Optimal Timing Strategies for Controlled $M/G/1$ Queueing System via Diffusion Approximation Approach

H. Okamura, T. Dohi and S. Osaki

Department of Industrial and Systems Engineering, Hiroshima University

Faculty of Engineering,

4-1 Kagamiyama 1 Chome, Higashi-Hiroshima 739, JAPAN

dohi@gal.sys.hiroshima-u.ac.jp

Abstract

In this paper, we consider the optimal timing strategies for a controlled single-server $M/G/1$ queueing system. The expected total discounted costs are formulated under two timing strategies and approximated by means of the diffusion approximation method. Numerical calculations are devoted to examine the accuracy of the approximation method. The results show that the diffusion approximation method is reasonably good for the typical queueing systems, especially, the $M/M/1$ single-server queue.

1 Introduction

Queueing control is one of the most important problems in the research field of Operations Research/Management Science. Especially, the optimal control policy for the $M/G/1$ queueing system can be applicable to many actual situations such as the production/inventory control, the telecommunication process control, computer science and so on. Heyman (1968), Sobel (1969) and Bell (1971) studied the optimal N -policy which activates the server when there are N customers waiting for service and deactivates the server when there are no customers in the system. In particular, Bell (1971) investigated the N -policy under the expected total discounted cost criterion, which is discounted continuously by a positive discount rate. However, it is noted that the exact expression of the expected total discounted cost needs the Laplace-Stieltjes transform of the service distribution function for the $M/G/1$ queue. Generally, the service time can not be often identified in many practical applications.

For this restricted problem, Kimura (1981) provided a robust approximation model for determining the optimal N -policy without assuming complete information on the service distribution function. In other words, assuming the existence of the first two moments on the service distribution, he applied the technique of diffusion approximation to express the expected total discounted cost for the queueing system over an infinite time horizon, which was developed by Kobayashi (1974). See, for

the nice survey in the research area of queueing control and its relation with vacation models, Teghem (1986) and Takagi (1991), respectively.

In this paper, we consider a situation where the server cannot continuously monitor the queue for an arrival when the server is not active. Such a situation may occur in the production/inventory system without a sensing device. It is assumed that the server scans the queue $T (> 0)$ time units after the end of the last busy period to determine if customers are present. To this end, if customers are found, a busy period begins and the server is kept in the active state until the system empty. On the other hand, if no customers are found, a busy period of length zero occurs. In either case, the next scan is made T units after the end of a busy period. Heyman (1977) called this plausible operation policy the T -policy. As proven in Heyman (1977), however, notice that the long-run average cost (but with no discounting) under the optimal T -policy is larger than that achieved by the optimal N -policy. This might motivate the queueing operation without the T -policy. But, as described previously, the assumption on continuous monitoring may be seen to be strong in many practical applications.

The main purpose of this paper is to obtain a diffusion approximation model for the $M/G/1$ queueing system with a removable server under the T -policy. First, we derive an approximation formula of the expected total discounted cost under the T -policy, using the similar technique with Kimura (1981). Here, we also give an exact expression of the expected total discounted cost with the Laplace-Stieltjes transform of the service distribution function, which was not derived by Heyman (1977). Next, we consider a combined operation policy consisted of both the T -policy and the N -policy. The results on the stopped Brownian motion process (see, *e.g.* Dohi, Kaio and Osaki (1995)) are directly applied to express the expected cost function. Finally, in numerical examples, we compare the approximation models with the exact ones and refer to the accuracy of the approximation.

2 A Timing Strategy for $M/G/1$ Queue

We consider an $M/G/1$ queueing system with a removable server. It is assumed that customers k ($k = 1, 2, \dots$) arrive according to a Poisson process $\{N(t), t \geq 0\}$ with $N(0) = 0$ and parameter $\lambda (> 0)$, and their service times S_k are nonnegative i.i.d. random variables having a distribution function $B(t)$ with finite mean $1/\mu (> 0)$ and finite variance $\sigma^2 (> 0)$. Define the Laplace-Stieltjes transform (L.S.T.) of the distribution function $B(t)$ as $\tilde{B}(\beta)$, where β is a complex number. Without loss of generality, the heavy traffic condition is assumed, i.e. the traffic intensity $\rho = \lambda/\mu$ is less than unity, and the number of customers in the system is zero at time $t = 0$.

