (PASTA) (sec. 11.2 in Heyman and Sobel [8]), we have

$$E[W^i] = (1-\rho)E[W^i|\text{idle}] + \rho E[W^i|\text{busy}] \qquad i = 1, 2, \cdots. \tag{4}$$

3.1. Conditional waiting time — idle case

If M arrives during an idle period, it has a chance to be selected for service immediately. Suppose that M arrives with k other messages in a batch to find the server idle. Denoting by Π^{I} the number of messages, other than M, included in the batch, we have

$$\pi_k^{\text{I}} = \text{Prob}\{\Pi^{\text{I}} = k\} = \frac{(k+1)g_{k+1}}{\sum_{j=0}^{\infty}(j+1)g_{j+1}} = \frac{(k+1)g_{k+1}}{g},$$

which yields

$$\Pi^{\text{I}}(z) = E[z^{\Pi^{\text{I}}}] = \frac{G^{(1)}(z)}{g}, \tag{5a}$$

$$E[\Pi^{\text{I}}] = \frac{g^{(2)}}{g}, \tag{5b}$$

$$E[(\Pi^{\text{I}})^2] - E[\Pi^{\text{I}}] = \frac{g^{(3)}}{g}. \tag{5c}$$

Next, let $W_k^*(s)$ be the LST of the CDF for the waiting time of M from the epoch that M gets a chance to be selected for service, on the condition that there are k messages excluding M in the system at the epoch. If M is selected for the next service immediately, which occurs with probability $1/(k+1)$, the waiting time is zero; otherwise M is delayed, which occurs with probability $k/(k+1)$, and it will be served after a later service completion (Figure 1). Denoting by j the number of messages which arrive during the service time, thus there being $k+j-1$ messages excluding M in the system when the service ends, we have the following recurrence formula

$$W_k^*(s) = \frac{1}{k+1} + \frac{k}{k+1}\sum_{j=0}^{\infty} B_j^*(s)W_{k+j-1}^*(s), \qquad k = 0, 1, 2, \cdots, \tag{6}$$

where $B_j^*(s)$ $(j = 0, 1, \ldots)$ denotes the joint LST of the CDF for a message service time and the probability that j messages arrive during that service time, and satisfies

$$\sum_{j=0}^{\infty} B_j^*(s)z^j = B^*[s+\lambda-\lambda G(z)]. \tag{7}$$

Equation (6) extends Kingman's result [9] which gives the formula for the M/G/1 queue. By following Takács [16], we obtain the first two moments as follows (Appendix A).

$$E[W_k] = \frac{kb}{2-\rho}, \tag{8a}$$

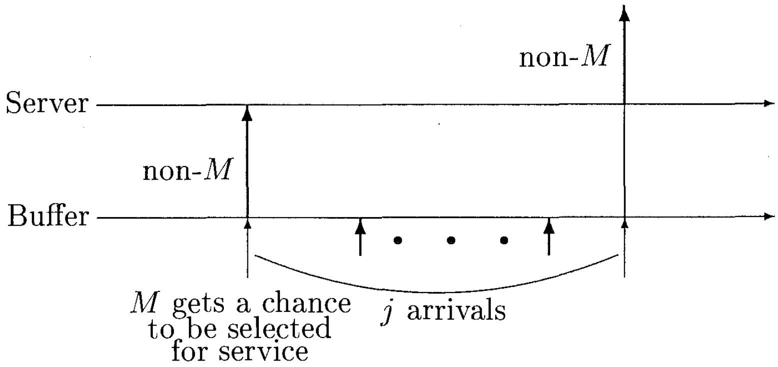

Figure 1: The conditional waiting time when M is not selected with prob. $k/(k+1)$.

$$E[W_k^2] = \frac{2k(k-1)b^2}{(2-\rho)(3-2\rho)} + \frac{k[(6-\rho)b^{(2)} + 2\lambda g^{(2)} b^3]}{(2-\rho)^2(3-2\rho)}. \tag{8b}$$

From (5), (6) and (8), we obtain

$$E[e^{-sW}|\text{idle}] = \sum_{k=0}^{\infty} \pi_k^{\mathrm{I}} W_k^*(s), \tag{9a}$$

$$E[W|\text{idle}] = \frac{g^{(2)} b}{(2-\rho)g}, \tag{9b}$$

$$E[W^2|\text{idle}] = \frac{2g^{(3)}b^2}{(2-\rho)(3-2\rho)g} + \frac{g^{(2)}[(6-\rho)b^{(2)} + 2\lambda g^{(2)} b^3]}{(2-\rho)^2(3-2\rho)g}. \tag{9c}$$

3.2. Conditional waiting time — busy case

If the server is busy at the arrival time of M, it is only after the completion of current service that M gets a chance to be selected for service. Let x be the length of the service which is going on when M arrives. First, we derive the waiting time conditioned on x. The waiting time of M consists of the remaining time of x with the LST $W_\alpha^*(s|x)$ and the time thereafter until the start of a service to M with the LST $W_\beta^*(s|x)$ (Figure 2). Note that the two conditional waiting times are independent of each other.

