Journal of the Operations Research
Society of Japan
Vol. 22, No. 1, March, 1979

DIFFUSION APPROXIMATION FOR GI/G/1 QUEUEING SYSTEMS WITH FINITE CAPACITY: I-THE FIRST OVERFLOW TIME

Toshikazu Kimura
Katsuhisa Ohno
and
Hisashi Mine
Kyoto University

(Received May 27, 1978)

Abstract A GI/G/1 queueing system with finite capacity is studied. The first overflow time, which means the time when the number of customers first exceeds the capacity, is analyzed by diffusion approximation. Approximate expressions for the distribution and moments of the first overflow time are derived explicitly. These results are modified so as to be more accurate for an M/G/1 system or a system with small capacity. Further these results are applied to the analysis of the maximum number of customers up to time t in a GI/G/1 system with infinite capacity. Finally, the accuracy of the diffusion approximations is examined numerically by using the analytical results for a GI/M/1 system.

1. Introduction

Consider a *GI/G/1* queueing system with finite capacity for which it is assumed that the maximum number of customers allowed in the system is equal to $N - 1$; that is, the total number of waiting places is $N - 2$. Further assume that interarrival and service times of customers are independent and identically distributed (i.i.d.) random variables with distribution functions (d.f.s) A and H, respectively. Denote the (mean, variance) of the d.f.s A and H by $(\frac{1}{\lambda}, \sigma_a^2)$ and $(\frac{1}{\mu}, \sigma_s^2)$, respectively.

Such a system has been studied as a model for many practical service facilities. In the design of a computer system, for example, one of the important problems is how much capacity is required for a buffer memory. This is because, if its capacity is too little, then overflows of customers (jobs) occur frequently in heavy traffic and the performance of the system deteriorates rapidly. On the other hand, if its capacity is too great, then most buffer memories remain unused. Thus detailed information on the jobs behaviour should be obtained.

For an $M/G/1$ and a $GI/M/1$ systems with finite capacity, stationary queue length distributions have been studied by Keilson [10]. The $M/G/1$ system has also been studied by Cohen [2,3]. An $E_k/G/1$ system with finite capacity has been analyzed by using the phase technique by Truslove [18,19] and its asymptotic behaviour has been studied by Hokstad [9]. An $M/M/1$ system with finite capacity has been discussed adequately by Finch [6].

In this paper, the overflows of customers in the $GI/G/1$ queueing system with finite capacity are dealt with under the heavy traffic condition. Overflows occurring as customers arrive to find waiting places fully occupied, are characteristic events of queueing system with finite capacity. Let $Q(t)$ denote the number of customers in the system at time t. The sample path of the process $Q(t)$ are defined as continuous from the left. Then, *the first overflow time*, defined by

$$(1.1)\qquad T(i,N) \equiv \inf\{t \geqq 0 \mid Q(t) = N,\ Q(0) = i\} \qquad N \geqq 2,\ 0 \leqq i \leqq N-1,$$

represents the time at which the number of customers first exceeds the capacity. Note that the first overflow time is invariant with work-conserving service disciplines [11,16]. Let $F_{iN}(t)$ and $f_{iN}(s)$ (Re $s \geqq 0$) denote the d.f. of $T(i,N)$ and its Laplace-Stieltjes transform (L.S.T.), respectively. That is,

$$(1.2)\qquad f_{iN}(s) \equiv \int_0^\infty e^{-st} dF_{iN}(t) , \qquad \text{Re } s \geqq 0 .$$

Saaty [14] has referred to the first overflow time in the analysis of a birth and death process. His result, which corresponds to the first overflow time for the $M/M/1$ system, can be rewritten as

$$(1.3)\qquad f_{iN}(s) = \frac{\lambda(\alpha_+^{i+1} - \alpha_-^{i+1}) - \mu(\alpha_+^{i} - \alpha_-^{i})}{\lambda(\alpha_+^{N+1} - \alpha_-^{N+1}) - \mu(\alpha_+^{N} - \alpha_-^{N})} ,$$

where

$$(1.4)\qquad \alpha_\pm \equiv \frac{(\lambda + \mu + s) \pm \sqrt{(\lambda + \mu + s)^2 - 4\lambda\mu}}{2\lambda} .$$

Consequently the mean first overflow time is given by

$$(1.5)\qquad \mathrm{E}[T(i,N)] = \begin{cases} \dfrac{N - i}{\lambda - \mu} + \dfrac{\lambda}{(\lambda - \mu)^2}\{(\dfrac{\mu}{\lambda})^{N+1} - (\dfrac{\mu}{\lambda})^{i+1}\}, & \text{if } \lambda \neq \mu \\ \dfrac{(N - i)(N + i + 1)}{2\lambda} , & \text{if } \lambda = \mu. \end{cases}$$

In this paper, the first overflow time defined by (1.1) is analyzed by using diffusion model approximation. That is, the d.f. and moments of the first overflow time are derived from the first passage time of a diffusion process approximating $Q(t)$. These results lead to some asymptotic properties of the first overflow time, which are relevant to fluid approximation. Further these results are applied to the analysis of the maximum number of customers in another system with infinite capacity, and yield Laplace transforms (L.T.s) of the d.f. and the mean maximum number of customers. Finally, the accuracy of the diffusion approximations is examined numerically by using the analytical results for a *GI/M/1* system.

2. Preliminaries for Diffusion Approximation

The basic arguement used in diffusion approximation is as follows: Let $\{x(t) \; ; \; t \geqq 0\}$ denote a homogeneous diffusion process and $p(x,t|x_0)$ its probability density function (p.d.f.), i.e.,

(2.1) $$p(x,t|x_0)dx = P\{x \leqq x(t) < x + dx | x(0) = x_0\}.$$

Then, $p(x,t|x_0)$ satisfies the following *Kolmogorov equations*:

(2.2) $$\frac{\partial p}{\partial t} = \frac{1}{2} a \frac{\partial^2 p}{\partial x^2} - b \frac{\partial p}{\partial x}$$

and

(2.3) $$\frac{\partial p}{\partial t} = \frac{1}{2} a \frac{\partial^2 p}{\partial x_0^2} + b \frac{\partial p}{\partial x_0} \; .$$

They are called *the forward (Fokker-Planck) equation*, and *the backward equation*, respectively. The diffusion parameters a and b denote the infinitesmal variance and the mean of the process, respectively. Diffusion approximation means to approximate queue characteristics of general queueing systems through the Kolmogorov equations. It is shown in [12] that diffusion approximation is efficient under the heavy traffic condition. The diffusion parameters a and b are determined by the elementary renewal theorem from asymptotic properties of the characteristic to be approximated [8]. For example, they are given by, for the queue length,

(2.4) $$a = \lambda^3\sigma_a^2 + \mu^3\sigma_s^2$$
$$b = \lambda - \mu \; .$$

Upon equations (2.2) and (2.3), the initial condition

(2.5) $$p(x,0|x_0) = \delta(x - x_0)$$

is imposed, where $\delta(\cdot)$ denotes the Dirac's delta function. Furthermore, appropriate boundary conditions are necessary in order to keep the process in the desired region. A reflecting barrier is used as a boundary condition at $x = 0$ for the queue length. The boundary condition for the reflecting barrier is given by, for (2.2),

$$\left. \frac{1}{2} a \frac{\partial p}{\partial x} - bp \right|_{x=0} = 0, \tag{2.6}$$

and for (2.3), given by

$$\left. \frac{\partial p}{\partial x_0} \right|_{x_0=0} = 0. \tag{2.7}$$

The reflecting barrier is effective under the heavy traffic condition [12].

3. Distribution and Moments of the First Overflow Time

Starting from the backward equation (2.3), we shall derive both the p.d.f. and d.f. of the first overflow time. Following Newell [12], we assume the heavy traffic condition. Since the heavy traffic condition means that the traffic intensity $\rho \equiv \lambda/\mu$ is close to unity, this condition is appropriate to the overflows because these are rare events when ρ is relatively small. Let $x(t)$ be the diffusion process which approximates the queue length $Q(t)$. Define $T_d(x_0,N)$ by

$$T_d(x_0,N) \equiv \inf\{t \geqq 0 \,|\, x(t) = N,\ x(0) = x_0\}, \qquad 0 \leqq x_0 < N, \tag{3.1}$$

where for notational convenience, the initial value i is setted as x_0. From the arguement in Section 2, it follows that $p(x,t|x_0)$ defined by (2.1) satisfies the Kolmogorov's backward equation (2.3) with the initial condition (2.5), where the diffusion parameters are given by (2.4). Moreover, since the process $x(t)$ is restricted within the interval $[0,N]$ and terminates when it reaches at $x = N$, the reflecting barrier is placed at the origin and an absorbing barrier is placed at $x = N$.