Following Heyman (1977), we consider the T -policy. Under this policy, the server is kept to be off from $t = 0$ to time $t = T (\geq 0)$ and customers arriving at the system during this period are cumulated in the buffer. At time T , the server is turned on, the service operation is started and continues until the customers vanish in the system (see Fig. 1). In this paper, we call this kind of control policy the *timing strategy*. Since the arrival process is the Poisson, the system is regenerative at the end of busy period. To this end, the queueing process under the T -policy is an ordinary renewal process, and one cycle is defined as the time interval from the end of busy period to the next one. This may be a special case for the queueing system with vacation periods. For more details, see Takagi (1991).

Fig. 1: Behavior of the queueing system under the T policy.

Define the length of busy period τ which is a random variable and the number of customers serviced during $[0, t)$ after starting service:

$$R(t) = \sup\{n : \sum_{k=1}^n S_k \leq t\}. \quad (1)$$

Also, we define

$$Y(t; x) = x + N(t) - R(t). \quad (2)$$

Then the number of customers in the system at time t , $\{Z(t), t \geq 0\}$ is the stochastic process as follows.

$$Z(t) = \begin{cases} N(t), & (0 \leq t < T) \\ Y(t - T; N(T)), & (T \leq t < T + \tau) \\ Z(t - T - \tau), & (T + \tau \leq t). \end{cases} \quad (3)$$

From Eq.(3), we represent the length of busy period:

$$\tau = \inf\{t > 0; Y(t; N(T)) \leq 0\}. \quad (4)$$

Next, consider the cost component. There are operating costs r_1 and r_2 ($r_2 \geq r_1 > 0$) for keeping the server off and on per unit time, switching costs $R_1 (> 0)$ and $R_2 (> 0)$ for turning the server on and turning it off, and the holding cost per unit time proportional to the number of customers in the system $h (> 0)$, respectively. Heyman (1977) dealt with the long-run average cost under the same model setting. In this respect, we assume that all costs are continuously discounted with an interest rate $\beta (> 0)$, i.e. one dollar paid at time $t (> 0)$ has a present value of $\exp(-\beta t)$. In our queueing system, the present value of a unit cost just after one cycle is

$$\delta_1(T) = E[e^{-\beta(T+\tau)}] = e^{-\beta T} E[e^{-\beta \tau}], \quad (5)$$

where E is the mathematical expectation operator. Thus, the expected total discounted cost for one cycle is

$$\begin{aligned} C_1(T) &= E[R_1 e^{-\beta T} + R_2 e^{-\beta(T+\tau)} \\ &\quad + \int_0^T r_1 e^{-\beta t} dt + \int_T^{T+\tau} r_2 e^{-\beta t} dt \\ &\quad + \int_0^{T+\tau} e^{-\beta t} [hZ(t)] dt] \\ &= e^{-\beta T} (R_1 + R_2 E[e^{-\beta \tau}]) \\ &\quad + \frac{r_1}{\beta} + e^{-\beta T} \left(\frac{r_2 - r_1}{\beta} - \frac{r_2}{\beta} E[e^{-\beta \tau}] \right) \\ &\quad + hE \left[\int_0^T e^{-\beta t} N(t) dt \right] \\ &\quad + h e^{-\beta T} E \left[\int_0^{\tau} e^{-\beta t} Y(t; N(T)) dt \right]. \end{aligned} \quad (6)$$

In order to obtain the explicit expression in Eq.(6), we suppose $\int_0^\infty e^{-\beta t} E[|N(t)|] dt < \infty$ and $\int_0^\infty e^{-\beta t} E[|Y(t; N(T))|] dt < \infty$. Using the first Dynkin formula (e.g. Karlin and Taylor (1981), p. 297), we have

$$\begin{aligned} &E \left[\int_0^T e^{-\beta t} N(t) dt \right] \\ &= \int_0^\infty e^{-\beta t} E[N(t)] dt \\ &\quad - e^{-\beta T} \int_0^\infty e^{-\beta t} E[N(t) | N(0) = N(T)] dt \\ &= \int_0^\infty e^{-\beta t} E[N(t)] dt \\ &\quad - e^{-\beta T} \int_0^\infty e^{-\beta t} E[N(T) + N(t)] dt \end{aligned} \quad (7)$$

and

$$E \left[\int_0^\tau e^{-\beta t} Y(t; N(T)) dt \right]$$

$$\begin{aligned}
&= \int_0^\infty e^{-\beta t} \mathbb{E}[Y(t; N(T))] dt \\
&\quad - \mathbb{E}[e^{-\beta \tau}] \int_0^\infty e^{-\beta t} \mathbb{E}[Y(t; 0)] dt \\
&= \int_0^\infty e^{-\beta t} \{ \mathbb{E}[N(T) + N(t)] - \mathbb{E}[R(t)] \} dt \\
&\quad - \mathbb{E}[e^{-\beta \tau}] \int_0^\infty e^{-\beta t} \{ \mathbb{E}[N(t)] - \mathbb{E}[R(t)] \} dt. \quad (8)
\end{aligned}$$