3.2.1. Derivation of $W_\alpha^*(s|x)$

Since an arrival point is uniformly distributed over $[0, x]$, we immediately have

$$W_\alpha^*(s|x) = \int_0^x e^{-sy} \frac{dy}{x} = \frac{1 - e^{-sx}}{sx}. \tag{10}$$

3.2.2. Derivation of $W_\beta^*(s|x)$

When the current service ends, M gets the first chance to be selected for service. Let $\pi_k^{\mathrm{B}}(x)$ be the probability that there are k messages excluding M in the system when the service x ends, and $\Pi^{\mathrm{B}}(z;x)$ be its GF. $\Pi^{\mathrm{B}}(z;x)$ is given as the product of the following three independent terms. The first is (3a), the PGF for the number of messages in the system when the service starts. The second is $e^{-\lambda[1-G(z)]x}$ which represents the PGF for the number of messages that arrive during the service, excluding those in the batch M belongs to. The

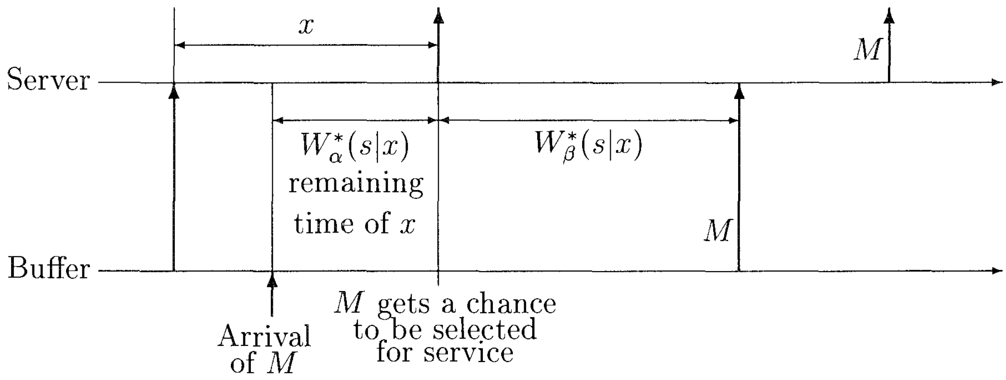

Figure 2: The conditional waiting time when M arrives during a busy period.

third is $G^{(1)}(z)/g$ which represents the PGF for the number of messages arriving with M in a batch and excluding M. Hence we have

$$\Pi^{\mathrm{B}}(z;x)=\sum_{k=0}^{\infty}\pi_k^{\mathrm{B}}(x)z^k=\frac{1-\rho}{g}\cdot\frac{[1-G(z)]e^{-\lambda[1-G(z)]x}}{B^*[\lambda-\lambda G(z)]-z}\cdot\frac{G^{(1)}(z)}{g}. \tag{11}$$

By using (11), we obtain

$$W^*_\beta(s|x)=\sum_{k=0}^{\infty}\pi_k^{\mathrm{B}}(x)W_k^*(s). \tag{12}$$

3.2.3. Unconditioning on x

The LST of the conditional waiting time distribution is given by

$$E[e^{-sW}|\text{busy},x]=W^*_\alpha(s|x)\cdot W^*_\beta(s|x). \tag{13}$$

We now uncondition this equation with respect to x. The probability that a message arrives during a service of length x is proportional to x as well as to the relative frequency of such length, thus given by $xdB(x)/b$ (sec. 5.2 in Kleinrock [10]). Substituting (10) and (12) into (13) and then unconditioning, we obtain

$$E[e^{-sW}|\text{busy}] = \int_0^\infty \frac{xdB(x)}{b}\frac{1-e^{-sx}}{sx}\sum_{k=0}^{\infty}\pi_k^{\mathrm{B}}(x)W_k^*(s). \tag{14}$$

From (8), (11), and (14), its first two moments are given by

$$E[W|\text{busy}]=\frac{(3-2\rho)g^{(2)}b}{2(1-\rho)(2-\rho)g}+\frac{b^{(2)}}{2(1-\rho)b}, \tag{15a}$$

$$\begin{aligned}E[W^2|\text{busy}] &= \frac{2b^{(3)}}{3(1-\rho)(2-\rho)b}+\frac{\lambda g[b^{(2)}]^2}{(1-\rho)^2(2-\rho)b}+\frac{2(4-3\rho)g^{(3)}b^2}{3(1-\rho)(2-\rho)(3-2\rho)g}\\ &+\frac{(18-26\rho+11\rho^2-\rho^3)g^{(2)}b^{(2)}}{(1-\rho)^2(2-\rho)^2(3-2\rho)g}+\frac{(4-\rho-4\rho^2+2\rho^3)[g^{(2)}b]^2}{(1-\rho)^2(2-\rho)^2(3-2\rho)g^2}.\end{aligned} \tag{15b}$$

