J. Operations Research Soc. of Japan
Vol. 19, No. 1, March, 1976

TRANSIENT BEHAVIOUR OF THE MEAN WAITING TIME AND ITS EXACT FORMS IN M/M/1 AND M/D/1

MASAO MORI
Ibaraki University

(Received October 4, 1974)

Abstract

The qualitative natures of the transient behaviours of the waiting times in a GI/G/1 system are studied first, when the waiting time of the initial customer is v. For instance, it is shown that the mean waiting time is increasing in n only when $v < l$ for some l.

Next this paper gives a way of calculating exact values of the mean waiting time of the n-th arrival in M/M/1 and M/D/1 and further proposes a simple scale of the rate of convergence to the equilibrium state.

1. Introduction and preliminaries.

Many authors have written papers on the subject of transient behaviours of queueing systems until now. Some of them have derived exact finite time solutions for simple queues (see Cohen [2]). Some of them have studied on the rate of convergence — the exponential convergence — to the equilibrium state (see Vere-Jones [17], Heathcote [10] and Miyazawa [12]). And some of them have checked the decreasing rate of the serial correlation coefficients of waiting times and have utilized the results to evaluate errors in estimating the mean waiting time by using sample averages on simulation experiments (see Daley [3] and Blomqvist [1]). Certainly their works give important and worthy informations on the subject to the practical researchers and the theoreticians.

From another standing point, Davis [5] introduced "*the build-up time*" as a rough scale of the rate of convergence to the equilibrium state. And, for a M/G/1 system, Morimura [13] conjectured that the queueing process may be considered to have almost reached at equilibrium state after the lapse of about two times of the build-up time T_b starting from the empty state, where $T_b = \int_0^\infty t \cdot d\mathrm{E}v(t) \,/\, \mathrm{E}v(\infty)$ and $v(t)$ is the virtual waiting time at t. This result also proposes a convienient and valuable measure to the practical researchers.

But it was reported by Hashida [9] that for the M/M/k system the build-up time T_i, starting from the state where there are i customers in the system initially, may take a negative value. For example T_i is negative in the case of the traffic intensity $\rho = 0.4$. The reason why T_i may take a negative value is because the mean system size does not necessarily increase monotonously and may be fluctuating. So we think that it is necessary to investigate the grobal properties of transient patterns of queueing processes.

In this paper we will first consider a GI/G/1 queueing system with first in first served (FIFS) discipline. Let t_n be the interarrival interval between the $(n-1)$-th and the n-th arriving customers; $\{t_n, n = 1, 2, \ldots\}$ is a sequence of positive i.i.d. variables. And the service time of the n-th arriving customer is denoted by s_n ; $\{s_n, n = 0, 1, 2, \ldots\}$ is also a sequence of positive i.i.d. variables. And we denote by $w_n(v)$ the waiting time of the n-th arriving customer when the system starts from the initial condition that there remain total residual service loads of v at $t = 0$, i.e. $w_0 = v$. Now we would want to know the qualitative nature of the transient behaviour of the process $\{w_n(v), n = 1, 2, \ldots\}$, and, in particular, we are interested in the mean waiting time $Ew_n(v)$ as functions of n for different initial values of v. For example, we can show $Ew_n(v)$ is increasing in n, if v is smaller than some value l.

In the section 3, we give the explicit forms of $P\{w_n(v) = 0\}$ in M/M/1 and M/D/1, and by using them we can calculate the exact values of $Ew_n(v)$ recursively to study the transient behaviour more concretely.

And for these case, we will examine numerically some scales of the rate of convergence to the equilibrium state, when the process starts from $w_0 = 0$. In those scales, $\tau_2 = (c_a^2 + \rho c_b^2) / (1 - \rho)^2$ seems to be a favourable one, where c_a and c_b are the coefficient of variation of the inter-arrival times and the service times respectively and ρ is the traffic intensity. Transforming the number n of the n-th arriving customer into $n' = n/\tau_2$, the values of $\gamma'_{n'} = [Ew_n(0) / Ew_\infty] \times 100$ % for M/D/1 behaves almost similarly to those values for M/M/1, where $\gamma'_{n'}$ is considered to represent a measure of measuring *"the closeness to the equilibrium state."* And it is shown that $\gamma'_{n'} \sim 95$ % for $n' = 2$ for both cases. Other time transformations such as stated above have been dealt with by Gaver [8] and Newell [14] in the study of diffusion approximations of queueing processes.

Here we are summing up preliminary definitions and a lemma for later use.