Notice that the following relationship is satisfied from the standard results on the $M/G/1$ queueing system:

$$\int_0^\infty e^{-\beta t} \mathbb{E}[N(t)] dt = \frac{\lambda}{\beta^2} \quad (9)$$

and

$$\int_0^\infty e^{-\beta t} \mathbb{E}[R(t)] dt = \frac{\tilde{B}(\beta)}{\beta[1 - \tilde{B}(\beta)]}. \quad (10)$$

From Eqs.(9) and (10), we obtain the expected total discounted cost for one cycle as follows.

$$\begin{aligned}
C_1(T) &= \frac{r_1}{\beta} + \frac{h\lambda}{\beta^2} + e^{-\beta T} \left[R_1 + \frac{r_2 - r_1}{\beta} - \frac{h\nu(\beta)}{\beta^2} \right. \\
&\quad \left. + \mathbb{E}[e^{-\beta \tau}] \left(R_2 - \frac{r_2}{\beta} - \frac{h(\lambda - \nu(\beta))}{\beta^2} \right) \right], \quad (11)
\end{aligned}$$

where

$$\nu(\beta) = \frac{\beta \tilde{B}(\beta)}{1 - \tilde{B}(\beta)}. \quad (12)$$

Now, let $\tilde{G}(\beta)$ denote the L.S.T. of the time interval at which a customer leaves from the $M/G/1$ queue. That is, we have

$$\tilde{G}(\beta) = \tilde{B}(\beta + \lambda - \lambda \tilde{G}(\beta)). \quad (13)$$

From Eq.(13), the L.S.T. of the length of busy period τ is

$$\mathbb{E}[e^{-\beta \tau}] = \sum_{n=0}^{\infty} e^{-\lambda T} \frac{(\lambda T)^n}{n!} \tilde{G}(\beta)^n = e^{-\lambda(1 - \tilde{G}(\beta))T}. \quad (14)$$

Hence, the expected total discounted cost over an infinite time span is

$$TC_1(T) = \sum_{n=0}^{\infty} C_1(T) \delta_1(T)^n = \frac{C_1(T)}{\bar{\delta}_1(T)}, \quad (15)$$

where, in general, $\bar{\delta}_1(\cdot) = 1 - \delta(\cdot)$.

3 Alternative Timing Strategy

Combining the timing strategy in Section 2 with the so-called N policy developed by Heyman (1968), we treat an alternative queueing control strategy. If the number of customers attains at a pre-specified level $N(\geq 0)$ within $[0, T)$, the server is turned on, otherwise, is done

at the scheduled time T . We suppose that a fixed switching cost $R_3(> 0)$ is incurred at each instant that $N(t)$ is beyond the level N at an arbitrary time $t \in [0, T)$.

In a fashion similar to the previous section, we define the following stopping times:

$$\tau_1 = \inf\{t > 0; N(t) \geq N\} \quad (16)$$

and

$$\tau_2(x) = \inf\{t > 0; Y(t; x) \leq 0\}. \quad (17)$$

Throughout this paper, we use the notation $\mathbb{E}[X; A]$ as a restricted expectation which means $\mathbb{E}[X I_A]$, where I_A is the indicator function of the set A . With this notation, the present value of a unit cost just after one cycle and the expected total discounted cost for one cycle are

$$\begin{aligned}
\delta_2(T) &= \mathbb{E}\left[e^{-\beta[\tau_1 + \tau_2(N)]}; \{\tau_1 \leq T\}\right] \\
&\quad + \mathbb{E}\left[e^{-\beta[T + \tau_2(N(T))]}; \{\tau_1 > T\}\right] \\
&= \mathbb{E}\left[e^{-\beta \tau_1}; \{\tau_1 \leq T\}\right] \mathbb{E}\left[e^{-\beta \tau_2(N)}\right] \\
&\quad + e^{-\beta T} \mathbb{E}\left[e^{-\beta \tau_2(N(T))}; \{\tau_1 > T\}\right] \quad (18)
\end{aligned}$$