3.3. Unconditional waiting time

By substituting (9) and (15) into (4), we get the first two moments of the waiting time

$$E[W] = \frac{\lambda g b^{(2)}}{2(1-\rho)} + \frac{g^{(2)} b}{2(1-\rho)g}, \tag{16a}$$

$$E[W^2] = \frac{2\lambda g b^{(3)}}{3(1-\rho)(2-\rho)} + \frac{[\lambda g b^{(2)}]^2}{(1-\rho)^2(2-\rho)} + \frac{2g^{(3)}b^2}{3(1-\rho)(2-\rho)g} + \frac{(1+\rho)g^{(2)}b^{(2)} + \lambda[g^{(2)}]^2 b^3}{(1-\rho)^2(2-\rho)g}. \tag{16b}$$

4. Waiting Time for the MV Model

In the MV model, if the server returns from a vacation to find no messages waiting, he begins another vacation immediately. A *regenerative point* (sec. 6.4 in Heyman and Sobel [8]) of this system is the epoch at which the system becomes empty and a vacation begins. The time interval between two such successive regenerative points is called a *regeneration cycle* (sec. 2.2 in Takagi [17]), whose length is denoted by V_c. The LST $V_c^*(s)$ of the CDF and the mean for V_c are given by

$$V_c^*(s) = V^*[s+\lambda-\lambda\Theta_g^*(s)], \tag{17a}$$

$$E[V_c] = \frac{E[V]}{1-\rho}, \tag{17b}$$

where $\Theta_g^*(s)$ is the LST of the CDF for the length Θ_g of a busy period initiated with the service times of the messages included in a batch, and satisfies the equation

$$\Theta_g^*(s) = G\{B^*[s+\lambda-\lambda\Theta_g^*(s)]\} \tag{18}$$

which gives

$$E[\Theta_g] = \frac{bg}{1-\rho}. \tag{19}$$

A vacation always appears once in a regeneration cycle, thus

$$\text{Prob[vacation]} = \frac{E[V]}{E[V_c]} = 1-\rho, \tag{20}$$

which gives

$$\text{Prob[busy]} = \rho. \tag{21}$$

Hence we have

$$E[W^i] = (1-\rho)E[W^i|\text{vacation}] + \rho E[W^i|\text{busy}] \qquad i = 1, 2, \cdots. \tag{22}$$

4.1. Conditional waiting time — vacation case

We can derive the conditional waiting time similarly to Section 3.2. Letting x be the length of a vacation which is effective when M arrives, we have

$$E[e^{-sW}|\text{vacation}] = \int_0^\infty \frac{x dV(x)}{E[V]} \frac{1-e^{-sx}}{sx} \sum_{k=0}^{\infty} \pi_k^{V}(x) W_k^*(s), \tag{23}$$

where $\pi_k^{\mathrm{V}}(x)$ denotes the probability that there are k messages excluding M in the system when a vacation of length x ends. Those messages consist of the following two types of messages. The first is the group of messages that arrive during the vacation, excluding those in the batch M belongs to. The second is the group of messages arriving in the same batch as M, excluding M. Thus we have

$$\sum_{k=0}^{\infty} \pi_k^{\mathrm{V}}(x) z^k = e^{-\lambda[1-G(z)]x} \cdot \frac{G^{(1)}(z)}{g}. \tag{24}$$

The first two moments of (23) are given by

$$E[W|\text{vacation}] = \frac{g^{(2)}b}{g(2-\rho)} + \frac{(2+\rho)E[V^2]}{2(2-\rho)E[V]}, \tag{25a}$$

$$\begin{aligned} E[W^2|\text{vacation}] &= \frac{2g^{(3)}b^2}{(2-\rho)(3-2\rho)g} + \frac{g^{(2)}[(6-\rho)b^{(2)} + 2\lambda g^{(2)} b^3]}{(2-\rho)^2(3-2\rho)g} \\ &+ \frac{[(6-\rho)\lambda g^2 b^{(2)} + (6+5\rho-2\rho^2)g^{(2)}b]E[V^2]}{(2-\rho)^2(3-2\rho)gE[V]} + \frac{2(3+\rho+\rho^2)E[V^3]}{3(2-\rho)(3-2\rho)E[V]}. \end{aligned} \tag{25b}$$