Let $F(t|x_0,N)$ and $f(s|x_0,N)$ denote the d.f. of $T_d(x_0,N)$ and its L.S.T. . Then

$$\begin{aligned} \int_0^N p(x,t|x_0)dx &= P\{T_d(x_0,N) > t\} \\ &= 1 - F(t|x_0,N). \end{aligned} \tag{3.2}$$

From (2.3), (2.5) and (3.2), it is easily shown that $f(s|x_0,N)$ satisfies the following ordinary differential equation [4, p.230]: for $0 < x_0 < N$, (3.3) $$\frac{1}{2} a \frac{d^2 f}{dx_0^2} + b \frac{df}{dx_0} = sf.$$

The boundary conditions at $x_0 = 0$ and $x_0 = N$ are:

(3.4) $$\left. \frac{df}{dx_0} \right|_{x_0=0} = 0$$

and

(3.5) $$f(s|N,N) = 1.$$

From (3.3) through (3.5), a caluculation similar to [4, p.233] leads to

(3.6) $$f(s|x_0,N) = exp\{\frac{b}{a}(N - x_0)\} \frac{b\sinh cx_0 + ac\cosh cx_0}{b\sinh cN + ac\cosh cN},$$

where

(3.7) $$c \equiv \sqrt{b^2 + 2as} /a.$$

Clearly, the L.T. $F^*(s|x_0,N)$ of the probability that the overflows occur within t is given by

(3.8) $$F^*(s|x_0,N) = \frac{1}{s} f(s|x_0,N).$$

Next, we shall derive the moments of $T_d(x_0,N)$. For $n = 0,1,\dots,$ let $m_n(x_0,N)$ denote the n-th moment of $T_d(x_0,N)$, where for $n = 0$ we assume $m_0(x_0,N) = 1$. Usually, $m_n(x_0,N)$ can be obtained by

(3.9) $$m_n(x_0,N) = (-1)^n \left. \frac{d^n}{ds^n} f(s|x_0,N) \right|_{s=0}.$$

However, this is very difficult because $f(s|x_0,N)$ given by (3.6) has a complicated form. Therefore, we shall adopt another method [4, p.232]. That is, since $f(s|x_0,N)$ is the moment generating function of $T_d(x_0,N)$, we have

(3.10) $$f(s|x_0,N) = \sum_{n=0}^{\infty} m_n(x_0,N) \frac{(-s)^n}{n!}.$$

Substituting (3.10) in (3.3) and equating coefficients of powers of s, we obtain for $n = 1,2,\dots,$

(3.11) $$\frac{1}{2} a \frac{d^2 m_n}{d x_0^2} + b \frac{dm_n}{dx_0} = - n\, m_{n-1}.$$

Similarly we can deduce the boundary conditions for (3.11) from (3.4) and (3.5). That is, for $n = 1,2,\dots,$

(3.12) $$\left. \frac{dm_n}{dx_0} \right|_{x_0=0} = 0$$

and

(3.13) $$m_n(N,N) = 0.$$

For $n = 1,2$, solving (3.11) with (3.12) and (3.13), we obtain the mean and the variance of $T_d(x_0,N)$ as follows:

(3.14)
$$E[T_d(x_0,N)] \equiv m_1(x_0,N)$$
$$= \begin{cases} \frac{1}{b}(N - x_0) + \frac{a}{2b^2}\{exp(-2bN/a) - exp(-2bx_0/a)\}, & \text{if } b \neq 0 \\ \frac{1}{a}(N^2 - x_0^2), & \text{if } b = 0, \end{cases}$$

and for the second moment of $T_d(x_0,N)$,

(3.15)
$$m_2(x_0,N)$$
$$= \begin{cases} \frac{1}{b^2}(N - x_0)^2 + \frac{a^2}{b^3}[(N - x_0) + 4Nexp(-2bN/a) \\ \quad - (N + x_0)\{exp(-2bN/a) + exp(-2bx_0/a)\}] \\ \quad + \frac{a^2}{2b^4}\{2 + exp(-2bN/a)\}\{exp(-2bN/a) - exp(-2bx_0/a)\}, & \text{if } b \neq 0 \\ \frac{1}{3a^2}(N^2 - x_0^2)(5N^2 - x_0^2), & \text{if } b = 0. \end{cases}$$

The derivation of the above results is shown in Appendix. From (3.14) and (3.15), the variance of $T_d(x_0,N)$ is given by

(3.16)
$$V[T_d(x_0,N)] \equiv m_2(x_0,N) - \{m_1(x_0,N)\}^2$$
$$= \begin{cases} \frac{a^2}{4b^4}\{exp(-2bN/a) + exp(-2bx_0/a) + 4\}\{exp(-2bN/a) - exp(-2bx_0/a)\} \\ \quad + \frac{a}{b^3}[(N - x_0) + 2\{Nexp(-2bN/a) - x_0exp(-2bx_0/a)\}], & \text{if } b \neq 0 \\ \frac{2}{3a^2}(N^4 - x_0^4), & \text{if } b = 0. \end{cases}$$

Remark 3.1. Sweet and Hardin [17] have derived the p.d.f. which corresponds to that of $T_d(x_0,N)$ by solving the Kolmogorov's forward equation (2.2) with similar boundary conditions. However, because their solution has an infinite series form, it seems impossible to derive the moments of $T_d(x_0,N)$ in simple forms.

4. Modifications of the Boundary Condition

In the preceding section we have considered the reflecting barrier as the boundary condition at the origin. However, if $\rho < 1$, then discontinuity of the path $Q(t)$ at the origin quite effects the solution, and hence (3.14) seems to be slightly underestimated from the true value. That is, if the process $Q(t)$ reaches to the boundary $x = 0$, the process remains at $x = 0$ for a residual interarrival time. When a new customer arrives at the system, the process jumps to $x = 1$ and thereafter the process starts from scratch. Under the assumption that the arrival process is the recurrent process, it may be difficult to obtain the distribution of the sojourn time at $x = 0$. However, if the arrival process is a Poisson process, then it is known that the sojourn time is exponentially distributed with parameter λ. Feller [5] has called this process an elementary return process.

Taking account of the sojourn time at the origin, we consider two modifications of the boundary condition. The first one is a heuristic modification of changing the location of the reflecting barrier. The second one is a modification of using the elementary return process. It follows from the definition of the elementary return process that the latter is appropriate for the *M/G/1* system. In the following the process $Q(t)$ having the reflecting barrier at the origin is called the reflecting barrier process. Moreover, when A denotes a characteristic for the reflecting barrier process, then $\hat{A}$ and $\tilde{A}$ denote that for the modified reflecting barrier process and the elementary return process, respectively.

4.1. Modified reflecting barrier process

Since it may be difficult in general to obtain the distribution of the sojourn time at the origin, it is assumed that the sojourn times have the same distribution as the stationary residual life [13] of the interarrival time. Under this assumption the mean sojourn time is given by $(\lambda^2\sigma_a^2 + 1)/2\lambda$. Let us consider a modification of changing the location of the reflecting barrier at $x = 0$ in accordance with this time. It should be noted that this modification is equivalent to changing the location of the absorbing barrier at $x = N$, since the diffusion process is spartially homogeneous. Place the reflecting barrier at $x = -\Delta$, or equivalently, place the absorbing barrier at $x = N + \Delta$, where

$$(4.1) \qquad \Delta \equiv \frac{\lambda^2\sigma_a^2 + 1}{4}.$$

Then, the relations $\hat{f}(s|x_0,N) = f(s|x_0+\Delta,N+\Delta)$, $E[\hat{T}_d(x_0,N)] = E[T_d(x_0+\Delta,N+\Delta)]$ and $V[\hat{T}_d(x_0,N)] = V[T_d(x_0+\Delta,N+\Delta)]$ imply that

(4.2) $$\hat{f}(s|x_0,N) = exp\{\frac{b}{a}(N - x_0)\} \frac{b\sinh c(x_0 + \Delta) + ac\cosh c(x_0 + \Delta)}{b\sinh c(N + \Delta) + ac\cosh c(N + \Delta)},$$

