Journal of the Operations Research Society of Japan
Vol. 22, No. 4, December 1979

DIFFUSION APPROXIMATION FOR GI/G/1 QUEUEING SYSTEMS WITH FINITE CAPACITY: II - THE STATIONARY BEHAVIOUR

Toshikazu Kimura
Katsuhisa Ohno
and
Hisashi Mine
Kyoto University

(Received September 14, 1978)

Abstract This paper is concerned with stationary behaviour of the GI/G/1(N) queueing system. Approximate formulae for the stationary loss probability and the stationary distribution of the number of customers are explicitly derived from using diffusion approximation and renewal theory. Moreover, approximate formulae for the mean number of customers are also obtained. These results are then modified so as to be more accurate for small N. It is shown from numerical results that these approximate formulae are sufficiently accurate for a practical use.

1. Introduction

In a previous paper Part I [8], the *first overflow time* for a *GI/G/1* queueing system with finite capacity was investigated by diffusion approximation. The first overflow time means the time at which the number of customers in the system first exceeds the capacity. In Part I, approximate expressions for the distribution and moments of the first overflow time are derived explicitly.

The present paper is also concerned with the *GI/G/1* system with finite capacity. In particular, its behaviour after the first overflow time is analyzed by diffusion approximation. It is assumed that overflowed customers are cleared. The same notation as in Part I will be used throughout this paper. That is, the d.f.s of the interarrival times and the service times are denoted by A and H, respectively, such that

$$\frac{1}{\lambda} = \int_0^\infty x dA(x), \qquad \sigma_a^2 = \int_0^\infty (x - \frac{1}{\lambda})^2 dA(x),$$

$$\frac{1}{\mu} = \int_0^\infty x dH(x) \quad \text{and} \quad \sigma_s^2 = \int_0^\infty (x - \frac{1}{\mu})^2 dH(x).$$

The maximum number of customers allowed in the system is, however, given by N for the notational convenience (unlike the choice of the capacity in Part I). Hence, this system is represented by $GI/G/1(N)$. Using the method of diffusion approximation, we shall investigate the stationary behaviour of the $GI/G/1(N)$ system.

In the analysis of a multiprogramming computer system, Gaver and Shedler [5] have studied the $GI/G/1(N)$ system to obtain *CPU* utilization by using diffusion approximation. Gelenbe [6] has also studied the same model as in Gaver and Shedler by modifying boundary conditions imposed on the diffusion process approximating the number of customers in the system. Since the boundary conditions used by Gelenbe are the *elementary return* stated in Part I, his solutions fairly agree with the exact ones for the $M/M/1(N)$ system. His boundary conditions, however, are not appropriate for the other systems. Furthermore, his analysis is concerned with obtaining only the utilization of the system.

This paper presents a new approach to boundary conditions in diffusion approximation and is concerned with various queueing characteristics of the $GI/G/1(N)$ system. In Section 2, under an appropriate assumption, it is shown that the behaviour of the system is precisely expressed by using renewal theory. From this consideration, we derive the probability that the capacity is fully occupied at time t. In Section 3, the stationary distribution of the number of customers and its mean are derived. In Section 4, the results obtained in Section 3 are modified so as to be more accurate for a system with small capacity. Finally, in Section 5, numerical results are provided to show the accuracy of diffusion approximation.

2. Loss Probability

By using renewal theory, we shall investigate the probability that the capacity is fully occupied. In the following, the state in which the capacity is fully occupied is called the *loss* state, and its probability is called the loss probability. Since customers arriving in the loss state are lost, in order to analyze the loss of customers, it is enough only to know whether the system is in the loss state or not. That is, the states of the system can be classified into two states, *loss* and *non-loss* states. Let $Q(t)$ denote the number of customers in the system at time t, and assume that $Q(0) = x_0$. Denote by $T^{(1)}$ the first entrance time to the loss state. Let $T_b^{(n)}$ ($n = 1, 2, \ldots$) be the length of the n^{th} loss period in which the capacity is fully occupied. Moreover, let $T^{(n)}$ ($n = 2, 3, \ldots$) be the length of the period from the end of the $(n - 1)^{st}$ loss period to the beginning of the n^{th} one. Then, because of the restriction on the number of customers, the path $Q(t)$ behaves in the following manner (see Figure 1). That is, the system enters first into the loss state at time $T^{(1)}$ after starting from $Q(0) = x_0$, and then the number of customers jumps into $N - 1$ at time $T^{(1)} + T_b^{(1)}$ when a customer in service departs from the system. After that, the process restarts from $N - 1$ and enters into the loss state at time $T^{(1)} + T_b^{(1)} + T^{(2)}$. Thereafter loss and non-loss states repeat alternately in the same manner as above.

Since the behaviour of the path $Q(t)$ during the periods $T^{(n)}$ coincides with that for the standard $GI/G/1$ system with infinite capacity, the process $Q(t)$ during $T^{(n)}$ can be approximated by a diffusion process $x(t)$ with the following diffusion parameters given in Section 2 of Part I:

$$b \equiv \lim_{\Delta t \downarrow 0} \frac{E[x(t + \Delta t) - x(t) \mid x(t)]}{\Delta t} = \lambda - \mu,$$

and

$$a \equiv \lim_{\Delta t \downarrow 0} \frac{\mathrm{Var}[x(t + \Delta t) - x(t) \mid x(t)]}{\Delta t} = \lambda^3\sigma_a^2 + \mu^3\sigma_s^2.$$