4.2. Conditional waiting time — busy case

By an argument similar to that in Section 3.2, we get the conditional waiting time if the server is busy when M arrives:

$$E[e^{-sW}|\text{busy}] = \int_0^{\infty} \frac{x dB(x)}{b} \frac{1-e^{-sx}}{sx} \sum_{k=0}^{\infty} \pi_k^{\mathrm{B}}(x) W_k^*(s), \tag{26}$$

where, by using (3b), $\pi_k^{\mathrm{B}}(x)$ for the MV model is given by

$$\sum_{k=0}^{\infty} \pi_k^{\mathrm{B}}(x) z^k = \frac{(1-\rho)\{1 - V^*[\lambda - \lambda G(z)]\} e^{-\lambda[1-G(z)]x}}{\lambda g E[V][B^*[\lambda - \lambda G(z)] - z]} \cdot \frac{G^{(1)}(z)}{g}. \tag{27}$$

The first two moments of (26) are given by

$$E[W|\text{busy}] = \frac{(3-2\rho)g^{(2)}b}{2(1-\rho)(2-\rho)g} + \frac{b^{(2)}}{2(1-\rho)b} + \frac{\rho E[V^2]}{2(2-\rho)E[V]}, \tag{28a}$$

$$\begin{aligned} E[W^2|\text{busy}] &= \frac{2g^{(3)}b^2}{(2-\rho)(3-2\rho)g} + \frac{g^{(2)}[(6-\rho)b^{(2)} + 2\lambda g^{(2)}b^3]}{(2-\rho)^2(3-2\rho)g} \\ &+ \frac{[(6-\rho)\lambda g^2 b^{(2)} + (6+5\rho-2\rho^2)g^{(2)}b]b^{(2)}}{(2-\rho)^2(3-2\rho)gb} + \frac{2(3+\rho+\rho^2)b^{(3)}}{3(2-\rho)(3-2\rho)b} \\ &+ \left(\frac{\lambda^2 g^2 b^{(2)}}{2(1-\rho)} + \frac{g^{(2)}}{2g(1-\rho)} + \frac{\lambda g E[V^2]}{2E[V]}\right)\left(\frac{b^{(2)}}{2-\rho} + \frac{(6-\rho)b^{(2)} + 2\lambda g^{(2)} b^3}{(2-\rho)^2(3-2\rho)}\right) \\ &+ \frac{2b^2}{(2-\rho)(3-2\rho)}\left(\frac{\lambda^2 g^2 E[V^3]}{3E[V]} + \frac{\lambda^3 g^3 b^{(2)} E[V^2]}{2(1-\rho)E[V]} + \frac{(2-\rho)\lambda g^{(2)} E[V^2]}{2(1-\rho)E[V]} + \frac{\lambda^3 g^3 b^{(3)}}{3(1-\rho)}\right. \\ &\left. + \frac{(\lambda^2 g^2 b^{(2)})^2}{2(1-\rho)^2} + \frac{\lambda (g^{(2)})^2 b}{2g(1-\rho)^2} + \frac{g^{(3)}}{3g(1-\rho)} + \frac{(3-\rho)\lambda^2 g g^{(2)} b^{(2)}}{2(1-\rho)^2}\right) \\ &+ \frac{2\rho^2 g^{(2)} b^{(2)}}{g(1-\rho)(2-\rho)(3-2\rho)} + \frac{2(g^{(2)})^2 b^2}{g^2(1-\rho)(2-\rho)(3-2\rho)} + \frac{2\rho g^{(2)} b E[V^2]}{g(2-\rho)(3-2\rho)E[V]} \\ &+ \frac{2\rho b^{(2)}}{(2-\rho)(3-2\rho)}\left(\frac{\lambda^2 g^2 b^{(b)}}{1-\rho} + \frac{g^{(2)}}{g(1-\rho)} + \frac{\lambda g E[V^2]}{E[V]}\right). \end{aligned} \tag{28b}$$

4.3. Unconditional waiting time

Substituting (25) and (28) into (22), we get the first two moments of the waiting time

$$E[W] = \frac{\lambda g b^{(2)}}{2(1-\rho)} + \frac{g^{(2)} b}{2(1-\rho)g} + \frac{E[V^2]}{2E[V]}, \tag{29a}$$

$$E[W^2] = \frac{2\lambda g b^{(3)}}{3(1-\rho)(2-\rho)} + \frac{[\lambda g b^{(2)}]^2}{(1-\rho)^2(2-\rho)} + \frac{2g^{(3)}b^2}{3(1-\rho)(2-\rho)g}$$
$$+ \frac{(1+\rho)g^{(2)}b^{(2)} + \lambda[g^{(2)}]^2 b^3}{(1-\rho)^2(2-\rho)g} + \frac{[g^{(2)}b + \lambda g^2 b^{(2)}]E[V^2]}{g(1-\rho)(2-\rho)E[V]} + \frac{2E[V^3]}{3(2-\rho)E[V]}. \tag{29b}$$

