A Single-Server Queueing System with Modified Service Mechanism: An Application of the Diffusion Process to the System Performance Measure Formulas

by

Yoshitaka Takahashi*

1. Introduction

We consider a single-server GI/G/1 queueing system with modified service mechanism. By modified service mechanism, we mean the customers who initiate a busy period may have a service time distribution different from that of the customers who do not initiate a busy period, i.e., the service time distribution (H_0) for the customers arriving to find the system idle may be different from the service time distribution (H_1) for the customers arriving to find the system busy.

A queueing system with modified service mechanism will be referred as modified service system in short. A (standard) queueing system is a special case of the modified service system, since the standard system immediately follows if we set $H_0 = H_1$ in the modified system. A set-up time queueing system or sometimes called as a warm-up time queueing system can be regarded as a special case of the modified service system. A vacation queuing system can be also regarded as a special case.

A modified service system seems to be useful in some practical situations. A typical

* Yoshitaka Takahashi is Professor of Informatics and Management Science at Faculty of Commerce, Waseda University, where he has taught since 2000. Prior to that, he was Senior Member of Technical Staff, Supervisor at NTT Laboratories. He obtained his Ph.D. degree in Computer Science from the Tokyo Insitute of Technology in 1990.

situation arises if provision has to be made for the warming up of the servicing machinery; see Bhat [1]. In the computer-communication field, we encounter another typical situation, where a single server works on primary and secondary customers and our focus is on the primary customers' queueing behavior; see e.g., Doshi [2], Kreinin [4], Niu et al. [7], and Takagi [12, 13].

There has been much interest in studying modified service systems. Bhat [1], Welch [16], and Yeo [17] have originally investigated a modified M/G/1 system, generalizing the results for the standard M/G/1 system. Pakes [8, 9] has treated a modified GI/M/1 system and a modified GI/G/1 system to derive the number of customers served during the busy period; see also Lemoine [5] and Minh [6] for limit theorems in the modified GI/G/1 system.

Since even the mean performance measures (e.g., the mean waiting time) cannot be easily obtained by using these previous results [4, 5, 6, 9], approximate approaches and techniques are important for analyzing the modified GI/G/1 queueing system performance. To the author's knowledge, however, there exists no literature on approximations for the mean performance measures. The main purpose of this paper is to propose an approximation on the mean performance measures in the modified GI/G/1 queueing system.

The rest of the paper is organized as follows. Section 2 describes a modified GI/G/1 queueing system and gives the notation. Section 3 is devoted to studying qualitative relationships among the performance measures in the system. Section 4 develops the diffusion process approximation for the virtual waiting time in the system. We propose a new mean (actual) waiting-time approximation formula through the qualitative relationships in Section 3 and the diffusion approximation. Section 5 considers special examples and presents comparisons with exact and simulation results, confirming the accuracy of the proposed approximation.

2. Modified Service System

We consider a stochastic service system, assuming the followings.

i) Customers arrive independently each other at a single server queueing system.

ii) The inter-arrival time of the customers is an independent, and identically distributed (iid) random variable (rv), A. The arrival rate is denoted by λ .

iii) If a customer arrives and finds the server idle, its service time is an iid rv, H_0; while if a customer arrives and finds the server busy, its service time is an iid rv, H_1.

iv) The capacity of the waiting room (queueing capacity) is infinite.

We denote the n-th moment of an rv, say X, by $x^{(n)} = E(X^n)$. We also use $x = E(X)$ instead of $x^{(1)} = E(X^1)$. We denote the coefficient of variation (cv) of rv X by cx, i.e.,

$$c_X^2 = \frac{(E(X^2) - E(X)^2)}{E(X)^2} = \frac{(x^{(2)} - x^2)}{x^2} \tag{1}$$

For example, the arrival rate is now given by $\lambda = 1 / E(A)$. The squared cv of the service time for the customers who finds the server busy is given by

$$c^2_{H_1} = \frac{(h_1^{(2)} - h_1^2)}{h_1^2}$$

We further assume that our modified GI/G/1 queueing system is stationary, so that

$$\lambda h_1 < 1, \text{ and } \lambda(h_1 - h_0) < 1 \tag{2}$$

The latter condition in (2) will be verified through the subsequent sections.