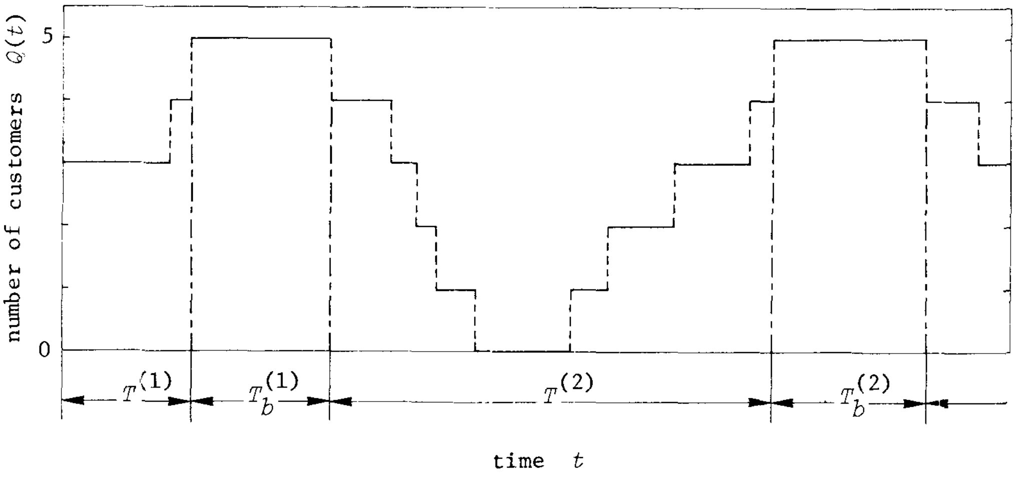

Figure 1. The behaviour of a sample path $Q(t)$ ($x_0 = 3$, $N = 5$).

Since $T^{(1)}$ is the time at which the number of customers first reaches the capacity level N, it can be approximated by the first overflow time

$$T_d(x_0,N) \equiv \inf\{t \geqq 0 \mid x(t) = N,\ x(0) = x_0\},$$

which is analyzed in Part I. Similarly, $T^{(n)}$ $(n = 2,3,\ldots)$ can be approximated by $T_d(N-1,N)$. Now, assume that for all $x \in [0, N)$ and $n = 1,2,\ldots,$ $T_b^{(n)}$ are *independent* of $T_d(x,N)$ and *i.i.d.* with a d.f. B. Then approximating $T^{(1)}$ by $T_d(x_0,N)$ and $T^{(n)}$ $(n = 2,3,\ldots)$ by $T_d(N-1,N)$, respectively, yields that a sequence of the beginning of loss and non-loss states are *renewal points*. That is, the loss and non-loss states make *a delayed alternating renewal process* (DARP) [12].

Remark 2.1. The above assumption on the independence does not always hold for the $GI/G/1(N)$ system. Because the loss periods terminate at customers' departure epochs, the d.f. B is closely relevant to the d.f. H. For the $M/M/1(N)$ system, it follows from the memoryless property that the above assumption certainly holds, that is, the d.f. B is the negative exponential distribution with mean $1/\mu$.

Let $P_N(t|x_0)$ denote the probability of the loss state at time t, given that $x(0) = x_0$. Furthermore, for $\mathrm{Re}\ s \geqq 0$, $p_N(s|x_0)$ denotes its Laplace-Stieltjes transform (L.S.T.). As in Part I, denote by $F(t|x_0,N)$ and $f(s|x_0,N)$ the d.f. of $T_d(x_0,N)$ and its L.S.T., respectively. Setting $F_0(t) = F(t|x_0,N)$ and $F_1(t) = F(t|N-1,N)$, we have the following renewal-type equation on $P_N(t|x_0)$:

$$(2.1)\qquad P_N(t|x_0) = F_0(t) - \int_0^t F_0(t-u)dB(u) + \int_0^t P(t-u)d\{F_0 * B * F_1(u)\},$$

where $*$ denotes convolution and $P(t)$ the probability of the loss state at time t in an alternating renewal process which is generated by deleting the first renewal from the DARP. The probability $P(t)$ also satisfies the renewal-type equation

$$(2.2)\qquad P(t) = 1 - B(t) + \int_0^t P(t-u)d\{B * F_1(u)\}.$$

Taking the L.S.T. of (2.1) and (2.2), we obtain

$$(2.3)\qquad p(s) = \frac{1 - b(s)}{1 - b(s)f(s|N-1,N)},$$

and therefore

$$(2.4)\qquad p_N(s|x_0) = \frac{\{1 - b(s)\}f(s|x_0,N)}{1 - b(s)f(s|N-1,N)}.$$

From (2.4), the stationary loss probability, $P_N \equiv \lim_{t\to\infty} P_N(t|x_0)$ is given by

$$(2.5) \qquad P_N = \frac{E[T_b]}{E[T_b] + E[T_d(N-1,N)]},$$

where T_b is a typical random variable representing the length of a loss period.

Remark 2.2. In order to obtain good fitness for the $M/G/1(N)$ system, it is desirable that $\underset{\sim}{f}$ in (2.4) and $E[T_d(N-1,N)]$ in (2.5) are respectively replaced by $\tilde{f}$ and $E[\tilde{T}_d(N-1,N)]$, where $\sim$ means the modification stated in Section 4.2 of Part I. In particular, for the $M/M/1(N)$ system, substituting (I.4.9)† and $E[T_b] = 1/\mu$ in (2.5), we have

$$(2.6) \qquad \tilde{P}_N = \begin{cases} \dfrac{\rho(1-\rho)}{exp\{2(N-1)(1-\rho)/(1+\rho)\} - \rho^2}, & \text{if } \rho \neq 1 \\ \dfrac{1}{N+1}, & \text{if } \rho = 1. \end{cases}$$

Therefore, it follows that, for $\rho = 1$ and/or $N = 1$, $\tilde{P}_N$ agrees with the exact solution for the $M/M/1(N)$ system [4].