and

$$\begin{aligned}
C_2(T) &= \mathbb{E}\left[R_3 e^{-\beta \tau_1} + R_2 e^{-\beta[\tau_1 + \tau_2(N)]}\right. \\
&\quad + \int_0^{\tau_1} r_1 e^{-\beta t} dt + \int_{\tau_1}^{\tau_1 + \tau_2(N)} r_2 e^{-\beta t} dt \\
&\quad + e^{-\beta \tau_1} \int_0^{\tau_1 + \tau_2(N)} e^{-\beta t} [hY(t; N)] dt \\
&\quad + \int_0^{\tau_1} e^{-\beta t} [hN(t)] dt; \{\tau_1 \leq T\}\right] \\
&\quad + \mathbb{E}\left[R_1 e^{-\beta T} + R_2 e^{-\beta[T + \tau_2(N(T))]} \right. \\
&\quad + \int_0^T r_1 e^{-\beta t} dt + \int_T^{T + \tau_2(N(T))} r_2 e^{-\beta t} dt \\
&\quad + e^{-\beta T} \int_0^{T + \tau_2(N(T))} e^{-\beta t} [hY(t; N(T))] dt \\
&\quad \left. + \int_0^T e^{-\beta t} [hN(t)] dt; \{\tau_1 > T\}\right], \quad (19)
\end{aligned}$$

respectively. Further, we apply the first Dynkin formula to Eq.(19). Consequently, we have

$$\begin{aligned}
C_2(T) &= \frac{r_1}{\beta} + \frac{h\lambda}{\beta^2} \\
&\quad + \mathbb{E}\left[e^{-\beta \tau_1}; \{\tau_1 \leq T\}\right] \left[R_3 + \frac{r_2 - r_1}{\beta} - \frac{h\nu(\beta)}{\beta^2} \right. \\
&\quad \left. + \mathbb{E}\left[e^{-\beta \tau_2(N)}\right] \left(R_2 - \frac{r_2}{\beta} - \frac{h(\lambda - \nu(\beta))}{\beta^2} \right) \right] \\
&\quad + e^{-\beta T} \left[\left(R_1 + \frac{r_2 - r_1}{\beta} - \frac{h\nu(\beta)}{\beta^2} \right) \Pr\{\tau_1 > T\} \right. \\
&\quad \left. + \mathbb{E}\left[e^{-\beta \tau_2(N(T))}; \{\tau_1 > T\}\right] \right]
\end{aligned}$$

$$\times \left(R_2 - \frac{r_2}{\beta} - \frac{h(\lambda - \nu(\beta))}{\beta^2} \right), \quad (20)$$

where

$$E[e^{-\beta\tau_1}; \{\tau_1 \leq T\}] = \int_0^T e^{-(\beta+\lambda)t} \frac{\lambda(\lambda t)^{N-1}}{(N-1)!} dt, \quad (21)$$

$$\Pr\{\tau_1 > T\} = \sum_{k=0}^{N-1} e^{-\lambda T} \frac{[\lambda T]^k}{k!}, \quad (22)$$

$$E[e^{-\beta\tau_2(N)}] = \tilde{G}(\beta)^N \quad (23)$$

and

$$E[e^{-\beta\tau_2(N(T))}; \{\tau_1 > T\}] = \sum_{k=0}^{N-1} e^{-\lambda T} \frac{[\lambda \tilde{G}(\beta) T]^k}{k!}. \quad (24)$$

Finally, the expected total discounted cost over an infinite time span becomes

$$TC_2(T) = C_2(T)/\bar{\delta}_2(T). \quad (25)$$

It may be noted that the expected cost $TC_2(T)$ can be regarded as a function with two decision variables T and N . However, we found through numerical calculations that

$$\min_{T,N} TC_2(T, N) \geq \lim_{T \rightarrow \infty} TC_2(T, N) \quad (26)$$

for integer N , although it has not been proven analytically. In other words, the N policy is optimal under the condition that the number of customers in the queue is observable, even if the expected total discounted cost is adopted. Therefore, we pay no attention to the situation in which the number of customers is monitored and countable continuously, and do not compare the alternative timing strategy with a fixed level N with the ordinary N policy.