5. Waiting Time for SV Model

In the SV model, if the server returns from a vacation to find no messages waiting, the system becomes idle. A regenerative point of this system is the epoch at which the system becomes empty and a vacation begins as in the MV model. The generation cycle is again the time interval between two such successive regenerative points. The LST of the CDF and the mean for the length V_c of a regeneration cycle are obtained by extending the result for the single arrival model (sec. 2.2 in Takagi [17] with correction needed) as

$$V_c^*(s) = V^*(s+\lambda)I^*(s)\Theta_g^*(s) + V^*[s+\lambda-\lambda\Theta_g^*(s)] - V^*(s+\lambda), \tag{30a}$$

$$E[V_c] = \frac{V^*(\lambda) + \lambda E[V]}{\lambda(1-\rho)}, \tag{30b}$$

where $\Theta_g^*(s)$ satisfies (18), and $I^*(s)$ is the LST of the length I of an idle period, given by

$$I^*(s) = \frac{\lambda}{s+\lambda}, \qquad E[I] = \frac{1}{\lambda}. \tag{30c}$$

Since a vacation appears exactly once in a regeneration cycle, we have

$$\text{Prob[vacation]} = \frac{E[V]}{E[V_c]} = \frac{(1-\rho)\lambda E[V]}{V^*(\lambda) + \lambda E[V]}. \tag{31a}$$

The system enters an idle period of mean length $1/\lambda$ if no messages arrive during a vacation. Thus we have

$$\text{Prob[idle]} = \frac{V^*(\lambda)/\lambda}{E[V_c]} = \frac{(1-\rho)V^*(\lambda)}{V^*(\lambda) + \lambda E[V]}, \tag{31b}$$

which gives

$$\text{Prob[busy]} = 1 - \text{Prob[vacation]} - \text{Prob[idle]} = \rho. \tag{31c}$$

From (31), we have

$$E[W^i] = \frac{(1-\rho)V^*(\lambda)}{V^*(\lambda) + \lambda E[V]} E[W^i|\text{idle}] + \frac{(1-\rho)\lambda E[V]}{V^*(\lambda) + \lambda E[V]} E[W^i|\text{vacation}] + \rho E[W^i|\text{busy}] \tag{32}$$

for $i = 1, 2, \cdots$.

The conditional waiting time distributions when M arrives to find the server idle or on vacation for the SV model equal those in Section 3.1 and Section 4.1. Thus it remains us to derive the conditional waiting time distribution when M arrives during a busy period. By the same argument as in Section 3.2, we have

$$E[e^{-sW}|\text{busy}] = \int_0^\infty \frac{xdB(x)}{b} \frac{1-e^{-sx}}{sx} \sum_{k=0}^{\infty} \pi_k^{\text{B}}(x) W_k^*(s), \tag{33}$$

where, by using (3c), $\pi_k^{\text{B}}(x)$ for the SV model is given by

$$\sum_{k=0}^{\infty} \pi_k^{\text{B}}(x) z^k = \frac{1-\rho}{g(V^*(\lambda)+\lambda E[V])} \cdot \frac{1-V^*[\lambda-\lambda G(z)]+[1-G(z)]V^*(\lambda)}{B^*[\lambda-\lambda G(z)]-z} \cdot \frac{G^{(1)}(z)}{g} \cdot e^{-\lambda[1-G(z)]x}. \tag{34}$$

The first two moments of (33) are given by

$$E[W|\text{busy}] = \frac{(3-2\rho)g^{(2)}b}{2(1-\rho)(2-\rho)g} + \frac{b^{(2)}}{2(1-\rho)b} + \frac{\rho\lambda E[V^2]}{2(2-\rho)(V^*(\lambda)+\lambda E[V])}, \tag{35a}$$