Remark 2.3. It is worth while noting that

$$(2.7) \qquad \lim_{N\to\infty} P_N = \lim_{N\to\infty} \tilde{P}_N = \begin{cases} \dfrac{(\lambda-\mu)E[T_b]}{(\lambda-\mu)E[T_b]+1}, & \text{if } \rho > 1 \\ 0, & \text{if } \rho \leqq 1. \end{cases}$$

In particular, for the $M/M/1(N)$ system,

$$(2.8) \qquad \lim_{N\to\infty} P_N = \lim_{N\to\infty} \tilde{P}_N = \begin{cases} 1 - \dfrac{\mu}{\lambda}, & \text{if } \rho > 1 \\ 0, & \text{if } \rho \leqq 1, \end{cases}$$

which is consistent with the analytical result [4].

To derive the total number of times that the system is in the loss state, it suffices to consider a delayed renewal process in which the renewal points are epochs of the ends of the loss states. Consequently, the interevent times are given by $T_b^{(n)} + T_d(N-1,N)$. Let $N_i(t|x_0)$ denote the i^{th} moment of the total number of times the system is in the loss state up to time t starting from $x(0) = x_0$, and $n_i(s|x_0)$ its L.S.T.. Then using renewal theory [1, pp. 97-98], for $i = 1,2$, we obtain

† Equations from Part I will be denoted by (I.•.•).

(2.9) $$n_1(s|x_0) = \frac{f(s|x_0,N)}{1 - b(s)f(s|N-1,N)},$$

and

(2.10) $$n_2(s|x_0) = \frac{1 + b(s)f(s|N-1,N)}{\{1 - b(s)f(s|N-1,N)\}^2} f(s|x_0,N).$$

From (2.9), the elementary renewal theory leads to

(2.11) $$\lim_{t\to\infty} \frac{N_1(t|x_0)}{t} = \frac{1}{E[T_b] + E[T_d(N-1,N)]}.$$

The right hand side of (2.11) represents the average number of times per unit time that the system is in the loss state.

3. The Number of Customers in the System

In this section we shall investigate the stationary distribution of the number of customers in the $GI/G/1(N)$ system. For the $M/M/1(N)$ system, this problem has been solved by Finch [4]. For the $M/G/1(N)$ and the $GI/M/1(N)$ systems, it has been studied by Keilson [7] and a few researchers.

Let define $r(x,t|x_0,N)$ as

(3.1) $$r(x,t|x_0,N)dx \equiv P\{x \leqq x(t) < x + dx|x(0) = x_0\}.$$

Then, by formalizing $x(t)$ as the DARP similarly to Section 2, it is easily shown that $r(x,t|x_0,N)$ satisfies the following renewal-type equation:

(3.2) $$r(x,t|x_0,N) = p(x,t|x_0) + \int_0^t r(x,t-u|N-1,N)d\{F_0 * B(u)\},$$

where

(3.3) $$p(x,t|x_0)dx \equiv P\{x \leqq x(t) < x + dx,\ T_d(x_0,N) > t|x(0) = x_0\}.$$

Taking the Laplace transform (L.T.) of (3.2), we have

(3.4) $$r^*(x,s|x_0,N) = p^*(x,s|x_0) + f(s|x_0,N)b(s)r^*(x,s|N-1,N),$$

where * represents the L. T. of the corresponding function. Setting $x_0 = N - 1$ in (3.4) and arranging it lead to

(3.5) $$r^*(x,s|x_0,N) = p^*(x,s|x_0) + k(s|N)p^*(x,s|N-1),$$

where

(3.6) $$k(s|N) \equiv \frac{b(s)f(s|x_0,N)}{1 - b(s)f(s|N-1,N)}.$$

It remains only to develop an expression for $p^*(x,s|x_0)$. It is easily shown that $p(x,t|x_0)$ satisfies the partial differential equation (I.2.2) with the initial condition (I.2.5), a conservative boundary at $x = 0$ and an absorbing barrier at $x = N$. Hence, $p^*(x,s|x_0)$ satisfies the following ordinary differential equation:

$$\frac{1}{2} a \frac{d^2p^*}{dx^2} - b \frac{dp^*}{dx} - sp^* = - \delta(x - x_0) \tag{3.7}$$

with the boundary condition for the absorbing barrier at $x = N$

$$p^*(N,s|x_0) = 0. \tag{3.8}$$

If we place a reflecting barrier at $x = 0$, the corresponding boundary condition is given by

$$\left.\frac{1}{2} a \frac{dp^*}{dx} - bp^* \right|_{x=0} = 0. \tag{3.9}$$

Next, we shall state without proof the following lemma [2, 3] which derives the solution of (3.7) with (3.8) and (3.9).

Lemma. *For* $x \in (r_1, r_2)$, *if we let* $\xi_i(x)$ $(i = 1,2)$ *be a non-negative solution of the homogeneous differential equation*

$$\frac{1}{2} a \frac{d^2p^*}{dx^2} - b \frac{dp^*}{dx} - sp^* = 0 \tag{3.10}$$

with one boundary condition at $x = r_i$, *then a solution of* (3.7) *with two boundary conditions at* $x = r_1$ *and* $x = r_2$ *is given by*

$$p^*(x,s|x_0) = \frac{2}{a} G(x,x_0), \tag{3.11}$$

where $G(x,x_0)$ *is the Green function of* (3.10) *and is given by*

$$G(x,x_0) = \begin{cases} \dfrac{\xi_1(x)\xi_2(x_0)}{\xi_1'(x_0)\xi_2(x_0) - \xi_1(x_0)\xi_2'(x_0)}, & \text{for } r_1 < x \leqq x_0 < r_2 \\[2ex] \dfrac{\xi_1(x_0)\xi_2(x)}{\xi_1'(x_0)\xi_2(x_0) - \xi_1(x_0)\xi_2'(x_0)}, & \text{for } r_1 < x_0 < x < r_2. \end{cases} \tag{3.12}$$