In stead of that, we focus on the other problem. As Kimura (1981) pointed out, it is necessary for the actual calculation of the expected total discounted costs $TC_1(T)$ and $TC_2(T)$ to know complete information on the distribution $B(t)$. This implies that the functional in Eq.(13) must be solved. It is, however, impossible to estimate the exact distribution $B(t)$ in many actual situations. Consequently, the expected total discounted costs should be approximated by means of the diffusion approximation method depending only on the first two moments of $B(t)$. In the following section, we apply the diffusion approximation method to the analysis of the queueing control problems under the timing strategies.

4 Diffusion Approximation Methods

First, we approximate the arrival process for the idle period and the $M/G/1$ process for the busy period to the diffusion processes $\{W_i(t); t \geq 0\}$, ($i = 1, 2$), respectively. The infinitesimal means a_i and the diffusion parameters b_i ($i = 1, 2$) for the diffusion processes

$\{W_i(t); t \geq 0\}$ are defined as follows:

$$a_i = \lim_{\Delta t \rightarrow 0} \frac{E[W_i(t + \Delta t) - W_i(t)|W_i(t)]}{\Delta t}, \quad (27)$$

$$b_i = \lim_{\Delta t \rightarrow 0} \frac{\text{Var}[W_i(t + \Delta t) - W_i(t)|W_i(t)]}{\Delta t}. \quad (28)$$

$i = 1, 2.$

Kimura (1981) obtained the approximate expressions: $a_1 = b_1 = \lambda$, $a_2 = \lambda - \mu$, $b_2 = \lambda + \mu^3 \sigma^2$. From these results, we have

$$\int_0^\infty e^{-\beta t} E[N(t)] dt \approx \frac{a_1}{\beta^2} \quad (29)$$

and

$$\int_0^\infty e^{-\beta t} E[N(t) - R(t)] dt \approx \frac{a_2}{\beta^2}. \quad (30)$$

From the well-known property on the Brownian motion process, it is easily shown

$$\begin{aligned} \Pr\{\tau > t | N(T) = x\} \\ \approx \Phi\left[\frac{x - a_1 T}{\sqrt{b_1 T}}\right] - e^{\frac{2a_1 x}{b_1}} \Phi\left[\frac{-x - a_1 T}{\sqrt{b_1 T}}\right], \end{aligned} \quad (31)$$

where $\Phi[\cdot]$ is the standard normal distribution function. Equation (31) provides the following simple representation:

$$\begin{aligned} E[e^{-\beta\tau}] &\approx \int_{-\infty}^\infty \int_0^\infty e^{-\beta s} p(s, x) q(x, T) ds dx \\ &= e^{-\frac{a_1^2 - \xi^2}{2b_1} T}, \end{aligned} \quad (32)$$

where

$$\begin{aligned} p(t, x) &\equiv \frac{\partial}{\partial t} \Pr\{\tau \leq t | N(T) = x\} \\ &= \frac{|x|}{\sqrt{2\pi b_2 t^3}} e^{-\frac{(x + a_2 t)^2}{2b_2 t}}, \end{aligned} \quad (33)$$

$$q(x, t) = \frac{1}{\sqrt{2\pi b_1 t}} e^{-\frac{(x - a_1 t)^2}{2b_1 t}} \quad (34)$$

and

$$\xi = -a_1 + \frac{b_1}{b_2} \left(a_2 + \sqrt{a_2^2 + 2\beta b_2} \right). \quad (35)$$

Using the results above, we obtain the approximate expressions:

$$\begin{aligned} C_{\text{app1}}(T) &= \frac{r_1}{\beta} + \frac{h a_1}{\beta^2} \\ &\quad + e^{-\beta T} \left[R_1 + \frac{r_2 - r_1}{\beta} - \frac{h(a_1 - a_2)}{\beta^2} \right. \\ &\quad \left. + e^{-\frac{a_1^2 - \xi^2}{2b_1} T} \left(R_2 - \frac{r_2}{\beta} - \frac{h a_2}{\beta^2} \right) \right] \end{aligned} \quad (36)$$

and

$$\delta_{\text{app1}}(T) = e^{-(\beta + \frac{a_1^2 - \xi^2}{2b_1}) T}. \quad (37)$$

Thus, we have

$$TC_1(T) \approx TC_{\text{app1}}(T) = C_{\text{app1}}(T)/\bar{\delta}_{\text{app1}}(T). \quad (38)$$