(4.3) $$E[\hat{T}_d(x_0,N)] = \begin{cases} \frac{1}{b}(N - x_0) + \frac{a}{2b^2} exp(-2b\Delta/a)\{exp(-2bN/a) - exp(-2bx_0/a)\}, & \text{if } b \neq 0 \\ \frac{1}{a}(N^2 - x_0^2) + \frac{2\Delta}{a}(N - x_0), & \text{if } b = 0, \end{cases}$$

and

(4.4) $$V[\hat{T}_d(x_0,N)] = \begin{cases} \frac{a^2}{4b^4} exp(-2b\Delta/a)[exp\{-2b(N + \Delta)/a\} + exp\{-2b(x_0 + \Delta)/a\} + 4] \\ \quad\cdot\{exp(-2bN/a) - exp(-2bx_0/a)\} + \frac{a}{b^3}\big((N - x_0) \\ \quad + 2[(N + \Delta)exp\{-2b(N + \Delta)/a\} - (x_0 + \Delta)exp\{-2b(x_0 + \Delta)/a\}]\big), & \text{if } b \neq 0 \\ \frac{2}{3a^2}\{(N + \Delta)^4 - (x_0 + \Delta)^4\}, & \text{if } b = 0. \end{cases}$$

Further, it follows from (1.5) that for $\rho = 1$, the value of $E[\hat{T}_d(x_0,N)]$ for the *M/M/1* system agrees with the exact one. It will be shown in Section 6 that the modified solutions are fairly accurate.

4.2. Elementary return process

Using the elementary return process, we shall precisely describe the boundary condition at $x = 0$ for the *M/G/1* system. That is, for equation (3.3), we replace the boundary condition (3.4) by

(4.5) $$\tilde{f}(s|0,N) = \frac{\lambda}{\lambda + s} \tilde{f}(s|1,N),$$

(see [4, p.232]). Solving (3.3) with the boundary conditions (4.5) and (3.5), we obtain

(4.6) $$\tilde{f}(s|x_0,N) = exp\{\frac{b}{a}(N - x_0)\} \frac{(\lambda+s)\sinh cx_0 - \lambda exp(-b/a)\sinh c(x_0-1)}{(\lambda+s)\sinh cN - \lambda exp(-b/a)\sinh c(N-1)},$$

By the method described in Section 3, the mean of $\tilde{T}_d(x_0,N)$ can be derived as follows: for $n = 1,2,...,$ the substitution of (3.10) into (4.5) leads to

(4.7) $$\tilde{m}_n(0,N) = \tilde{m}_n(1,N) + \frac{n}{\lambda}\tilde{m}_{n-1}(0,N).$$

This equation represents the boundary condition of (3.11) for the elementary return at $x = 0$. In particular, for $n = 1$ we have

$$\tilde{m}_1(0,N) = \tilde{m}_1(1,N) + \frac{1}{\lambda} . \tag{4.8}$$

It seems that this expression suggests an average behaviour of the elementary return process. Solving (3.11) for $n = 1$ with the boundary conditions (4.8) and (3.13), we obtain

$$\mathrm{E}[\tilde{T}_d(x_0,N)] \equiv \tilde{m}_1(x_0,N) \tag{4.9}$$

$$= \begin{cases} \frac{1}{b}(N - x_0) + \frac{1}{\rho b\{1 - exp(-2b/a)\}}\{exp(-2bN/a) - exp(-2bx_0/a)\}, & \text{if } b \neq 0 \\ \frac{1}{a}(N^2 - x_0^2) + (\frac{1}{\lambda} - \frac{1}{a})(N - x_0), & \text{if } b = 0. \end{cases}$$

The derivation of this equation is shown in Appendix.

Remark 4.1. It is easily shown that under the heavy traffic condition (3.14), (4.3) and (4.9) are equivalent except for the term $o(\frac{1}{\lambda})$. This is because the sojourn time at $x = 0$ decreases to zero as the value of increases, and because the process hardly depends on the boundary conditions at $x = 0$ under the heavy traffic.

Remark 4.2. Fluid approximation [11] is a rough approximation which disregards both the randomness of the arrival and the service processes and the discreteness of the path $Q(t)$. In fluid approximation, we regard cumulative numbers of arrivals and departures as fluids which pour into a reservoir and pour out from it, respectively. Therefore, if $b \leqq 0$, then the amount of fluid in the reservoir never exceeds its capacity, whereas if $b > 0$, then the capacity will be exceed within a finite time since the amount of fluid increases by b per unit time. Let $T_f(x_0,N)$ be the fluid approximation of $T(x_0,N)$. Then from the above arguement, we obtain

$$T_f(x_0,N) = \begin{cases} \frac{1}{b}(N - x\ ), & \text{if } b > 0 \\ + \infty & \text{if } b \leqq 0. \end{cases} \tag{4.10}$$

Note that for $b > 0$, $T_f(x_0,N)$ corresponds to the first term of (3.14), (4.3) and (4.9).

Remark 4.3. From (3.14), (4.3) and (4.9), for $b > 0$, we have

$$\lim_{a\downarrow 0} \mathrm{E}[T_d(x_0,N)] = \lim_{a\downarrow 0} \mathrm{E}[\hat{T}_d(x_0,N)] = \lim_{a\downarrow 0} \mathrm{E}[\tilde{T}_d(x_0,N)] = T_f(x_0,N). \tag{4.11}$$

Under the heavy traffic condition, we have

$$\mathrm{E}[T_d(x_0,N)] = T_f(x_0,N) + o(\frac{1}{\lambda}),$$
$$\mathrm{E}[\hat{T}_d(x_0,N)] = T_f(x_0,N) + o(\frac{1}{\lambda}),$$
and $$\mathrm{E}[\tilde{T}_d(x_0,N)] = T_f(x_0,N) + o(\frac{1}{\lambda}).$$
Furthermore,
$$\mathrm{V}[T_d(x_0,N)] = o(\frac{1}{\lambda}),$$
and $$\mathrm{V}[\hat{T}_d(x_0,N)] = o(\frac{1}{\lambda}).$$
These relations imply that the solution by diffusion approximation converges to that by fluid approximation as $\lambda \to +\infty$.

5. Application of the First Overflow Time

As an application of the first overflow time, we shall investigate a transient behaviour of the maximum number of customers in a system with infinite capacity. Suppose that the arrival and the service processes of this system are identical with the model discussed so far. The maximum number of customers up to time t is defined as

(5.1) $$M(t|x_0) \equiv \sup_{0 \leqq \tau \leqq t} \{Q(\tau)|Q(0) = x_0\}.$$

Let $M_d(t|x_0)$ denote the diffusion approximation of $M(t|x_0)$. Then we can derive the d.f. of $M_d(t|x_0)$ from the following relation [13],

(5.2) $$\{M_d(t|x_0) \leqq y\} = \{T_d(x_0,y) \geqq t\}.$$

Hence,

(5.3) $$L(y,t|x_0) \equiv P\{M_d(t|x_0) \leqq y\}$$
$$= 1 - F(t|x_0,y).$$

For $\mathrm{Re}\ s \geqq 0$, let $L^*(y,s|x_0)$ denote the L.T. of $L(y,t|x_0)$. From (5.3), we obtain

(5.4) $$L^*(y,s|x_0) = \frac{1}{s}\{1 - f(s|x_0,y)\}.$$

Moreover, the mean maximum number of customers is given by

(5.5) $$\mathrm{E}[M_d(t|x_0)] \equiv \int_{x_0}^{\infty} y d_y L(y,t|x_0)$$
$$= x_0 U(t) + \int_{x_0}^{\infty} F(t|x_0,y)dy,$$

where

$$U(t) \equiv \begin{cases} 1, & \text{if } t > 0 \\ 0, & \text{if } t \leqq 0. \end{cases}$$

Combining (3.8) and (5.5) yields

$$(5.6) \qquad \int_0^\infty e^{-st} \mathrm{E}[M_d(t|x_0)]dt = \frac{1}{s}\{x_0 + \int_{x_0}^\infty f(s|x_0,y)dy\}.$$

Remark 5.1. In (5.4), the reflecting barrier process is used. However, if the arrival process is a Poisson process, then $\tilde{f}$ had better been replaced by f. Further for small N, f may be replaced by $\hat{f}$.