Putting $r_1 = 0$ and $r_2 = N$ and using the boundary condition (3.8) and (3.9), this lemma yields

$$\begin{aligned} \xi_1(x) &= exp(bx/a)\{exp(cx) + \frac{ac-b}{ac+b} exp(-cx)\} \\ \xi_2(x) &= exp(bx/a)\{exp\{-c(x-2N)\} - exp(cx)\}, \end{aligned} \tag{3.13}$$

where $c \equiv \sqrt{b^2 + 2as}/a$. Consequently, we obtain

(3.14) $$p^*(x,s|x_0) = \frac{exp\{b(x-x_0)/a\}}{ac\{(ac+b)exp(2cN) + (ac-b)\}} \cdot \Big((ac+b)\{exp\{-c(|x-x_0|-2N)\} - exp\{c(x+x_0)\}\} + (ac-b)\{exp\{-c(x+x_0-2N)\} - exp\{c|x-x_0|\}\}\Big).$$

Let define $r(x|N)$ as

(3.15) $$r(x|N) \equiv \lim_{t\to\infty} r(x,t|x_0,N),$$

and $K(N)$ as

(3.16) $$K(N) \equiv \lim_{s\downarrow 0} sk(s|N) = \frac{1}{E[T_b] + E[T_d(N-1,N)]}.$$

It is noted that $K(N)$ agrees with (2.11). Then, from (3.5), (3.6) and (3.14), we obtain

(3.17) $$r(x|N) = \lim_{s\downarrow 0} sp^*(x,s|x_0,N) = \begin{cases} K(N)\,\frac{1}{b}\,\big(exp\{b(x+1-N-|x+1-N|)/a\} - exp\{2b(x-N)/a\}\big), & \text{if } b \neq 0 \\ K(N)\,\frac{1}{a}\,\{(-x+1+N) - |x+1-N|\}, & \text{if } b = 0. \end{cases}$$

From the relation with the first overflow time, it can be shown that $\{r(x|N), P_N\}$ constructs a proper distribution truncated at $x = N$. Because, it follows from (I.3.2) that

(3.18) $$f(s|x_0,N) = 1 - s\int_0^N p^*(x,s|x_0)dx.$$

Differenciating (3.18) with respect to s and letting $s \downarrow 0$, we have

(3.19) $$E[T_d(x_0,N)] = \lim_{s\downarrow 0}\int_0^N p^*(x,s|x_0)dx.$$

Hence, combining (2.5), (3.17) and (3.19) leads to

(3.20) $$\int_0^N r(x|N)dx + P_N = 1.$$

Now, from $r(x|N)$, we shall derive a stationary discrete distribution of the number of customers in the system. For $k = 0,1,\ldots,N-1$, let P_k denote the stationary probability that there are k customers in the system. Then we define P_k $(k = 0,1,\ldots,N-1)$ as

$$(3.21) \qquad P_k \equiv \int_k^{k+1} r(x|N)dx,$$

and it is clear from (3.20) that $\sum_{k=0}^{N} P_k = 1$. Substituting (3.17) in (3.21), we obtain, for $k = 0,1,\ldots,N-2$,

$$(3.22) \qquad P_k = \begin{cases} K(N)\dfrac{a}{2b^2}\{exp(2b/a) - 1\}^2 exp\{2b(k-N)/a\}, & \text{if } b \neq 0 \\ K(N)\dfrac{2}{a}, & \text{if } b = 0, \end{cases}$$

and for $k = N-1$,

$$(3.23) \qquad P_{N-1} = \begin{cases} K(N)\dfrac{a}{2b^2}\{exp(-2b/a) + 2b/a - 1\}, & \text{if } b \neq 0 \\ K(N)\dfrac{1}{a}, & \text{if } b = 0. \end{cases}$$

Remark 3.1. From (3.22), it follows that for $k = 0,1,\ldots,$

$$(3.24) \qquad \lim_{N\to\infty} P_k = \begin{cases} \{1 - exp(2b/a)\}exp(2bk/a), & \text{if } \rho < 1 \\ 0, & \text{if } \rho \geqq 1. \end{cases}$$

These limits agree with the first-order results obtained by Kobayashi [11].

Let L denote the mean number of customers in the system, i.e.,

$$(3.25) \qquad L \equiv \sum_{k=0}^{N} k \cdot P_k.$$

Then, from (2.5), (3.22) and (3.23), we have

$$(3.26) \qquad L = \begin{cases} K(N)\left(\dfrac{a}{2b^2}\left\{exp\{-2b(N-1)/a\} + \dfrac{2b(N-1)}{a} - 1\right\} + N\cdot E[T_b]\right), & \text{if } b \neq 0 \\ K(N)\left\{\dfrac{(N-1)^2}{a} + N\cdot E[T_b]\right\}, & \text{if } b = 0. \end{cases}$$

Next, we briefly discuss the stationary waiting time. It is assumed that waiting times are only defined for customers who are admitted to the system. Let W_q denote the mean waiting time in this sense. Then, we have

$$(3.27) \qquad W_q = \frac{L - (1 - P_0)}{\lambda(1 - P_N)}.$$

This equation can be easily obtained by the standard arguement for the Little's formula.

Remark 3.2. It is possible to define the mean number of customers by

$$(3.28) \qquad L_c \equiv \int_0^N x r(x|N)dx + N\cdot P_N.$$

From (2.5) and (3.17), we obtain

$$(3.29) \qquad L_c = \begin{cases} K(N)\left(\frac{1}{2b}\left\{2N - 1 - \frac{a}{b} - \frac{a^2}{2b^2}\{1 - exp(2b/a)\}exp(-2bN/a)\right\}\right. \\ \qquad\qquad\qquad\qquad \left. + N\cdot\mathrm{E}[T_b]\right), & \text{if } b \neq 0 \\ K(N)\left\{\frac{(3N^2 - 3N + 1)}{3a} + N\cdot\mathrm{E}[T_b]\right\}, & \text{if } b = 0. \end{cases}$$

We shall investigate the accuracy of L and L_c in Section 5.

4. Modification of the Boundary Condition

In Section 3 we have considered the reflecting barrier as the boundary condition at the origin. However, since the behaviour of the reflecting barrier process differs from that of the process $Q(t)$ in the neighbourhood of the origin, some modifications are necessary, in particular, for the case of small N.