3. Qualitative Results

We use the following symbols in the subsequent analysis.

B: busy period

I: idle period

p_0: idle probability, i.e., the probability that the system is idle

L: the number of customers in the system (including server)

L_q: the number of customers in the queue

V: virtual waiting time

W: (actual) waiting time including service time (system sojourn time)

W_q: (actual) waiting time in the queue

$W_q^c(0)$: probability of delay, i.e., the probability that an arriving customer has to wait

$W_q(0)$: the probability that the waiting time in the queue of an arriving customer is zero

We then have

$$W_q^c(0) = P_0(\, W_q > 0\,),$$

and

$$W_q(0) = P_0(\, W_q = 0\,),$$

where P_0 is the Palm distribution with respect to the arrival point process; see Kawashima et al. [3]. It follows that

$$W_q^c(0) + W_q(0) = 1 \tag{3}$$

We denote by E_0 the expectation with respect to the Palm distribution, P_0; while we denote by E the (ordinary) expectation with respect to the probability measure P under which our modified GI/G/1 queueing system is stationary. It should be noted that the (actual) waiting time sequence is stationary under P_0 (not under P), while the queue length process and the virtual waiting time process are stationary under P.

Applying the level-crossing argument by Rice [10] to our modified GI/G/1 queueing system, we have

$$E(B)p_0 = (\, 1 - p_0\,)E(I) \tag{4}$$

Denoting by p_n the stationary probability that there are n customers in the system, we have

$$E(L) = \Sigma_{n>0}\, np_n, \text{ and } E(L_q) = \Sigma_{n>0}\,(n - 1)p_n\,, \tag{5}$$

yielding

$$E(L) - E(L_q) = 1 - p_0\,. \tag{6}$$

Similarly, we have the mean waiting time relationship as

$$E_0(W) - E_0(W_q) = W_q{}^c(0)h_1 + W_q(0)h_0. \tag{7}$$

Applying Little's law which relates E and E_0,

$$E(L) = \lambda E_0(W), \tag{8a}$$

and

$$E(L_q) = \lambda E_0(W_q). \tag{8b}$$

From (6) together with (3), (7), and (8a),(8b), we obtain

$$p_0 = 1 - \lambda h_0 - \lambda W_q{}^c(0)\,(h_1 - h_0) \tag{9}$$

The rate conservation law approach [3] is now applied to find

$$E(V) = o_0 \frac{h_0{}^{(2)}}{2h_0} + o_1 \frac{h_1{}^{(2)}}{2h_1} + \lambda h_1 E_0(W_q), \tag{10}$$

where o_0 is the probability that a test customer arrives at the system and finds the customer in service initiated a busy period, and o_1 is the probability that a test customer arrives at the system and finds the customer in service did not initiate a busy period, i.e.,

$$o_0 = \frac{h_0}{E(B) + E(I)} \tag{11a}$$

and

$$o_1 = \frac{E(B) - h_0}{E(B) + E(I)} \tag{11b}$$

4. Proposed Approximation

We are finally in a position to propose an approximation for the mean performance measures in the modified GI/G/1 queueing system.

We approximate the virtual waiting time process by a diffusion process with elementary return boundary at x = 0. As in Takahashi [14], solving the time-stationary diffusion equation, we have the idle probability:

$$p_0 = \frac{1 - \lambda h_1}{1 - \lambda(h_1 - h_0)} \tag{12}$$

Substituting (12) into (9), we obtain an approximate formula on the probability of delay:

$$W_q^c(0) = \frac{\lambda h_0}{1 - \lambda(h_1 - h_0)} \tag{13}$$

The mean virtual waiting time approximate formula is then

$$E(V) = \frac{\lambda}{2[1 - \lambda(h_1 - h_0)]} \cdot \frac{h_0^{(2)} + \lambda h_0 h_1^2 (C_A^2 + C_{H1}^2)}{(1 - \lambda h1)} \tag{14}$$