Next, we develop a diffusion approximation method for an alternative queueing strategy in Section 3. Similarly, Eqs. (29) and (30) are also true for the extended model. On the other hand, the probability distribution functions for the stopping times τ_1 and $\tau_2(x)$ defined in Eqs. (16) and (17) are approximated as

$$\Pr\{\tau_1 > t\} \approx \Phi\left[\frac{N - a_1 T}{\sqrt{b_1 T}}\right] - e^{\frac{2a_1 N}{b_1}} \Phi\left[\frac{-a_1 T - N}{\sqrt{b_1 T}}\right] \quad (39)$$

and

$$\Pr\{\tau_2(x) > t\} \approx \Phi\left[\frac{x + a_2 T}{\sqrt{b_2 T}}\right] - e^{-\frac{2a_2 x}{b_2}} \Phi\left[\frac{a_2 T - x}{\sqrt{b_2 T}}\right]. \quad (40)$$

Using Eqs. (39) and (40), we obtain

$$\begin{aligned} E[e^{-\beta\tau_1}; \{\tau_1 \leq T\}] &\approx \int_0^T e^{-\beta t} p_1(t) dt \\ &= e^{\frac{a_1 - \psi_1}{b_1} N} \left(\Phi\left[\frac{\psi_1 T - N}{\sqrt{b_1 T}}\right] \right. \\ &\quad \left. + e^{\frac{2\psi_1 N}{b_1}} \Phi\left[\frac{-\psi_1 T - N}{\sqrt{b_1 T}}\right] \right), \quad (41) \end{aligned}$$

$$E[e^{-\beta\tau_2(N)}] \approx \int_0^\infty e^{-\beta t} p_2(t, N) dt = e^{-\frac{a_2 + \psi_2}{b_2} N} \quad (42)$$

and

$$\begin{aligned} E[e^{-\beta\tau_2(N(T))}; \{\tau_1 > T\}] &\approx \int_{-\infty}^N \int_0^\infty e^{-\beta s} p_2(s, x) ds q(x, T) dx \\ &= e^{-\frac{a_1^2 - \xi^2}{2b_1} T} \left(\Phi\left[\frac{N + \xi T}{\sqrt{b_1 T}}\right] \right. \\ &\quad \left. - e^{-\frac{2\xi N}{b_1}} \Phi\left[\frac{-N + \xi T}{\sqrt{b_1 T}}\right] \right), \quad (43) \end{aligned}$$

where

$$p_1(t) \equiv \frac{\partial}{\partial t} \Pr\{\tau_1 \leq t\} = \frac{|N|}{\sqrt{2\pi b_1 t^3}} e^{-\frac{(N - a_1 t)^2}{2b_1 t}}, \quad (44)$$

$$p_2(t, x) \equiv \frac{\partial}{\partial t} \Pr\{\tau_2(x) \leq t\} = \frac{|x|}{\sqrt{2\pi b_2 t^3}} e^{-\frac{(x + a_2 t)^2}{2b_2 t}}, \quad (45)$$

$$\psi_1 = \sqrt{a_1^2 + 2\beta b_1} \quad (46)$$

and

$$\psi_2 = \sqrt{a_2^2 + 2\beta b_2}. \quad (47)$$

Substituting Eqs.(41)-(43) into Eqs.(18) and (20), we can evaluate the approximation formula for the expected total discounted cost over an infinite time span.

5 Numerical Comparison

Of our interest is to evaluate the diffusion approximation methods proposed in Section 4 in terms of their precision. Numerical comparison is carried out for two examples; $M/M/1$ and $M/D/1$ queueing systems. The cost parameters used are the following:

$$\begin{aligned} \mu &= 1.0 & \beta &= 0.01 & r_1 &= 1 & r_2 &= 5 \\ R_1 &= 10 & R_2 &= 10 & h &= 1 \end{aligned}$$

The accuracy of the approximate solutions is measured by the relative error:

$$\frac{(\text{approximate value}) - (\text{exact value})}{(\text{exact value})} \times 100(\%).$$

Table 1: Approximation of $M/M/1$ queueing system under timing strategy (method 1).

ρ	exact optimal		approximate optimal		error (%)
	T^*	$TC_1(T^*)$	T^*	$TC_{A_1}(T^*)$	
0.1	20.07	324.01	20.07	324.01	0.0001
0.2	13.23	440.38	13.23	440.38	0.0002
0.3	10.06	534.77	10.06	534.77	0.0004
0.4	8.05	618.25	8.05	618.25	0.0006
0.5	6.59	697.00	6.59	697.00	0.0012
0.6	5.41	777.47	5.41	777.49	0.0024
0.7	4.40	870.27	4.40	870.32	0.0051
0.8	3.51	997.57	3.51	997.70	0.0123
0.9	2.71	1209.63	2.71	1210.02	0.0319

Table 2: Approximation of $M/D/1$ queueing system under timing strategy (method 1).