In this section, we shall modify P_k $(k = 0,1,\ldots,N)$ and L derived in Section 3 on the basis of the same consideration as in Part I. For $k = 0,1,\ldots,N$, let p_k denote the stationary probability that the number of customers in the $M/M/1(N)$ system is equal to k. Then, from [4], we have

$$(4.1) \qquad p_k = \begin{cases} \frac{(1-\rho)\rho^k}{1-\rho^{N+1}}, & \text{if } \rho \neq 1 \\ \frac{1}{N+1}, & \text{if } \rho = 1. \end{cases}$$

Note that if $\rho = 1$, p_k are all equivalent. On the other hand, applying (2.5), (3.22) and (3.23) to the $M/M/1(N)$ system yields for $\rho = 1$

$$(4.2) \qquad P_k = \begin{cases} \frac{2}{2N+1}, & \text{for } k = 0,1,\ldots,N-2 \text{ and } N \\ \frac{1}{2N+1}, & \text{for } k = N-1. \end{cases}$$

Equations (4.1) and (4.2) show that the value of P_{N-1} is half of the other P_k, and that all of P_k fairly differ from p_k especially for the case of small N. The reason for the former is considered as an effect of the discontinuity of $x(t)$ at $x = N$, and that for the latter as that of the reflecting barrier at $x = 0$. In the following, we shall modify the boundary condition at the origin and the definition of P_{N-1}.

From the arguments in Section 4.1 of Part I, let us move the reflecting barrier at $x = 0$ to $x = -\Delta$, or equivalently, move the absorbing barrier at $x = N$ to $x = \hat{N} \equiv N + \Delta$, where

(4.3) $$\Delta \equiv \frac{\frac{\lambda^2\sigma^2}{a} + 1}{4}.$$

Define a modification of P_k denoted by $\hat{P}_k$ as follows:

(4.4) $$\hat{P}_k = \begin{cases} \int_k^{k+1} r(x|\hat{N})dx, & \text{for } k = 0,1,\ldots,N-2 \\ \int_{N-1}^{\hat{N}} r(x|\hat{N})dx, & \text{for } k = N-1 \\ \dfrac{E[T_b]}{E[T_b] + E[T_d(\hat{N}-1,\hat{N})]}, & \text{for } k = N. \end{cases}$$

Then, it is clear from (3.20) that $\sum_{k=0}^{N} \hat{P}_k = 1$. Substituting (3.17) in (4.4), we obtain for $k = 0,1,\ldots,N-2$ and N,

(4.5) $$\hat{P}_k = P_k \Big|_{N=\hat{N}},$$

and for $k = N-1$,

(4.6) $$\hat{P}_{N-1} = \begin{cases} K(\hat{N}) \frac{a}{2b^2} \left(exp(-2b\Delta/a)\{exp(-2b/a) - 1\} + 2b/a\right), & \text{if } b \neq 0 \\ K(\hat{N}) \frac{1}{a} (2\Delta + 1), & \text{if } b = 0, \end{cases}$$

where $P_{\hat{k}}$ are given by (3.22). It follows from (4.2) that for the $M/M/1(N)$ system, $\hat{P}_k = p_k$ if $\rho = 1$.

Next, we shall modify the mean number of customers in the system by using $\{\hat{P}_k\}$. Define the modification of L denoted by $\hat{L}$ as

(4.7) $$\hat{L} \equiv \sum_{k=0}^{N} k \cdot \hat{P}_k.$$

Then, from (4.5) and (4.6), we obtain

(4.8) $$\hat{L} = \begin{cases} K(\hat{N}) \left(\frac{a}{2b^2}\{exp\{-2b(N+\Delta-1)/a\} - exp(-2b\Delta/a) + 2b(N-1)/a\} + N \cdot E[T_b]\right), & \text{if } b \neq 0 \\ K(\hat{N})\{ \frac{1}{a} (N - 1)(N + 2\Delta - 1) + N \cdot E[T_b]\}, & \text{if } b = 0. \end{cases}$$

It will be shown from numerical results that $\hat{L}$ is more accurate than L for small N.

Remark 4.1. For general arrival processes having small value of the coefficient of variation, there are almost no differences between $\{P_k\}$ and $\{\hat{P}_k\}$, or between L and $\hat{L}$, since the sojourn time at $x = 0$ becomes shorter as the coefficient of variation of the process becomes smaller.

5. Numerical Results

In order to examine the accuracy of the approximate formulae obtained in Sections 2 through 4, they are numerically compared with the known analytical results as well as those of a GPSS simulation. For the $M/M/1(N)$ system, exact expressions for the stationary loss probability and the mean number of customers have been derived in [4]. Hence, using the approximate and exact results for the $M/M/1(N)$ system, we have Figures 2 and 3 which illustrate the stationary loss probability and the mean number of customers in the $M/M/1(N)$ system, respectively. In order to show the accuracy of these results in detail, the relative errors for the $M/M/1(N)$ system are given in Table 1. Since it is difficult to obtain usable results for the general case, the approximate results are compared with the results obtained from a GPSS simulation. In this case, it is necessary to assume the expression of the d.f. B,