Remark 5.2. Because of the complicated form of $f(s|x_0,y)$, to integrate the right hand side of (5.6) is very difficult, but if $b = 0$ for the reflecting barrier process, then the integration can be executed. That is, from (3.6) and (5.6), for $b = 0$ we obtain

$$(5.7) \qquad \int_0^\infty e^{-st} \mathrm{E}[M_d(t|x_0)]dt = \frac{1}{s}[x_0 + \frac{1}{c}\{\pi - 2\arctan(e^{cx_0})\}\cosh cx_0].$$

Moreover, if $x_0 = 0$ in (5.7), then the analytical inversion of the transform leads to

$$(5.8) \qquad \mathrm{E}[M_d(t|0)] = \sqrt{\frac{a\pi t}{2}}.$$

Therefore it can be observed that if the number of customers is equal to zero at $t = 0$ and the traffic intensity $\rho = 1$, the mean maximum number of customers up to time t is proportional to the square root of t.

6. Comparisons with the Analytical Results

6.1. Analytical results for a GI/M/1 system

It is quite difficult to analyze the first overflow time for the *GI/G/1* queueing system. However, for the case that the service time distribution is negative exponential, that is, for the *GI/M/1* system, the first overflow time can be investigated by using the regenerative property. Denote the i-th arrival epoch to the system by t_i $(i = 1,2,\ldots)$. Assume that the inter-arrival times $a_i \equiv t_{i+1} - t_i$ $(i = 1,2,\ldots)$ are i.i.d. random variables with d.f. A. It should be noted that the distribution of $a_0 \equiv t_1$ does not necessarily agree with A. Since the initial time $t = 0$ may not be an arrival epoch, it is assumed to keep generality that the distribution of a_0 is A_0. For this system, the process $Q(t)$ is regenerative with respect to the delayed renewal process $\{N_t \; ; \; t \geqq 0\}$ with

$$N_t \equiv \max\{n \mid t_n < t\}, \qquad \text{for } t > 0$$
$$N_0 \equiv 0. \tag{6.1}$$

Since the service times are negative exponentially distributed, it follows that for $i = 0,1,\ldots,N-2$, $n \geqq 1$,

$$q_{ij}(t) \equiv P\{Q(t_n + t) = j \mid t_n + t < t_{n+1},\, Q(t_n) = i\} \tag{6.2}$$
$$= \begin{cases} \displaystyle\int_0^t \frac{(\mu x)^i}{i!} e^{-\mu x} \mu dx, & \text{for } j = 0 \\ \displaystyle\frac{(\mu t)^{i+1-j}}{(i+1-j)!} e^{-\mu t}, & \text{for } j = 1,\ldots,i+1 \\ 0, & \text{for } j = i+2,\ldots\ , \end{cases}$$

(see, e.g., [2, p.204]) and that for $i = 0,1,\ldots,N-1$,

$$p_{ij}(t) \equiv P\{Q(t) = j \mid t < t_1,\, Q(0) = i\} \tag{6.3}$$
$$= \begin{cases} \delta_{0j}, & \text{for } i = 0 \\ q_{i-1,j}(t), & \text{for } i \neq 0, \end{cases}$$

where δ denotes the Kronecker's delta.

Since the overflows occur at the arrival epoch and $Q(t)$ is left-continuous, if $Q(t_n) = N - 1$, then $T(\cdot,N) = t_n$. Furthermore, the value of $Q(t)$ never increases between two successive arrivals or during $(0,t_1)$. Hence, the d.f. of the first overflow time can be derived from that of a entrance time to $N - 1$ on the epochs just before arrivals. For the sake of convenience, let $Q_n = Q(t_n)$. Define the d.f. of the entrance time to $N - 1$ from the first arrival epoch as

$$\Phi^{(1)}_{j,N-1}(t) \equiv P\{Q_2 = N - 1,\, t_2 < t \mid Q_1 = j\},$$
$$\Phi^{(n)}_{j,N-1}(t) \equiv P\{Q_{n+1} = N - 1,\, t_{n+1} < t,\, Q_h \neq N - 1,\, h = 2,\ldots,n \mid Q_1 = j\} \qquad \text{for } n \geqq 2, \tag{6.4}$$

and

$$\Phi_{j,N-1}(t) \equiv \sum_{n=1}^{\infty} \Phi^{(n)}_{j,N-1}(t). \tag{6.5}$$

It is verified from the renewal theoretic consideration that there is a relation between $F_{iN}(t)$ and $\Phi_{j,N-1}(t)$ such that for $i = 0,1,\ldots,N-1$,

$$F_{iN}(t) = \int_0^t p_{i,N-1}(u)dA_0(u) + \sum_{j=0}^{N-2} \int_0^t p_{ij}(u)\Phi_{j,N-1}(t - u)dA_0(u). \tag{6.6}$$

Taking the L.S.T. of both sides of (6.6) yields

$$(6.7) \qquad f_{iN}(s) = \pi_{i,N-1}(s) + \sum_{j=0}^{N-2} \pi_{ij}(s)\phi_{j,N-1}(s),$$

where

$$(6.8) \qquad \pi_{ij}(s) \equiv \int_0^\infty e^{-st} p_{ij}(t)dA_0(t),$$

and

$$(6.9) \qquad \phi_{j,N-1}(s) \equiv \int_0^\infty e^{-st} d\Phi_{j,N-1}(t).$$

For $\Phi_{j,N-1}^{(n)}(t)$, $j = 0,1,\ldots,N-2$, the following integral relations hold.

$$(6.10) \qquad \Phi_{j,N-1}^{(1)}(t) = \int_0^t q_{j,N-1}(u)dA(u),$$

and for $n \geqq 2$,

$$(6.11) \qquad \Phi_{j,N-1}^{(n)}(t) = \sum_{k=0}^{N-2} \int_0^t q_{jk}(u)\Phi_{k,N-1}^{(n-1)}(t-u)dA(u).$$

Combining (6.5), (6.10) and (6.11) leads to

$$(6.12) \qquad \Phi_{j,N-1}(t) = \int_0^t q_{j,N-1}(u)dA(u) + \sum_{k=0}^{N-2} \int_0^t q_{jk}(u)\Phi_{k,N-1}(t-u)dA(u).$$

Taking the L.S.T. of (6.12) yields

$$(6.13) \qquad \phi_{j,N-1}(s) = \alpha_{j,N-1}(s) + \sum_{k=0}^{N-2} \alpha_{jk}(s)\phi_{k,N-1}(s),$$

where

$$(6.14) \qquad \alpha_{jk}(s) \equiv \int_0^\infty e^{-st} q_{jk}(t)dA(t).$$

Consequently, the above results can be summarized as follows:

Proposition 1. *Let* $\{\phi_j(s)\ ;\ j = 0,1,\ldots,N-2\}$ *be a solution of a system of linear equations*

$$\phi_i(s) = \alpha_{i,N-1}(s) + \sum_{j=0}^{N-2} \alpha_{ij}(s)\phi_j(s) \qquad i = 0,1,\ldots,N-2$$

with $\alpha_{ij}(s)$ *given by* (6.14). *Then, the L.S.T. of the d.f. of the first overflow time for the GI/M/1 system is given by*

$$f_{iN}(s) = \pi_{i,N-1}(s) + \sum_{j=0}^{N-2} \pi_{ij}(s)\phi_j(s),$$

where $\pi_{ij}(s)$ *are given by* (6.8).

Denote, for $i = 0,1,\ldots,N-1$, $j = 0,1,\ldots,N-2$,

$$(6.15) \qquad \overline{\alpha}_{ij} \equiv -\frac{d}{ds}\alpha_{ij}(s)\bigg|_{s=0}$$

and

$$(6.16)\qquad \overline{\pi}_{ij} \equiv -\frac{d}{ds}\pi_{ij}(s)\Big|_{s=0}.$$

Then, from Proposition 1, the next proposition holds.