ρ	exact optimal		approximate optimal		error (%)
	T^*	$TC_1(T^*)$	T^*	$TC_{A_1}(T^*)$	
0.1	20.06	323.72	20.01	319.26	-1.3789
0.2	13.21	438.55	13.18	429.32	-2.1048
0.3	10.04	529.59	10.01	515.56	-2.6506
0.4	8.02	607.10	8.00	588.26	-3.1035
0.5	6.54	675.87	6.52	652.27	-3.4915
0.6	5.35	739.94	5.33	711.72	-3.8141
0.7	4.31	805.54	4.30	773.03	-4.0351
0.8	3.37	887.43	3.36	851.63	-4.0349
0.9	2.48	1031.20	2.48	995.28	-3.4835

Tables 1 and 2 present numerical comparisons based on the diffusion approximation method (method 1) developed in Section 2 for the optimal timing strategy, where TC_{A_1} denotes the approximated cost value. From these results, the approximation for the $M/M/1$ queueing system is better than that for the $M/D/1$ system. Notice that the above method 1 approximates directly the stochastic processes during both the busy and idle periods. In order to approximate consistently the queueing process, it is sufficient to approximate only the $M/G/1$ busy process (method 2). This means

$$G(\beta) \approx e^{-\frac{a_2 + \psi_2}{b_2}}. \quad (48)$$

It should be noted that such a treatment was not done by Kimura (1981).

Tables 3 and 4 are the results based on the method 2, where TC_{B_1} denotes the revised approximation cost value. The results show that the method 2 is better than the method 1 for both the $M/M/1$ and $M/D/1$ queueing systems and satisfy our intuition.

Next, we numerically evaluate the alternative timing control strategy based on the diffusion approximations. The parameters used are the following:

Table 3: Approximation of $M/M/1$ queueing system under timing strategy (method 2).

ρ	exact optimal		approximate optimal		error (%)
	T^*	$TC_1(T^*)$	T^*	$TC_{B_1}(T^*)$	
0.1	20.07	324.01	20.07	323.99	-0.0048
0.2	13.23	440.38	13.23	440.34	-0.0082
0.3	10.06	534.77	10.06	534.71	-0.0117
0.4	8.05	618.25	8.05	618.15	-0.0157
0.5	6.59	697.00	6.59	696.85	-0.0205
0.6	5.41	777.47	5.41	777.27	-0.0265
0.7	4.40	870.27	4.40	869.98	-0.0337
0.8	3.51	997.57	3.51	997.17	-0.0409
0.9	2.71	1209.63	2.71	1209.13	-0.0413

Table 4: Approximation of $M/D/1$ queueing system under timing strategy (method 2).

ρ	exact optimal		approximate optimal		error (%)
	T^*	$TC_1(T^*)$	T^*	$TC_{B_1}(T^*)$	
0.1	20.06	323.72	20.01	319.24	-1.3839
0.2	13.21	438.55	13.18	429.28	-2.1134
0.3	10.04	529.59	10.01	515.49	-2.6631
0.4	8.02	607.10	8.00	588.16	-3.1206
0.5	6.54	675.87	6.52	652.11	-3.5147
0.6	5.35	739.94	5.33	711.48	-3.8460
0.7	4.31	805.54	4.30	772.67	-4.0805
0.8	3.37	887.43	3.36	851.02	-4.1029
0.9	2.48	1031.20	2.48	994.15	-3.5929

Table 5: Approximation of $M/M/1$ queueing system under alternative timing strategy (method 1).

N	exact optimal		approximate optimal		error (%)
	T^*	$TC_2(T^*)$	T^*	$TC_{A_2}(T^*)$	
1	∞	374.82	∞	375.77	-0.2538
2	∞	302.88	∞	303.99	-0.3648
3	39.60	306.97	∞	309.09	-0.6907
4	24.96	316.69	23.25	321.34	-1.4686
5	21.69	321.37	20.60	323.57	-0.6846
6	20.61	323.17	20.15	323.95	-0.2417
7	20.24	323.77	20.08	324.00	-0.0706
8	20.12	323.95	20.07	324.01	-0.0178

Table 6: Approximation of $M/D/1$ queueing system under alternative timing strategy (method 1).