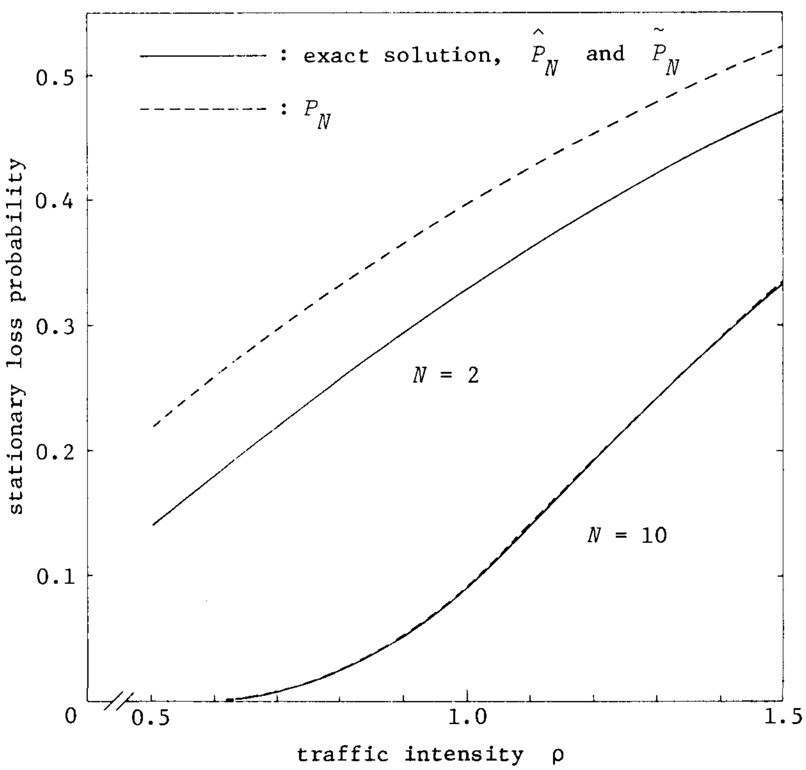

Figure 2. The stationary loss probability for the $M/M/1(N)$ system.

especially that of the mean $E[T_b]$. From the argument on formalizing the DARP, we consider the following two definitions of the d.f. B.

(1) The d.f. B is defined as that of stationary residual service times, i.e.,

$$B(t) = \mu \int_0^t \{1 - H(u)\} du \quad \text{and} \quad E[T_b] = \frac{\mu^2\sigma_s^2 + 1}{2\mu} .$$

(2) The d.f. B is defined as that of service times, i.e.,

$$B(t) = H(t) \quad \text{and} \quad E[T_b] = \frac{1}{\mu} .$$

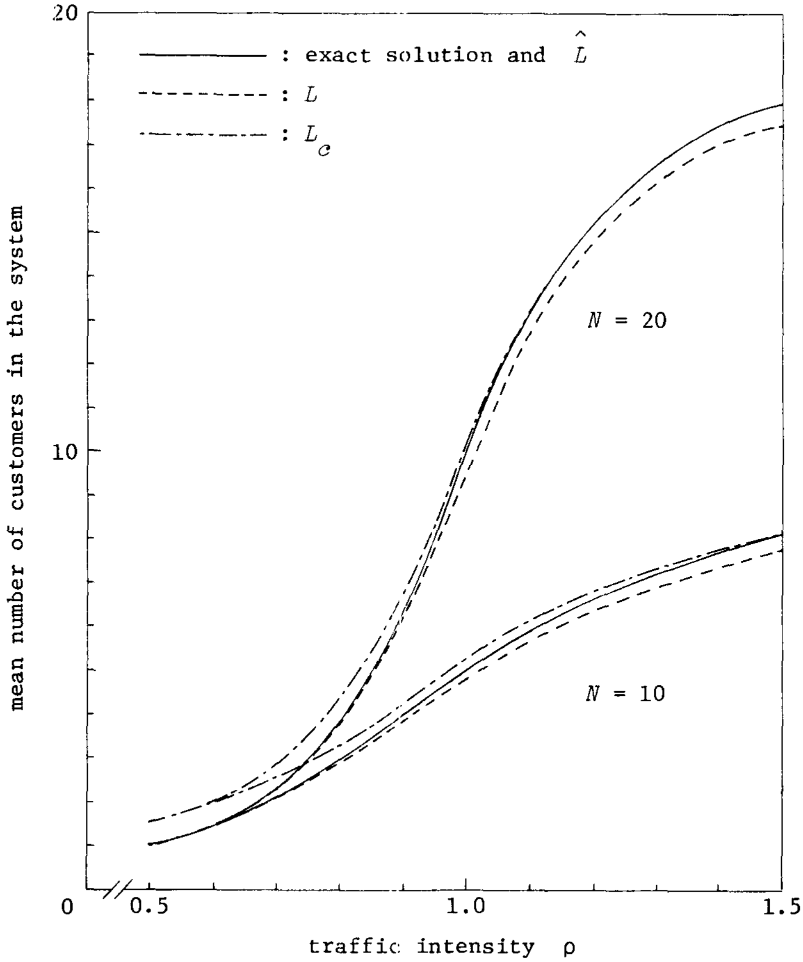

Figure 3. The mean number of customers in the $M/M/1(N)$ system.

Table 1. The relative errors for the $M/M/1(N)$ system.

(a) The stationary loss probability.

N	ρ	exact	P_N	relative error(%)	$\hat{P}_N$	relative error(%)	$\tilde{P}_N$	relative error(%)
2	0.8	0.26230	0.33417	(27.405)	0.26218	(-0.043)	0.26279	(0.189)
	1.0	0.33333	0.40000	(20.000)	0.33333	(0)	0.33333	(0)
	1.2	0.39560	0.45698	(15.516)	0.39690	(0.328)	0.39588	(0.069)
10	0.8	0.02349	0.02675	(13.898)	0.02367	(0.786)	0.02370	(0.911)
	1.0	0.09091	0.09523	(4.761)	0.09091	(0)	0.09091	(0)
	1.2	0.19258	0.19567	(1.615)	0.19278	(0.100)	0.19272	(0.070)

(b) The mean number of customers in the system.