Proposition 2. *Let* $\{\overline{\phi}_j,\ {}_2\overline{\phi}_j\ ;\ j=0,1,\ldots,N-2\}$ *be a solution of two systems of linear equations*

$$(6.17)\qquad \overline{\phi}_i = \frac{1}{\lambda} + \sum_{j=0}^{N-2}\alpha_{ij}(0)\overline{\phi}_j$$

and

$$(6.18)\qquad {}_2\overline{\phi}_i = \{(\sigma_a^2 + \frac{1}{\lambda^2}) + 2\sum_{j=0}^{N-2}\overline{\alpha}_{ij}\overline{\phi}_j\} + \sum_{j=0}^{N-2}\alpha_{ij}(0)\,{}_2\overline{\phi}_j,$$

for $i = 0,1,\ldots,N-2.$

Then, the first and second moments of the first overflow time for the GI/M/1 system are given by

$$(6.19)\qquad \mathrm{E}[T(i,N)] = \nu_1 + \sum_{j=0}^{N-2}\pi_{ij}(0)\overline{\phi}_j,$$

and

$$(6.20)\qquad \mathrm{E}[\{T(i,N)\}^2] = \nu_2 + 2\sum_{j=0}^{N-2}\overline{\pi}_{ij}\overline{\phi}_j + \sum_{j=0}^{N-2}\pi_{ij}(0)\,{}_2\overline{\phi}_j,$$

respectively, where $\nu_i \equiv \int_0^\infty x^i dA_0(x)$, $i = 1,2.$

Proof: Differenciating (6.7) and (6.13) with respect to s and using the relations

$$\sum_{j=0}^{N-1}\alpha_{ij}(s) = \int_0^\infty e^{-st}dA(t)$$

and

$$\sum_{j=0}^{N-1}\pi_{ij}(s) = \int_0^\infty e^{-st}dA_0(t),$$

we obtain the desired results immediately. □

Note that the above results are also applied to the analysis of the maximum number of customers in the system by the same method as in Section 5. Final forms corresponding to (5.4) and (5.6) are respectively

$$(6.21)\qquad \int_0^\infty e^{-st}P\{M(t|i) \leqq j\}dt = \frac{1}{s}\{1 - f_{ij}(s)\},$$

and

$$(6.22)\qquad \int_0^\infty e^{-st}\mathrm{E}[M(t|i)]dt = \frac{1}{s}\{i + \sum_{i=j+1}^{\infty}f_{ij}(s)\}.$$

Remark 6.1. Let $\alpha(s)$ and I be the $(N-1)\times(N-1)$ matrices of elements $\alpha_{ij}(s)$ and δ_{ij}, respectively. Then, it is well known that for $\mathrm{Re}\ s \geqq 0$, the matrix $I - \alpha(s)$ is nonsingular. Hence, each system of linear equations

(6.13), (6.17) and (6.18) has a unique solution.

Remark 6.2. If the arrival distribution A is the Erlang distribution with phase k, i.e.,

$$A(t) = \int_0^t \frac{(\lambda k)^k x^{k-1}}{(k-1)!} e^{-\lambda k x} dx,$$

then the explicit form of $\alpha_{ij}(s)$ can be caluculated as

$$(6.23) \qquad \alpha_{ij}(s) = \begin{cases} \left(\dfrac{\lambda}{s+\lambda k}\right)^k - \sum_{j=1}^{N-1} \alpha_{ij}(s), & \text{for } j = 0 \\ \binom{i-j+k}{k-1} \left(\dfrac{\mu}{s+\mu+\lambda k}\right)^{i-j+1} \left(\dfrac{\lambda k}{s+\mu+\lambda k}\right)^k, & \text{for } j = 1,\dots,i+1 \\ 0, & \text{for } j = i+2,\dots . \end{cases}$$

Moreover, if $A_0 = A$, then we have

$$(6.24) \qquad \pi_{ij}(s) = \begin{cases} \delta_{j0} \left(\dfrac{\lambda k}{s+\lambda k}\right)^k, & \text{for } i = 0 \\ \alpha_{i-1,j}(s), & \text{for } i \neq 0. \end{cases}$$

6.2. Numerical results

In order to examine the accuracy of the diffusion approximation, we shall numerically compare the approximate results obtained in Section 3 through 5 with the above analytical results. For computational convenience, consider the standard $E_k/M/1$ system in which the d.f.s A and A_0 are the common Erlang distribution with phase k. The mean first overflow time given by (3.14), (4.3) and (4.9), and its relative errors for the $M/M/1$ system are shown in Tables 1 and 2, where the relative error means

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

Tables 3 and 4 show the variance of the first overflow time given by (3.16) and (4.4), and its relative errors for the $M/M/1$ system. Tables 5 and 6 show the mean and the variance of the first overflow time and their relative errors for the $E_2/M/1$ system and for the $E_5/M/1$ system, respectively. From these tables, it is concluded about the accuracy of the diffusion approximation with respect to the moments that:

i) When the initial value i is close to N, all the moments $E[T_d(i,N)]$ $E[\hat{T}_d(i,N)]$, $E[\tilde{T}_d(i,N)]$, $V[T_d(i,N)]$ and $V[\hat{T}_d(i,N)]$ become accurate.

ii) For small k, $E[\hat{T}_d(i,N)]$ is more accurate than $E[T_d(i,N)]$.

iii) For large k, $E[T_d(i,N)]$ is more accurate than $E[\hat{T}_d(i,N)]$ for the case $\rho > 1.0$.

iv) The relative errors of $V[T_d(i,N)]$ and $V[\hat{T}_d(i,N)]$ are greater than those of $E[T_d(i,N)]$ and $E[\hat{T}_d(i,N)]$. It should be noted, however, that the relative errors of the corresponding standard deviation are about the same as those of the means.

Tables 1 through 6 are dealt with only the case $\rho = 0.95$ and $\rho = 1.05$. In order to examine the accuracy of the diffusion approximation for the other value of ρ, the mean and the variance of the first overflow time are shown for the $E_2/M/1$ system with $i = 0$ in Figure 1. It follows from Figure 1 that the approximate solutions behave similarly to the exact ones even in the light traffic.

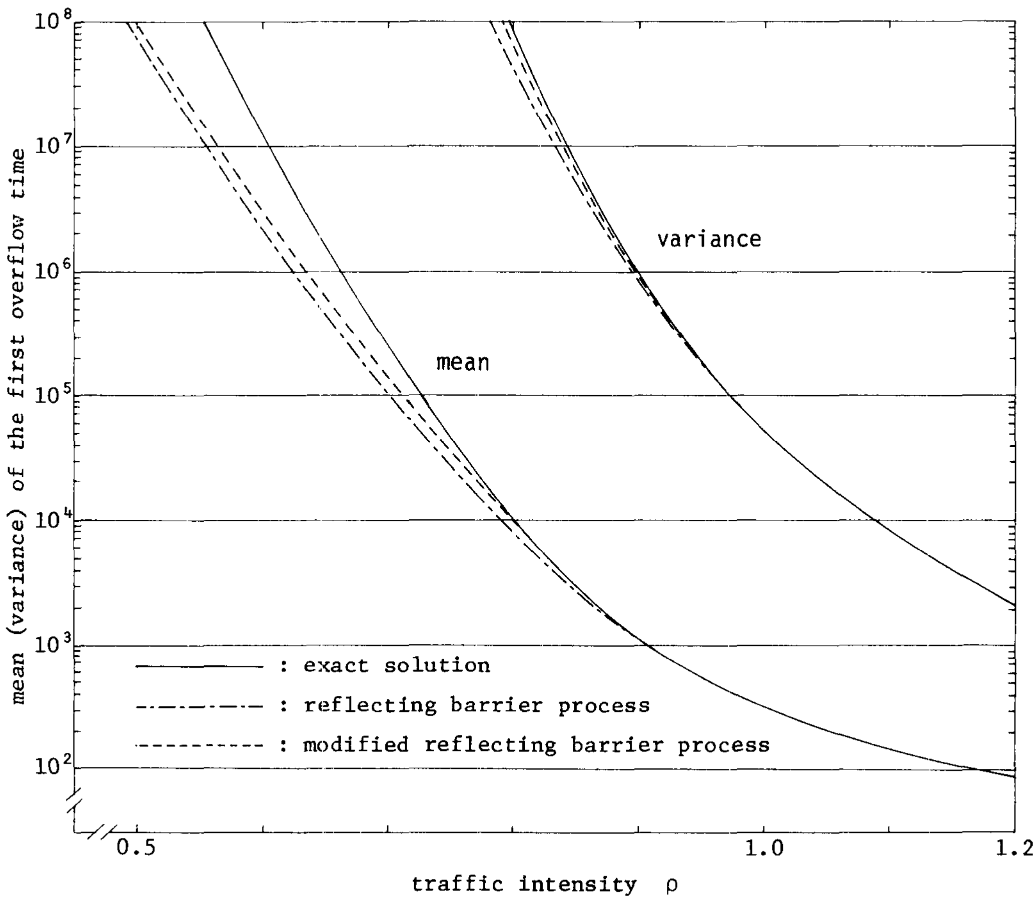

Figure 1. The mean and the variance of the first overflow time for the $E_2/M/1$ system ($N = 21$, $i = 0$, $\frac{1}{\mu} = 1.0$).