N	exact optimal		approximate optimal		error (%)
	T^*	$TC_2(T^*)$	T^*	$TC_{A_2}(T^*)$	
1	∞	374.42	∞	370.43	1.0648
2	∞	302.57	∞	298.97	1.1887
3	39.56	306.68	∞	304.34	0.7652
4	24.94	316.41	23.13	316.61	-0.0637
5	21.68	321.09	20.53	318.83	0.7043
6	20.60	322.89	20.08	319.21	1.1407
7	20.23	323.49	20.01	319.25	1.3094
8	20.11	323.66	20.01	319.26	1.3615

$$\begin{matrix} \mu = 1.0 & \rho = 0.1 & \beta = 0.01 & r_1 = 1 & r_2 = 5 \\ R_1 = 10 & R_2 = 10 & R_3 = 15 & h = 1 \end{matrix}$$

Tables 5-8 are the comparison results under the alternative timing strategy, where TC_2 , TC_{A_2} and TC_{B_2} denote the exact minimum cost value, the approximated cost values based on the methods 1 and 2, respectively. From these, we observe that the alternative timing strategy has the same tendencies as the ordinary one in terms of the diffusion approximation.

Acknowledgment: The authors would like thank Professor T. Kimura, Department of Business Administration, Hokkaido University, for his helpful guidance in the area of queueing approximations. This work was partially supported by a Grant-in-Aid for Scientific Re-

Table 7: Approximation of $M/M/1$ queueing system under alternative timing strategy (method 2).

N	exact optimal		approximate optimal		error (%)
	T^*	$TC_2(T^*)$	T^*	$TC_{B_2}(T^*)$	
1	∞	374.82	∞	374.80	0.0047
2	∞	302.88	∞	302.87	0.0054
3	39.60	306.97	39.60	306.95	0.0051
4	24.96	316.69	24.96	316.67	0.0049
5	21.69	321.37	21.69	321.35	0.0048
6	20.61	323.17	20.61	323.16	0.0048
7	20.24	323.77	20.24	323.76	0.0048
8	20.12	323.95	20.12	323.93	0.0048

Table 8: Approximation of $M/D/1$ queueing system under alternative timing strategy (method 2).

N	exact optimal		approximate optimal		error (%)
	T^*	$TC_2(T^*)$	T^*	$TC_{B_2}(T^*)$	
1	∞	374.42	∞	369.46	1.3253
2	∞	302.57	∞	297.84	1.5615
3	39.56	306.68	39.29	302.16	1.4760
4	24.94	316.41	24.85	311.93	1.4152
5	21.68	321.09	21.61	316.62	1.3930
6	20.60	322.89	20.54	318.42	1.3861
7	20.23	323.49	20.17	319.01	1.3843
8	20.11	323.66	20.05	319.18	1.3840

search from the Ministry of Education, Science and Culture of Japan under Grant No. 08780427.

References

[1] C. E. Bell: "Characterization and computation of optimal policies for operating an $M/G/1$ queueing system with removable server", *Ope. Res.*, 19, 208/218 (1971).

[2] T. Dohi, N. Kaio and S. Osaki: "Optimal control of a finite dam with sample path constraint", *Math. Comput. Modelling*, 22, 45/51 (1995).

[3] D. P. Heyman: "Optimal operating policies for $M/G/1$ queueing systems", *Ope. Res.*, 16, 362/382 (1968).

[4] D. P. Heyman: "The T -policy for the $M/G/1$ queue", *Management Sci.*, 23, 775/778 (1977).

[5] S. Karlin and H. M. Taylor: *A Second Course in Stochastic Processes*, Academic Press, N. Y. (1981).

[6] T. Kimura: "Optimal control of an $M/G/1$ queueing system with removable server via diffusion approximation", *Euro. J. Opl. Res.*, 8, 390/398 (1981).

[7] H. Kobayashi: "Application of the diffusion approximation to queueing networks I, equilibrium queue distributions", *J. ACM.*, 21, 316/328 (1974).

[8] M. J. Sobel: "Optimal average-cost policy for a queue with start-up and shut-down costs", *Ope. Res.*, 17, 145/162 (1969).

[9] H. Takagi: *Queueing Analysis*, Vol. 1, Elsevier Sci. Pub. The Netherlands (1991).

[10] J. Jr. Teghem: "Control of the service process in a queueing system", *Euro. J. Opl. Res.*, 23, 141/158 (1986).