N	ρ	exact	L	relative error(%)	$\hat{L}$	relative error(%)	L_c	relative error(%)
10	0.8	2.9663	2.9100	(-1.897)	2.9750	(0.293)	3.3762	(13.817)
	1.0	5.0000	4.8095	(-3.809)	5.0000	(0)	5.2540	(5.079)
	1.2	6.7107	6.4294	(-4.191)	6.7045	(-0.091)	6.8289	(1.761)
20	0.8	3.8045	3.8030	(-0.040)	3.8196	(0.396)	4.2829	(12.575)
	1.0	10.000	9.7805	(-2.195)	10.000	(0)	10.252	(2.520)
	1.2	15.467	15.092	(-2.424)	15.454	(-0.083)	15.560	(0.253)

Note that these two definitions are equivalent for the $GI/M/1(N)$ system. Although the definition (1) seems to be more natural than (2), it follows from many simulations that the approximate results using (1) are quite underestimated, and that those using (2) are fairly accurate for all cases of simulations. It seems that the validity of using (2) is closely relevant to the assumption on the independence between $T_d(x_0,N)$ and $T_b^{(n)}$. Therefore, we adopt the latter definition (2) in executing the calculations of the approximate formulae. These results are shown in Figures 4 and 5, in which it is assumed that the d.f.s A and H are the Erlang distributions with phase 2 and 5, respectively. Table 2 shows the relative errors for this $E_2/E_5/1(N)$ system. It is concluded from these figures and tables that

i) For the $M/M/1(N)$ system with small capacity N, the stationary loss probability P_N is overestimated. However, the modified loss probability $\tilde{P}_N$ and $\hat{P}_N$ almost agree with the exact one for such a case;

ii) The modified mean number of customers $\hat{L}$ is always plotted between L and L_c. The curve of $\hat{L}$ almost agrees with that of L in the light

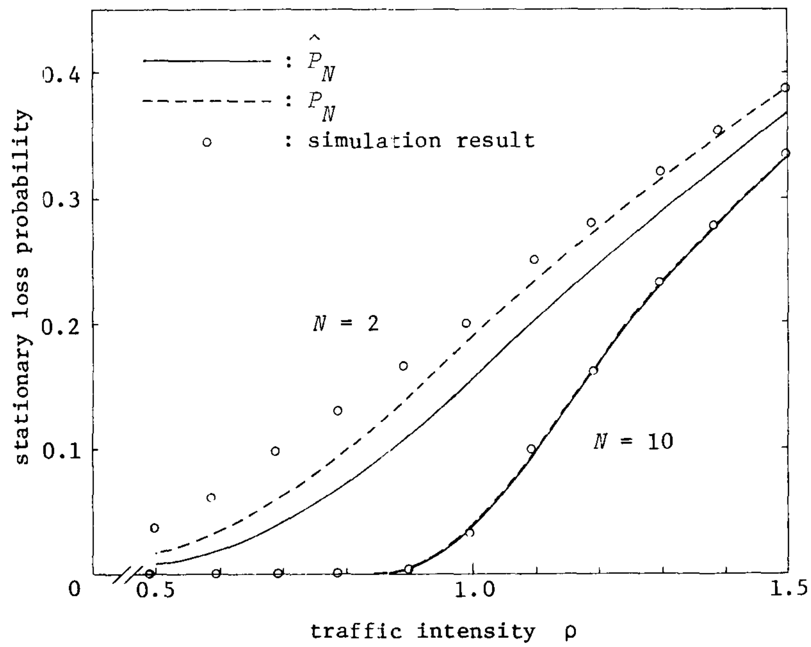

Figure 4. The stationary loss probability for the $E_2/E_5/1(N)$ system.

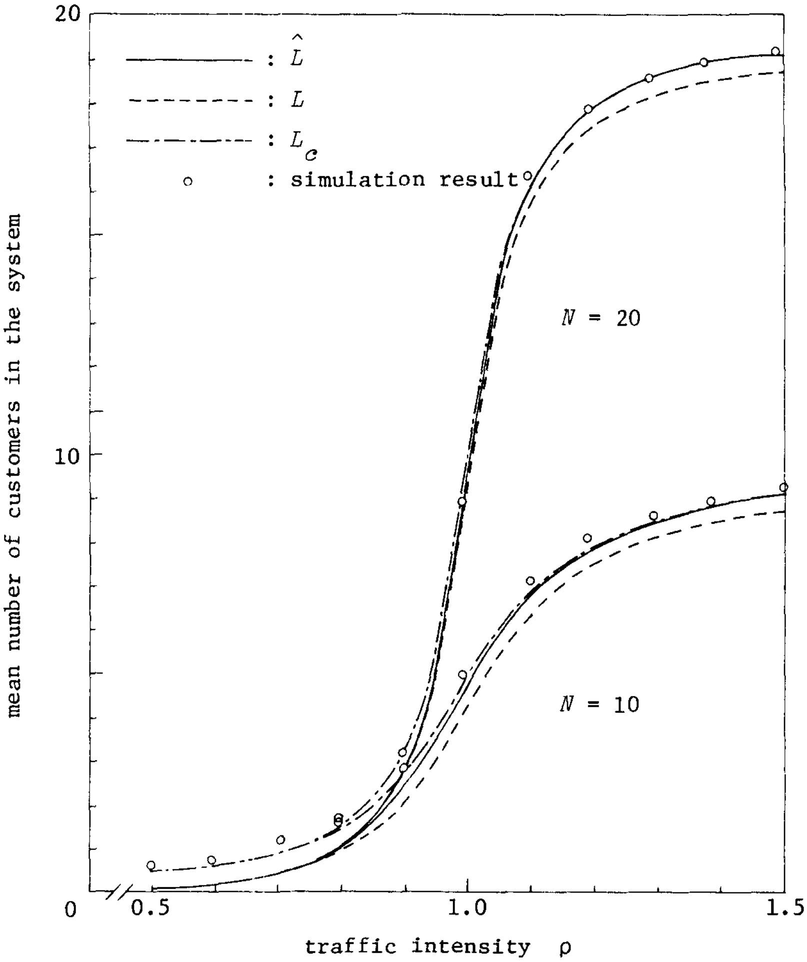

Figure 5. The mean number of customers in the $E_2/E_5/1(N)$ system.

traffic and L_c in the heavy traffic. For the $M/M/1(N)$ system, the modification $\hat{L}$ almost agrees with the exact one;

iii) The accuracy of the approximate formulae heavily depends not on the traffic intensity ρ but on the capacity N.