The probability that a test customer arrives at the system and finds the customer in service receives H_0 [or H_1] service time, o_0 [or o_1] is respectively approximated as

$$o_0 = \frac{\lambda h_0(1 - \lambda h_1)}{1 - \lambda(h_1 - h_0)} \tag{15a}$$

and

$$o_1 = \frac{\lambda^2 h_0 h_1}{1 - \lambda(h_1 - h_0)} \tag{15b}$$

Approximation (15) uses the fact that test customer arrivals see time averages just like Poisson input. The mean waiting time [$E_0(W_q)$] approximation is then obtained from (10) together with (11a), (11b), (14), (15a), and (15b).

5. Special Queueing Models

If we assume a Poisson input ($C_A = 1$) system with modified service mechanism, our proposed approximation on the mean waiting time is reduced to:

$$E_0(W_q) = \frac{\lambda(1-\lambda h_1)h_0^{(2)} + \lambda^2 h_0 h_1^{(2)}}{2[1-\lambda(h_1-h_0)](1-\lambda h_1)} \tag{16}$$

Equation (16) is seen to be consistent with the exact formula for the modified M/G/1 queuing system obtained by Welch [16] and Yeo [17].

If we assume the standard service ($H_0 = H_1 = H$) mechanism, our proposed approximation on the mean waiting time is reduced to:

$$E_0(W_q) = \frac{\lambda h^2 (C_A^2 + C_H^2)}{2(1-\lambda h)} \tag{17}$$

Equation (17) is now seen to be consistent with the approximate formula for the standard GI/GI/1 queueing system obtained by Sakasegawa [11] and Yu [18].

For the two-stage Erlang input and deterministic service ($C_A^2 = 0.5$, $C_H = 0$; $E_2/D/1$) system with modified service mechanism, Figure 1 shows how our approximation performs well for three different parameter patterns ($h_0 = 4$, $h_1 = 1$; $h_0 = 3$, $h_1 = 1.5$; $h_0 = 1$, $h_1 = 1.5$) by comparing our approximation results with simulation results.

6. Concluding Remarks

In the computer-communication field, we frequently encounter a situation in which a single server (processor) works on primary and secondary customers. If our focus is on the primary (delay-sensitive) customers' queueing behavior, this situation leads to a GI/G/1 queueing system with modified service mechanism, where the service time distribution (H_0) for the customers arriving to find the system idle may be different from the service time distribution (H_1) for the customers arriving to find the system busy. We have presented the qualitative relationships among the performance measures in the system. Applying the diffusion process approximation for the virtual waiting time together with these qualitative results, we have proposed a new approximate formula for the mean performance measures. For special cases, our approximation is seen to be consistent with the previously-obtained exact results [16, 17] for the M/G/1 queueing system with modified service mechanism, and it is further seen to be consistent with the previously-proposed approximate results [11, 18] for the GI/GI/1 queueing system with standard ($H_0 = H_1$) service mechanism. It is left for future work to treat a finite-capacity queueing system, which requires another boundary condition when applying the diffusion process as in Takahashi [15].

References

[1] U. N. Bhat, "On the busy period of a single server bulk queue with a modified service mechanism," *Calcutta Statist. Assoc. Bull.*, 13, pp. 163-171 (1964).

[2] B. T. Doshi, "Queueing systems with vacations - A survey," *Queueing Systems*, 1, pp. 29-66 (1986).

[3] K. Kawashima, F. Machihara, Y. Takahashi, and H. Saito, "*The Basis of Teletraffic Theory and Multimedia Network*," IEICE Ed., Corona, Tokyo (1995).

[4] A. Ya Kreinin, "Single-channel queueing system with warm up," *Automation & Remote Control*, 41, 6, pp. 771-776 (1980).

[5] A. J. Lemoine, "Limit theorems for generalized single server queues: The exceptional system," *SIAM J. Appl. Math.*, 28, pp. 596-606 (1975).