$$\begin{aligned}
E[W^2|\text{busy}] = & \frac{2g^{(3)}b^2}{(2-\rho)(3-2\rho)g} + \frac{g^{(2)}[(6-\rho)b^{(2)}+2\lambda g^{(2)}b^3]}{(2-\rho)^2(3-2\rho)g} \\
& + \frac{[(6-\rho)\lambda g^2 b^{(2)} + (6+5\rho-2\rho^2)g^{(2)}b]b^{(2)}}{(2-\rho)^2(3-2\rho)gb} + \frac{2(3+\rho+\rho^2)b^{(3)}}{3(2-\rho)(3-2\rho)b} \\
& + \frac{(6-\rho)b^{(2)}+2\lambda g^{(2)}b^3}{(2-\rho)^2(3-2\rho)} \left(\frac{\lambda^2 g^2 b^{(2)}}{2(1-\rho)} + \frac{g^{(2)}}{2g(1-\rho)} + \frac{\lambda^2 g E[V^2]}{2(V^*(\lambda)+\lambda E[V])} \right) \\
& + \frac{2b^2}{(2-\rho)(3-2\rho)} \left(\frac{\lambda^3 g^2 E[V^3]}{3(V^*(\lambda)+\lambda E[V])} + \frac{\lambda^4 g^3 b^{(2)} E[V^2]}{2(1-\rho)(V^*(\lambda)+\lambda E[V])} \right. \\
& \quad + \frac{(4-3\rho)\lambda^2 g^{(2)} E[V^2]}{2(1-\rho)(V^*(\lambda)+\lambda E[V])} + \frac{\lambda^3 g^3 b^{(3)}}{3(1-\rho)} + \frac{(\lambda^2 g^2 b^{(2)})^2}{2(1-\rho)^2} + \frac{(2-\rho)\lambda (g^{(2)})^2 b}{2g(1-\rho)^2} \\
& \quad \left. + \frac{g^{(3)}}{3g(1-\rho)} + \frac{(5-3\rho)\lambda^2 g g^{(2)} b^{(2)}}{2(1-\rho)^2} \right) \\
& + \frac{(3+2\rho)g^{(2)}b^{(2)}}{2g(1-\rho)(2-\rho)(3-2\rho)} + \frac{(3+2\rho)\lambda^2 g^2 (b^{(2)})^2}{2(1-\rho)(2-\rho)(3-2\rho)} \\
& + \frac{(3+2\rho)\lambda^2 g b^{(2)} E[V^2]}{2(2-\rho)(3-2\rho)(V^*(\lambda)+\lambda E[V])}.
\end{aligned} \tag{35b}$$

Substituting (9), (25) and (35) into (32), we obtain

$$E[W] = \frac{\lambda g b^{(2)}}{2(1-\rho)} + \frac{g^{(2)}b}{2(1-\rho)g} + \frac{\lambda E[V^2]}{2(V^*(\lambda)+\lambda E[V])}, \tag{36a}$$

$$\begin{aligned}
E[W^2] = & \frac{2\lambda g b^{(3)}}{3(1-\rho)(2-\rho)} + \frac{[\lambda g b^{(2)}]^2}{(1-\rho)^2(2-\rho)} + \frac{2g^{(3)}b^2}{3(1-\rho)(2-\rho)g} \\
& + \frac{(1+\rho)g^{(2)}b^{(2)} + \lambda[g^{(2)}]^2 b^3}{(1-\rho)^2(2-\rho)g} \\
& + \frac{[g^{(2)}b + \lambda g^2 b^{(2)}]\lambda E[V^2]}{g(1-\rho)(2-\rho)(V^*(\lambda)+\lambda E[V])} + \frac{2\lambda E[V^3]}{3(2-\rho)(V^*(\lambda)+\lambda E[V])}.
\end{aligned} \tag{36b}$$

6. Remarks and Numerical Examples

In this section we make a few remarks on the results in Sections 3 through 5. We also present numerical examples in Figures 3 and 4, which show the mean and the coefficient of variation of the waiting time for each model under ROS and FCFS disciplines as a function of ρ, where we assume that service times follow 3-stage Erlang distribution with mean 0.5, vacation times follow 2-stage Erlang distribution with mean 1.0, and batch sizes follow a geometric distribution with mean 2.

6.1. Comparison between ROS and FCFS

For each model, the mean waiting time under ROS equals that under FCFS; this is obvious from *Little's formula* (Little [14]) and the fact that the queue size distribution is invariant to the order of service. We can also confirm the following relationship on the second moment between the ROS and FCFS systems for each model,

$$E[W^2]_{\text{ROS}} = \frac{2}{2-\rho} E[W^2]_{\text{FCFS}}. \tag{37}$$

This agrees with the result for single arrival models, which was derived originally by Takács [16] and interpreted by Fuhrmann [7] for single arrival M/G/1 queues. We note that Fuhrmann's technique does not apply to batch arrival models. Therefore, the relationship in (37) is first established for batch arrival models in this paper.

6.2. Comparison of systems without vacations and with vacations

From (16), (29) and (36), we see that each moment in the vacation models consists of the corresponding moment in the NV model plus additional terms for each vacation model. Figures 3 and 4 show that as ρ approaches 1, the mean and the coefficient of variation of the waiting time distribution for the vacation models approach those of the NV model. This is because the probability that M arrives to find the server on a vacation gets smaller.

On the other hand, as ρ approaches 0, the mean and the coefficient of variation of the waiting time distribution for the SV model approach those of the NV model, because the probability that the server is idle becomes equally large for both models.