Furthermore, many other numerical results lead to the following statement:

v) All the moments become accurate for the case $\rho > 0.9$.

vi) The relative errors of the moments increase as N decreases. Even in such cases, $E[\hat{T}_d(i,N)]$ and $E[\tilde{T}_d(i,N)]$ are quite more accurate than $E[T_d(i,N)]$.

Next, we shall investigate the d.f. of the first overflow time. However, the expressions (1.2), (3.6), (4.2) and (4.6) for the L.S.T. of the d.f. are complicated and their analytical inversions appear to be impossible. Hence, we shall caluculate the d.f.s of the first overflow time by using a routine of numerical L.T. inversion. Two different methods are adopted to invert the L.S.T.s, that is, Gaver's method [7] and the method of Bellman, Kalaba and Lockett [1]. The reason for using different methods is due to a mutual check on the agreement of the results, since the performance of a method is highly dependent on the nature of the original function [1]. It follows from the numerical tests that Gaver's method is more accurate than the method of

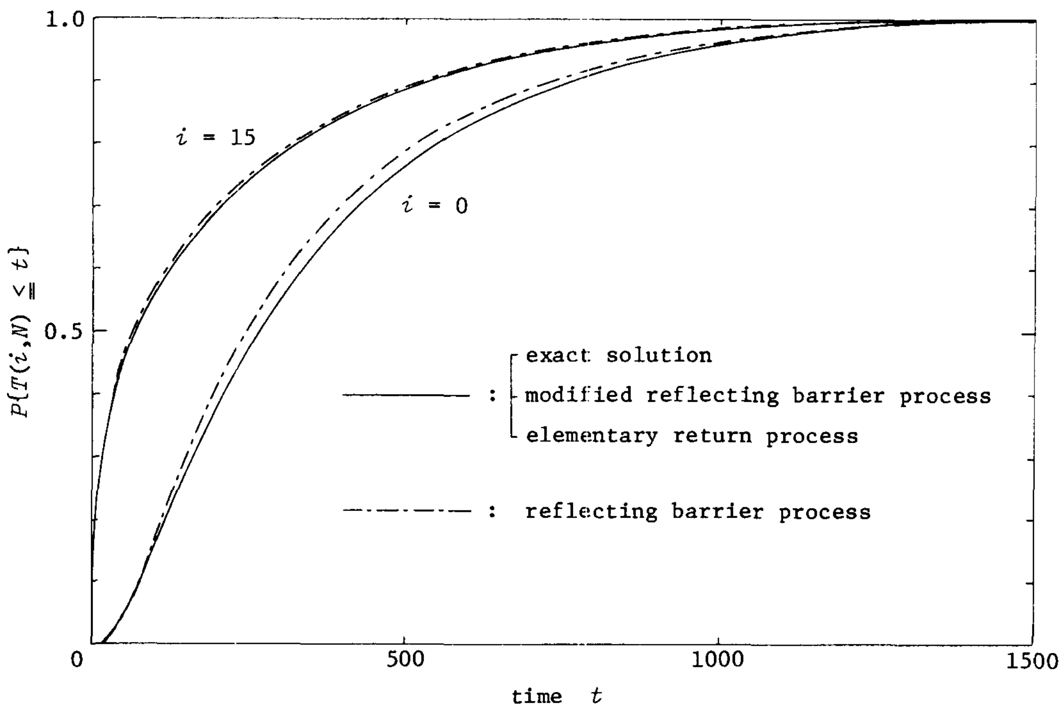

Figure 2. The d.f. of the first overflow time for the $M/M/1$ system ($N = 21$, $\frac{1}{\mu} = 1.0$, $\rho = 0.95$).

Bellman et al.. Therefore, the former is selscted to use in the following, and the latter is used for checking the results. Since the L.S.T. $f_{iN}(s)$ for $k \neq 1$ is derived from a solution of the system of linear equations, the numerical inversion seems to be fairly difficult. Hence, we restrict our caluculations to the case of $k = 1$, i.e., the *M/M/1* system. Figure 2 illustrates the d.f. of the first overflow time for the case of $N = 21$, $i = 0$ or 15, and $\rho = 0.95$.

By using the same routine of the L.T. inversion, we shall caluculate the d.f. of the maximum number of customers in the system. Inverting numerically (5.4) and (6.21), we obtain Figure 3 which shows the d.f. of the maximum number of customers in the *M/M/1* system at $t = 20$ or 40. The d.f. at any other time can be caluculated in much the same way.

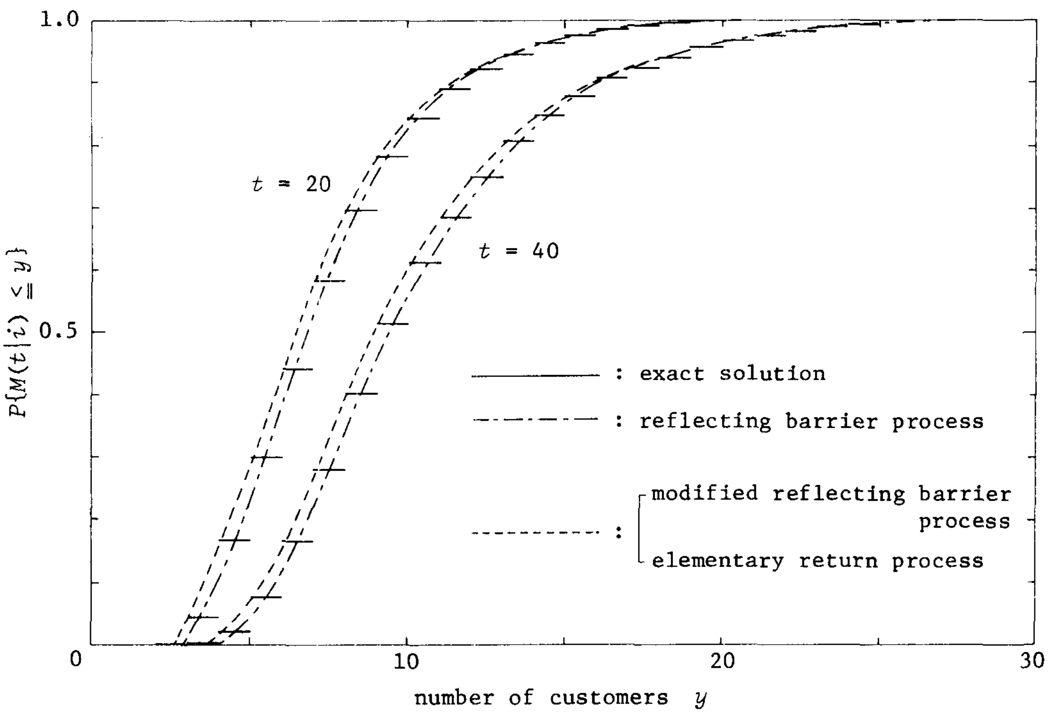

Figure 3. The d.f. of the maximum number of customers in the *M/M/1* system ($i = 0$, $\rho = 0.95$).

Table 1. The mean first overflow time and its relative errors for the *M/M/1* system ($N = 21$, $\frac{1}{\mu} = 1.0$).

ρ	i	$E[T(i,N)]$	$E[T_d(i,N)]$	relative error(%)	$E[\hat{T}_d(i,N)]$	relative error(%)	$E[\tilde{T}_d(i,N)]$	relative error(%)
0.95	0	354.530	334.896	(-5.538)	354.503	(-0.008)	354.438	(-0.025)
	5	337.588	320.906	(-4.941)	337.552	(-0.011)	337.496	(-0.027)
	10	286.457	273.598	(-4.489)	286.418	(-0.013)	286.370	(-0.030)
	15	191.142	183.235	(-4.137)	191.111	(-0.016)	191.081	(-0.032)
	20	38.726	37.232	(-3.858)	38.719	(-0.019)	38.713	(-0.035)
1.05	0	163.576	157.196	(-3.901)	163.528	(-0.030)	163.561	(-0.009)
	5	150.166	145.939	(-2.818)	150.129	(-0.025)	150.151	(-0.010)
	10	118.011	115.467	(-2.156)	117.986	(-0.022)	118.000	(-0.012)
	15	71.170	69.950	(-1.713)	71.156	(-0.019)	71.162	(-0.011)
	20	12.821	12.641	(-1.400)	12.818	(-0.017)	12.820	(-0.007)

Table 2. The mean first overflow time and its relative errors for the *M/M/1* system ($N = 5$, $\frac{1}{\mu} = 1.0$).