[6] D. L. Minh, "Analysis of the exceptional queueing system by the use of regenerative process and analytical methods," *Math. Operations Res.*, 5, pp. 147-159 (1990).

[7] Z. Niu and Y. Takahashi, "A finite capacity queue with exhaustive vacation / close-down / setup times and Markovian arrival processes," *Queueing Systems*, 31, pp. 1-23 (1999).

[8] A. G. Pakes, "A GI/M/1 queue with a modified service mechanism," *Ann. Inst. Statist. Math.*, 24, pp. 589-597 (1972).

[9] A. G. Pakes, "On the busy period of the modified GI/G/1 queue," *J. Appl. Probability*, 10, pp. 192-197 (1973).

[10] S. O. Rice, "Single server systems - I. Relations between some averages," *Bell Systems Tech. J.*, 41, pp. 269-278 (1962).

[11] H. Sakasegawa, "An approximation formula $L_q \sim \alpha\rho^{\beta} / (1 - \rho)$," *Ann. Inst. Statist. Math.*, 29A, pp. 67-75 (1977).

[12] H. Takagi (Ed), "*Stochastic Analysis of Computer and Communication Systems,*" North-Holland, Amsterdam (1990).

[13] H. Takagi, "*Queueing Analysis, Volume 1: Vacation and Priority Systems, Part 1,*" North-Holland, Amsterdam (1991).

[14] Y. Takahashi, "Diffusion approximation for the single server system with batch arrivals of multi-class calls," *IECE Trans.*, J69-A, 3, pp. 317-324 (1986).

[15] Y. Takahashi, "A branching Poisson process input finite-capacity queueing system for telecommunication networks," *The Waseda Commercial Review*, No. 391, pp. 491-501 (2001).

[16] P. D. Welch, "On a generalized M/G/1 queueing process in which the first customer of each busy period receives exceptional service," *Operations Res.*, 12, pp. 736-752 (1964).

[17] G. F. Yeo, "Single server queues with modified service mechanisms," *J. Austral. Math. Soc.*, 3, pp. 499-507 (1962).

[18] P. S. Yu, "On accuracy improvement and applicability conditions of diffusion approximation with application to modeling of computer systems," TR-129, *Digital Systems Lab., Stanford Univ.* (1977).

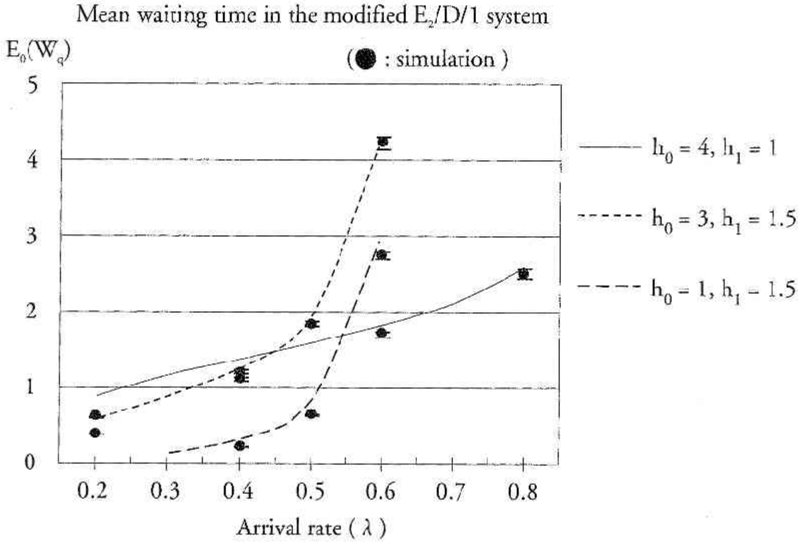

Figure 1: A performance comparison between our approximation and simulations.

Acknowledgement

The present research was partially supported by a Grant-in-Aid for Scientific Research (C) from Japan Society for the Promotion of Science under Grant No.1458049.