6.3. Limiting values of the coefficient of variation

As ρ gets close to 1, the coefficient of variation of the waiting time distributions under ROS becomes $\sqrt{3}$, while that under FCFS becomes 1. As ρ approaches 0, the coefficients of variation of the waiting time distributions for the NV and SV models converge to

$$\frac{\sqrt{6gg^{(2)}b^{(2)} + 4gg^{(3)}b^2 - 3(g^{(2)})^2b^2}}{\sqrt{3}g^{(2)}b},$$

and that for the MV model to

$$\frac{\sqrt{6gg^{(2)}b^{(2)} + 4\left(gg^{(3)}b^2 + \frac{g^2E[V^3]}{E[V]}\right) - 3\left((g^{(2)})^2b^2 + \left(\frac{gE[V^2]}{E[V]}\right)^2\right)}}{\sqrt{3}\left(g^{(2)}b + \frac{gE[V^2]}{E[V]}\right)}.$$

Note that the two expressions agree when $\frac{E[V^3]}{E[V]} \to 0$ and $\frac{E[V^2]}{E[V]} \to 0$.

Appendix A. Derivation of $E[W_k]$ and $E[W_k^2]$ in (8)

Taking the first derivative of (6) at $s = 0$, we get

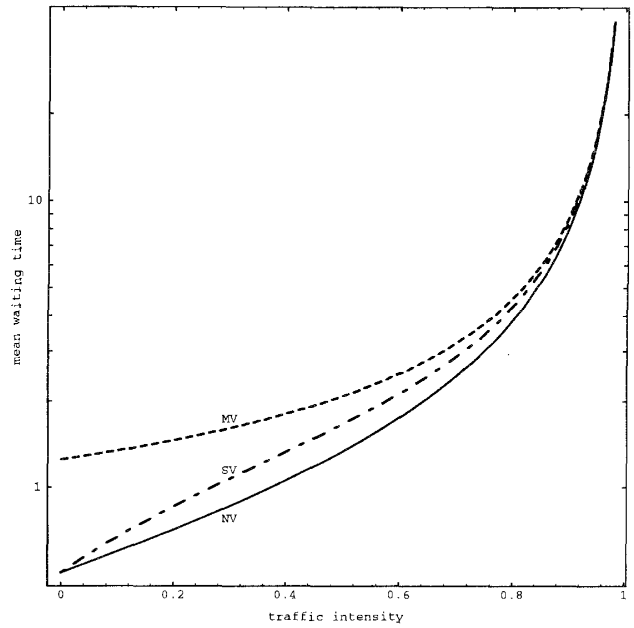

Figure 3: The mean waiting time.

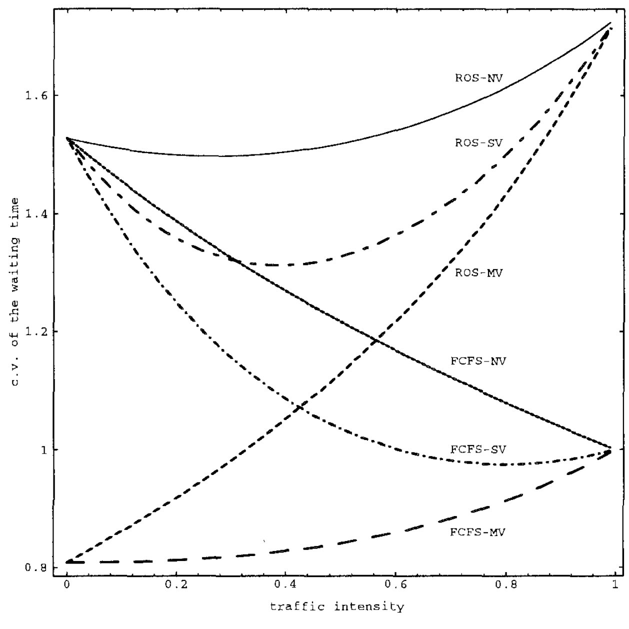

Figure 4: The coefficient of variation of the waiting time.

$$E[W_k] = \frac{k}{k+1}b + \frac{k}{k+1}\sum_{j=0}^{\infty} B_j^*(0)E[W_{k+j-1}]. \tag{38}$$

On the other hand, from (7) we have

$$\sum_{j=1}^{\infty} jB_j^*(0) = \frac{d}{dz}B^*[\lambda - \lambda G(z)]\Big|_{z=1} = \rho, \tag{39a}$$

$$\sum_{j=2}^{\infty} j(j-1)B_j^*(0) = \frac{d^2}{dz^2}B^*[\lambda - \lambda G(z)]\Big|_{z=1} = \lambda^2 g^2 b^{(2)} + \lambda g^{(2)} b, \tag{39b}$$

$$\sum_{j=1}^{\infty} jB_j^{(1)}(0) = \frac{d}{dz}B^{(1)}[\lambda - \lambda G(z)]\Big|_{z=1} = \lambda g b^{(2)}, \tag{39c}$$

where

$$B_j^{(1)}(s) \equiv \frac{d}{ds}B_j^*(s).$$

If we assume the form $E[W_k] = \alpha k$ where α is a constant, by substituting into (38) we get

$$\alpha k = \frac{k}{k+1}\left(b + \sum_{j=0}^{\infty} B_j^*(0)(k+j-1)\alpha\right). \tag{40}$$