ρ	i	$E[T(i,N)]$	$E[T_d(i,N)]$	relative error(%)	$E[\hat{T}_d(i,N)]$	relative error(%)	$E[\tilde{T}_d(i,N)]$	relative error(%)
0.95	0	16.942	13.990	(-17.423)	16.950	(0.052)	16.939	(0.016)
	1	15.890	13.468	(-15.236)	15.896	(0.042)	15.886	(0.017)
	2	13.729	11.867	(-13.561)	13.733	(0.034)	13.726	(0.018)
	3	10.401	9.128	(-12.239)	10.404	(0.029)	10.400	(0.019)
	4	5.847	5.194	(-11.168)	5.848	(0.024)	5.846	(0.019)
1.05	0	13.410	11.261	(-16.026)	13.399	(-0.082)	13.408	(0.015)
	1	12.458	10.781	(-13.459)	12.449	(-0.072)	12.456	(0.014)
	2	10.598	9.371	(-11.574)	10.591	(-0.064)	10.597	(0.013)
	3	7.875	7.077	(-10.131)	7.870	(-0.058)	7.874	(0.013)
	4	4.329	3.940	(-8.991)	4.327	(-0.053)	4.328	(0.014)

Table 3. The variance of the first overflow time and its relative errors for the $M/M/1$ system ($N = 21,\ \frac{1}{\mu} = 1.0$).

ρ	i	$V[T(i,N)]$	$V[T_d(i,N)]$	relative error(%)	$V[\hat{T}_d(i,N)]$	relative error(%)
0.95	0	95996.9	85301.1	(-11.142)	95920.8	(-0.079)
	5	95792.4	85166.2	(-11.093)	95719.2	(-0.077)
	10	92662.7	82625.6	(-10.832)	92597.1	(-0.071)
	15	76274.2	68409.6	(-10.311)	76225.1	(-0.064)
	20	20242.8	18294.2	(-9.626)	20230.6	(-0.060)
1.05	0	15432.8	14258.3	(-7.610)	15397.6	(-0.228)
	5	15312.9	14176.5	(-7.421)	15280.1	(-0.214)
	10	14525.0	13172.7	(-6.765)	14102.1	(-0.187)
	15	10280.7	9674.9	(-5.893)	10264.4	(-0.159)
	20	2226.5	2114.3	(-5.039)	2223.5	(-0.135)

Table 4. The variance of the first overflow time and its relative errors for the $M/M/1$ system ($N = 5,\ \frac{1}{\mu} = 1.0$).

ρ	i	$V[T(i,N)]$	$V[T_d(i,N)]$	relative error(%)	$V[\hat{T}_d(i,N)]$	relative error(%)
0.95	0	204.446	134.941	(-33.996)	201.782	(-1.303)
	1	203.338	134.759	(-33.727)	200.849	(-1.224)
	2	196.336	131.895	(-33.822)	194.199	(-1.088)
	3	172.983	118.862	(-31.287)	171.376	(-0.929)
	4	116.004	81.950	(-29.365)	115.107	(-0.774)
1.05	0	119.891	81.797	(-31.774)	117.503	(-1.992)
	1	118.984	81.644	(-31.382)	116.755	(-1.873)
	2	113.798	79.448	(-30.185)	111.888	(-1.679)
	3	98.151	70.356	(-28.319)	96.718	(-1.460)
	4	63.613	46.994	(-26.125)	62.817	(-1.252)

Table 5. The mean and the variance of the first overflow time and their relative errors for the $E_2/M/1$ system (N = 21, $\frac{1}{\mu}$ = 1.0).

ρ	i	$E[T(i,N)]$	$E[T_d(i,N)]$	relative error(%)	$E[\hat{T}_d(i,N)]$	relative error(%)
0.95	0	530.331	510.009	(-3.832)	533.957	(0.684)
	5	509.539	490.971	(-3.644)	511.853	(0.454)
	10	439.555	423.899	(-3.562)	440.479	(0.210)
	15	300.035	289.409	(-3.542)	299.951	(-0.028)
	20	62.532	62.298	(-3.573)	62.366	(-0.266)
1.05	0	193.305	191.958	(-0.696)	197.497	(2.170)
	5	177.631	177.218	(-0.232)	180.686	(1.720)
	10	138.698	138.645	(-0.039)	140.621	(1.386)
	15	82.848	82.900	(0.063)	83.801	(1.150)
	20	14.767	14.784	(0.116)	14.911	(0.974)

ρ	i	$V[T(i,N)]$	$V[T_d(i,N)]$	relative error(%)	$V[\hat{T}_d(i,N)]$	relative error(%)
0.95	0	220291.4	205129.2	(-6.842)	225603.2	(2.411)
	5	220008.5	204966.7	(-6.837)	225258.8	(2.386)
	10	214458.3	199832.3	(-6.820)	219222.8	(2.222)
	15	180776.3	168452.2	(-6.817)	184080.7	(1.828)
	20	50210.5	46759.4	(-6.873)	50822.2	(1.218)
1.05	0	19993.4	20176.3	(0.890)	21298.7	(6.502)
	5	19853.5	20037.8	(0.928)	21117.2	(6.365)
	10	18258.0	18446.1	(1.030)	19330.0	(5.871)
	15	13125.2	13272.3	(1.121)	13810.0	(5.217)
	20	2786.2	2818.7	(1.168)	2913.6	(4.574)

Table 6. The mean and the variance of the first overflow time and their relative errors for the $E_5/M/1$ system ($N = 21$, $\frac{1}{\mu} = 1.0$).

ρ	i	$E[T(i,N)]$	$E[T_d(i,N)]$	relative error(%)	$E[\hat{T}_d(i,N)]$	relative error(%)
0.95	0	759.051	731.885	(-3.579)	761.293	(0.295)
	5	734.219	707.597	(-3.626)	733.832	(-0.053)
	10	642.784	618.404	(-3.793)	639.809	(-0.463)
	15	448.527	430.411	(-4.039)	444.464	(-0.906)
	20	96.220	92.024	(-4.361)	94.884	(-1.389)
1.05	0	217.732	220.665	(1.347)	225.547	(3.589)
	5	199.924	202.579	(1.328)	205.454	(2.776)
	10	154.949	156.766	(1.173)	158.314	(2.172)
	15	91.671	92.611	(1.025)	93.282	(1.757)
	20	16.179	16.324	(0.893)	16.414	(1.450)

ρ	i	$V[T(i,N)]$	$V[T_d(i,N)]$	relative error(%)	$V[\hat{T}_d(i,N)]$	relative error(%)
0.95	0	464572.9	436804.8	(-5.977)	473805.2	(1.987)
	5	464188.3	436389.6	(-5.989)	473269.7	(1.956)
	10	455003.3	427268.8	(-6.095)	462831.0	(1.720)
	15	391467.7	366298.5	(-6.429)	395688.9	(1.078)
	20	113583.5	105549.8	(-7.073)	113555.0	(-0.025)
1.05	0	237062.9	252597.4	(6.553)	263078.8	(10.974)
	5	235246.8	250536.5	(6.454)	260525.5	(10.699)
	10	215317.5	228345.3	(6.050)	236268.7	(9.730)
	15	152590.5	160948.0	(5.477)	165554.8	(8.496)
	20	31716.0	33263.4	(4.879)	34036.8	(7.317)

7. Concluding Remarks

By using diffusion approximation, we have investigated some basic properties of the first overflow time in a general queueing system. The d.f. and the first two moments of the first overflow time have been derived in the L.S.T. and explicit forms, respectively. For the case of Poisson arrival and/or small N, these results have been corrected by modifying the boundary condition at the origin. It is noted that under the heavy traffic condition, the solutions by diffusion approximation converge to those by fluid approximation. These results have been applied to the analysis of the maximum number of customers in the $GI/G/1(\infty)$ queueing system. The L.S.T. of its d.f. and the L.T. of its mean have been derived. It is shown from numerical results that they are fairly accurate in the heavy traffic.