Moreover, many other numerical results show that

iv) For the $M/M/1(N)$ system with $N \geqq 5$, the absolute errors of P_N are less than 0.01;

v) For the $M/E_k/1(N)$ system, $\hat{P}_N$ as well as $\tilde{P}_N$ is more accurate than P_N;

Table 2. The relative errors for the $E_2/E_5/1(N)$ system.

(a) The stationary loss probability.

N	ρ	simulation	P_N	relative error(%)	$\hat{P}_N$	relative error(%)
2	0.8	0.1313	0.1016	(-22.65)	0.0726	(-44.73)
	1.0	0.2078	0.1892	(-8.95)	0.1573	(-24.29)
	1.2	0.2811	0.2767	(-1.53)	0.2487	(-11.52)
10	0.8	0.00016	0.00035	(118.75)*	0.00027	(68.75)*
	1.0	0.0335	0.0355	(6.18)	0.0342	(2.30)
	1.2	0.1632	0.1679	(2.89)	0.1677	(2.76)

(b) The mean number of customers in the system.

N	ρ	simulation	L	relative error(%)	$\hat{L}$	relative error(%)	L_c	relative error(%)
10	0.8	1.832	1.042	(-43.12)	1.136	(-37.99)	1.486	(-18.88)
	1.0	4.968	4.467	(-10.08)	4.743	(-4.52)	4.941	(-0.54)
	1.2	8.098	7.573	(-6.48)	7.931	(-2.06)	7.989	(-1.34)
20	0.8	1.861	1.055	(-43.31)	1.147	(-38.36)	1.500	(-19.39)
	1.0	8.813	8.503	(-3.51)	8.765	(-0.54)	8.991	(2.01)
	1.2	17.888	17.501	(-2.16)	17.869	(-0.10)	17.918	(0.16)

* The large relative errors with asterisk may be mainly due to errors of exact values estimated by the simulation.

vi) For the $E_k/E_l/1(N)$ system $(k,l \neq 1)$, $\hat{P}_N$ and L_c are more accurate than P_N and L, respectively. In the heavy traffic, $\hat{L}$ is also accurate as well as L_c.

6. Conclusion

As a continued work of Part I, we have investigated the stationary behaviour of the $GI/G/1(N)$ system. Making the appropriate assumptions, we have formalized the process $x(t)$ as the DARP. From this consideration, we have derived the stationary loss probability P_N and the stationary distribution of the number of customers $\{P_k\}$. Moreover, approximate formulae (3.26), (3.29) for the mean number of customers have been derived from this distribution. It is shown from numerical results that they are relatively accurate even for the light traffic and/or small N.

Diffusion approximation is the method in which the queueing process is approximated by one diffusion process with appropriate boundary conditions. The approach adopted in this paper, however, is fairly different from the usual diffusion approximation. That is, the first step of the approach is to decompose the process into subprocesses which follow the same probability laws. The second step is to approximate the subprocesses by using appropriately selected diffusion processes which are mutually independent. Then, the final step is to connect these diffusion processes by using renewal theory, under the approximate assumption on the independence between diffusion processes. This approach enables us to take account of the effects of discontinuity at boundaries. It is also applicable to the analysis of queueing systems with complex structures, e.g., a *tandem* queueing system with a finite waiting room [10], a queueing system with a *removable server* [9] and so on. The first overflow time investigated in Part I plays an important role in this approach.

Applying the renewal theoretic approach to the boundary at the origin instead of the reflecting barrier used in this paper does not seem difficult. However, those results will not have simple forms and significant differences with the present results for the case of the heavy traffic.

Acknowledgment

The authors wish to thank the referees for their helpful suggestions.

References

[1] Cohen, J. W., *The Single Server Queue*, North-Holland, Amsterdam, 1969.

[2] Feller, W., "The Parabolic Differential Equations and the Associated Semigroups of Transforms," *Annals of Mathematics*, 55, 468-519 (1952).

[3] ________, "Diffusion Processes in One Dimension," *Transactions of the American Mathematical Society*, 77, 1-31 (1954).

[4] Finch, P. D., "The Effect of the Size of the Waiting Room on a Simple Queue," *Journal of the Royal Statistical Society*, *Series B*, 20, 182-186 (1958).

[5] Gaver, D. P., Jr. and Shedler, G. S., "Processor Utilization in Multiprogramming Systems via Diffusion Approximation," *Operations Research*, 21, 569-576 (1973).

[6] Gelenbe, E., "On Approximate Computer System Models," *Journal of the Association for Computing Machinery*, 22, 261-269 (1975).

[7] Keilson, J., "The Ergodic Queue Length Distribution for Queueing Systems with Finite Capacity," *Journal of the Royal Statistical Society*, *Series B*, 28, 190-201 (1966).

[8] Kimura, T., Ohno, K. and Mine, H., "Diffusion Approximation for GI/G/1 Queueing Systems with Finite Capacity: I - The First Overflow Time," *Journal of the Operations Research Society of Japan*, 22, 41-68 (1979).

[9] ____________________, "Approximate Analysis of Optimal Operating Policies for a GI/G/1 Queueing System," submitted for publication, (1979).

[10] ____________________, "Diffusion Approximation for Tandem Queueing Systems with a Finite, Intermediate Waiting Room," submitted for publication, (1979).

[11] Kobayashi, H., "Application of the Diffusion Approximation to Queueing Networks. I: Equilibrium Queue Distributions," *Journal of the Association for Computing Machinery*, 21, 316-328 (1974).

[12] Ross, S. M., *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, California, 1970.

Toshikazu KIMURA: Department of Applied Mathematics and Physics, Faculty of Engineering, Kyoto University, Yoshida-Honmachi, Sakyo-Ku, Kyoto 606, Japan.