Substituting (39) into (40) and manipulating, we obtain (8a). Similarly we can derive $E[W_k^2]$ by taking the second derivative of (6) at $s = 0$,

$$E[W_k^2] = \frac{k}{k+1}\left(b^{(2)} - 2\sum_{j=0}^{\infty} B_j^{(1)}(0)E[W_{k+j-1}] + \sum_{j=0}^{\infty} B_j^*(0)E[W_{k+j-1}^2]\right). \tag{41}$$

If we assume the form $E[W_k^2] = \beta k(k-1) + \gamma k$ where β and γ are constants, from (41) we get

$$\begin{aligned}\beta k(k-1) + \gamma k &= \frac{k}{k+1}\Bigg(b^{(2)} - 2\sum_{j=0}^{\infty} B_j^{(1)}(0)\frac{(k+j-1)b}{2-\rho} \\ &\quad + \sum_{j=0}^{\infty} B_j^*(0)[(k+j-1)(k+j-2)\beta + (k+j-1)\gamma]\Bigg). \end{aligned}\tag{42}$$

Substituting (39) into (42) and manipulating, we obtain

$$\begin{aligned}3(k-1)\beta + 2\gamma &= (k-1)\left(\frac{2b^2}{2-\rho} + 2\rho\beta\right) + \frac{(2+\rho)b^{(2)}}{2-\rho} \\ &\quad + (\lambda^2 g^2 b^{(2)} + \lambda g^{(2)} b)\beta + \rho\gamma. \end{aligned}\tag{43}$$

Since (43) is an identity with respect to k, solving for β and γ yields (8b).

References

[1] Y. Baba: On the $M^X/G/1$ queue with vacation time. *Operations Research Letters*, **5** (1986) 93–98.

[2] Y. Baba: On the $M^X/G/1$ queue with and without vacation time under nonpreemptive last-come first-served discipline. *Journal of the Operations Research Society of Japan*, **30** (1987) 150–159.

[3] P. J. Burke: Delays in single-server queues with batch input. *Operations Research*, **23** (1975) 830–833.

[4] B. Conolly: *Lecture Notes on Queueing Systems* (Ellis Horwood Limited, Sussex, England, 1975).

[5] R. B. Cooper: *Introduction to Queueing Theory*, Second Edition (North-Holland Publishing Company, New York, 1981) (First Edition: Macmillan, New York, 1972), re-published by the Continuing Engineering Education Program, The George Washington University, Washington D.C., 1990.

[6] B. T. Doshi: Queueing systems with vacations - a survey. *Queueing Systems*, **1** (1986) 29–66.

[7] S. W. Fuhrmann: Second moment relationships for waiting times in queueing systems with Poisson input. *Queueing Systems*, **8** (1991) 397–406.

[8] D. P. Heyman M. J. Sobel: *Stochastic Models in Operations Research, Volume I: Stochastic Processes and Operating Characteristics* (McGraw-Hill Publishing Company, New York, 1982).

[9] J. F. C. Kingman: On queues in which customers are served in random order. *Proceedings of the Cambridge Philosophical Society*, **58** (1962) 79–91.

[10] L. Kleinrock: *Queueing Systems, Volume 1: Theory* (John Wiley and Sons, New York, 1975).

[11] L. Kleinrock: *Queueing Systems, Volume 2: Computer Applications* (John Wiley and Sons, New York, 1976).

[12] L. Kleinrock and R. Gail: *Queueing Systems: Problems and Solutions* (John Wiley and Sons, New York, 1996).

[13] Y. Levy and U. Yechiali: Utilization of idle time in an M/G/1 queueing system. *Management Science*, **22** (1975) 202–211.

[14] J. D. C. Little: A proof for the queuing formula: $L = \lambda W$. *Operations Research*, **9** (1961) 383–387.

[15] M. Scholl and L. Kleinrock: On the M/G/1 queue with rest periods and certain service-independent queueing disciplines. *Operations Research*, **31** (1983) 705–719.

[16] L. Takács: Delay distributions for one line with Poisson input, general holding times, and various orders of service. *The Bell System Technical Journal*, **42** (1963) 487–503.

[17] H. Takagi: *Queueing Analysis, A Foundation of Performance Evaluation, Volume 1: Vacation and Priority Systems, Part 1* (Elsevier, Amsterdam, 1991).

[18] H. Takagi and S. Kudoh: Symbolic higher-order moments of the waiting time in an M/G/1 queue with random order of service. *Communications in Statistics-Stochastic Models*, **13** (1997) 167–179.

Hideaki Takagi
Institute of Policy and Planning Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba-shi
Ibaraki 305-8573, Japan
E-mail: `takagi@shako.sk.tsukuba.ac.jp`