It is of interest to extend the model to many server queueing system. It seems that the extension is not so difficult theoretically, since for many server system, it is only necessary to vary the diffusion parameters a and b appropriately [15]. It becomes, however, difficult to solve explicitly the corresponding differential equations because of spatial nonhomogeneity.

Acknowledgment

The authors would like to thank the referees for several valuable suggestions and clarifications.

Appendix

Derivation of (3.14)

Let $\overline{m}_1(x_0)$ denote a solution of the homogeneous differential equation

$$\text{(A.1)} \qquad \frac{1}{2}\, a\, \frac{d^2 m_1}{d\, x_0^2} + b\, \frac{dm_1}{dx_0} = 0.$$

When $b \neq 0$, by the usual quadrature we obtain

$$\text{(A.2)} \qquad \overline{m}_1(x_0) = A_1 + B_1 exp(-2bx_0/a),$$

where A_1 and B_1 are arbitary constants. Upon choosing $m_1(x_0) = -\, x_0/b$ as a particular solution of (3.11) for $n = 1$, a general solution of (3.11) is given by

$$\text{(A.3)} \qquad m_1(x_0) = A_1 + B_1 exp(-2bx_0/a) - \frac{1}{b}\, x_0.$$

Applying the boundary condition (3.12) to (A.3), we have

(A.4) $$B_1 = -a/2b^2.$$

Substituting (A.4) in (A.3) and the boundary condition (3.13) lead to

(A.5) $$A_1 = \frac{1}{b}\{\frac{a}{2b}\,exp(-2bN/a) + N\}.$$

From (A.3), (A.4) and (A.5), we obtain the desired result for $b \neq 0$.

When $b = 0$ in (A.1), it follows easily that

(A.6) $$m_1(x_0) = -\frac{1}{a}x_0^2 + C_1x_0 + D_1,$$

where C_1 and D_1 are arbitary constants. Applying the boundary conditions (3.12) and (3.13) to (A.6) yields $C_1 = 0$ and $D_1 = N^2/a$, and hence we obtain (3.14) for $b = 0$. □

Derivation of (3.15)

Let $\overline{m}_2(x_0)$ denote a homogeneous solution of (3.11) for $n = 2$. When $b \neq 0$, it is clear that $\overline{m}_2(x_0)$ has the same form as (A.2), that is,

(A.7) $$\overline{m}_2(x_0) = A_2 + B_2exp(-2bx_0/a),$$

where A_2 and B_2 are arbitary constants. Assume the form of a particular solution of (3.11) for $n = 2$ as

(A.8) $$m_2(x_0) = px_0^2 + qx_0 + rx_0exp(-2bx_0/a),$$

where p, q and r are arbitary constants. Substituting (A.8) and (3.14) in

(A.9) $$\frac{1}{2}a\frac{d^2m_2}{d\,x_0^2} + b\frac{dm_2}{dx_0} = -2m_1(x_0)$$

and comparing the coefficients of each term of x_0 in both sides, we have

(A.10) $$\begin{aligned} p &= 1/b^2 \\ q &= -a\{1 + exp(-2bN/a)\}/b^3 - 2N/b^2 \\ r &= -a/b^3. \end{aligned}$$

Hence, a general solution of (A.9) is given by

(A.11) $$m_2(x_0) = A_2 + B_2exp(-2bx_0/a) - \frac{1}{b^2}x_0\big((2N - x_0) + \frac{a}{b}\{1 + exp(-2bN/a) + exp(-2bx_0/a)\}\big).$$

Applying the boundary condition (3.12) to (A.11), we have

(A.12) $$B_2 = -\frac{a}{b^3}\big(N + \frac{a}{2b}\{2 + exp(-2bN/a)\}\big).$$

Substitution of (A.12) in (A.11) and the boundary condition (3.13) lead to

(A.13) $$A_2 = \frac{a}{b^3}\big(3N + \frac{a}{2b}\{2 + exp(-2bN/a)\}\big)exp(-2bN/a) + \frac{N^2}{b^2} + \frac{aN}{b^3}.$$

From (A.11), (A.12) and (A.13), we obtain the desired result for $b \neq 0$.

When $b = 0$ in (A.9), it follows easily that

(A.14) $$m_2(x_0) = \frac{1}{3a^2} x_0^4 - \frac{2}{a^2} N^2 x_0^2 + C_2 x_0 + D_2,$$

where C_2 and D_2 are arbitary constants. Applying the boundary conditions (3.12) and (3.13) to (A.14) yields $C_2 = 0$ and $D_2 = 5N^4/3a^2$, and hence we obtain (3.15) for $b = 0$. □

Derivation of (4.9)

When $b \neq 0$, for $n = 1$ the general solution of (3.11) has the following form given by (A.3):

(A.15) $$\tilde{m}_1(x_0) = A_3 + B_3 exp(-2bx_0/a) - \frac{1}{b} x_0,$$

where A_3 and B_3 are arbitary constants. Applying the boundary condition (4.8) to (A.15), we have

(A.16) $$B_3 = \frac{1}{1 - exp(-2b/a)} \left(\frac{1}{\lambda} - \frac{1}{b} \right).$$

Substitution of (A.16) into (A.15) and the boundary condition (3.13) lead to

(A.17) $$A_3 = \frac{N}{b} - \frac{1}{1 - exp(-2b/a)} \left(\frac{1}{\lambda} - \frac{1}{b} \right) exp(-2bN/a).$$

From (A.15), (A.16) and (A.17), we obtain the desired result for $b \neq 0$.

When $b = 0$, it follows from (A.6) that $\tilde{m}_1(x_0)$ has the form:

(A.18) $$\tilde{m}_1(x_0) = -\frac{1}{a} x_0^2 + C_3 x_0 + D_3,$$

where C_3 and D_3 are arbitary constants. Applying the boundary conditions (4.8) and (3.13) to (A.18) yields $C_3 = 1/a - 1/\lambda$ and $D_3 = N\{(N - 1)/a + 1/\lambda\}$. Hence we obtain (4.9) for $b = 0$ from the relations $b = \lambda - \mu$ and $\rho = \lambda/\mu$. □

References

[1] Bellman, R. E., Kalaba, R. E. and Lockett, J. A., *Numerical Inversion of the Laplace Transform*, American Elsevier, New York, 1966.

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

[3] ____________, "On the Busy Periods for the M/G/1 Queue with Finite and Infinite Waiting Room," *Journal of Applied Probability*, 8, 821-827 (1971).

[4] Cox, D. R. and Miller, H. D., *The Theory of Stochastic Processes*, John Wiley, New York, 1965.

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

[6] 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).

[7] Gaver, D. P., Jr., "Observing Stochastic Processes, and Approximate Transform Inversion," *Operations Research*, 14, 444-459 (1966).

[8] Heyman, D. P., "A Diffusion Model Approximation for the GI/G/1 Queue in Heavy Traffic," *The Bell System Technical Journal*, 54, 1637-1646 (1975).

[9] Hokstad, P., "Asymptotic Behaviour of the E_k/G/1 Queue with Finite Waiting Room," *Journal of Applied Probability*, 14, 358-366 (1977).

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

[11] Kleinrock, L., *Queueing Systems*, *Vol. II: Computer Applications*, Wiley-Interscience, New York, 1975.

[12] Newell, G. F., *Applications of Queueing Theory*, Chapman & Hall, Ltd., London, 1971.

[13] Ross, S. M., *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, 1969.

[14] Saaty, T. L., "Some Stochastic Processes with Absorbing Barriers," *Journal of the Royal Statistical Society*, *Series B*, 23, 319-334 (1961).

[15] Sunaga, T., Kondo, E. and Biswas, S. K., "An Approximation Method Using Continuous Models for Queueing Problems," *Journal of the Operations Research Society of Japan*, 21, 29-44 (1978).

[16] Suzuki, T. and Hayashi, K., "On Queue Disciplines," *ibid.*, 13, 43-58 (1970).

[17] Sweet, A. L. and Hardin, J. C., "Solutions for Some Diffusion Processes with Two Barriers," *Journal of Applied Probability*, 7, 423-431 (1970).

[18] Truslove, A. L., "Queue Length for the E_k/G/1 Queue with Finite Waiting Room," *Advances in Applied Probability*, 7, 215-226 (1975).

[19] ______________, "The Busy Period of the E_k/G/1 Queue with Finite Waiting Room," *ibid.*, 7, 416-430 (1975).

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