1. *(Esary, Proschan and Walkup* [6].*)*

Random variables $(X_1, X_2, \ldots, X_n)$ *are said to be associated if*

$$\text{(1.1)} \qquad \mathrm{cov}\{f(X_1, X_2, \ldots, X_n),\ g(X_1, X_2, \ldots, X_n)\} \geqq 0$$

for every pair of functions f *and* g *which are monotone non-decreasing in each arguments, provided that the covariance exists.*

It is noticed that mutually independent random variables are of course associated.

2. Let $\mathfrak{X}$ be the family of real valued random variables with finite mean on a probability space.

We define the following binary relation on $\mathfrak{X} \times \mathfrak{X}$, which are introduced by H. Stoyan and D. Stoyan [16] to investigate the order relationships between some queueing systems.

Definition. (Stoyan) *For any* X_1 *and* $X_2 \in \mathfrak{X}$, *we denote* $X_1 \subset X_2$ *if*

$$\int_c^\infty P\{X_1 > x\}dx \leqq \int_c^\infty P\{X_2 > x\}dx \tag{1.2}$$

for all c.

Lemma. *(a) If* $X_1 \subset X_2$ *and* $X_2 \subset X_3$, *then* $X_1 \subset X_3$ *(Trasitivity).*

(b) If X_3 *is independent of* X_1 *and* X_2, *then* $X_1 \subset X_2$ *implies* $X_1 + X_3 \subset X_2 + X_3$.

(c) If $X_1 \subset X_2$, *then* $\max(0, X_1) \subset \max(0, X_2)$.

(d) If $X_1 \subset X_2$, *then* $EX_1 \leqq EX_2$. *Particularly in the case of* $X_i \geqq 0$ *and* $EX_i^m < \infty$ *for* $i = 1, 2$, *we have* $EX_1^m \leqq EX_2^m$.

(e) Let $P\{X_1 = a\} = 1$. *Then we have* $X_1 \subset X_2$ *for any* X_2 *such that* $EX_2 = a$.

And many other interesting properties are shown in [16]. Here we add one new property which gives us useful information to compare queueing systems.

(f) Let $\{X_i\}$ *be a sequence of i.i.d. random variables. Then*

$$\ldots \subset Y_{i+1} \subset Y_i \subset \ldots \subset Y_2 \subset Y_1 \quad \text{where } Y_i = \sum_{j=1}^{i} X_j / i . \tag{1.3}$$

This property is proved by the similar way as deriving equation (3) in [1].

In this paper we often say that X_1 is smaller than X_2 *in the Stoyan's sense* if $X_1 \subset X_2$. Notice that this definition is a natural extention of the ordinary stochastic order, which we denote $X_1 \overset{st.}{\leqq} X_2$ if $P\{X_1 > x\} \leqq P\{X_2 > x\}$.

2. The transient patterns of the $\{w_n(v)\}$ process.

2.1 The $\{w_n(v)\}$ process.

Now let us consider a queueing system GI/G/1 with FIFS discipline in this section. Let w be a typical random variable which obeys to the equilibrium waiting time distribution $W(x)$ if it exists. Let $G(x)$ be the d.f. of $u_n = s_n - t_{n+1}$, and let $U_n = \sum_{i=0}^{n} u_i$. Hence we have

$$(2.1)\qquad \begin{aligned} &w_0(v) = v\,, \\ &w_{n+1}(v) = max(0,\ w_n(v) + u_n) \qquad for\ \ n \geqq 0, \end{aligned}$$

from which we can easily derive

$$(2.2)\qquad w_n(v) = max(0,\ u_{n-1},\ u_{n-1} + u_{n-2},\ \ldots,\ u_{n-1} + u_{n-2} + \ldots + u_0 + v).$$

For any fixed n, let us put $u_i^* = u_{n-1-i}$ and $U_i^* = \sum_{j=0}^{i} u_j^*$ for $i = 0, 1, 2, \ldots, n-1$. Then we can rewrite (2.2) as

$$\begin{aligned}(2.3)\qquad w_n(v) &= max(0,\ U_0^*,\ U_1^*,\ \ldots,\ U_{n-1}^* + v) \\ &= max(0,\ U_0^*,\ U_1^*,\ \ldots,\ U_{n-1}^*) \\ &\qquad + max(0,\ v + U_{n-1}^* - max(0,\ U_0^*, U_1^*,\ \ldots,\ U_{n-1}^*)) \\ &= w_n(0) + [v + mix(0,\ U_0,\ U_1,\ \ldots,\ U_{n-1})]^+ \end{aligned}$$

with probability one, where $[a]^+ = max(0,\ a)$. And it is noteworthy that the last term is stochastically equal to $[v - \hat{w}_n]^+$, i.e.

$$(2.4)\qquad M(v) \equiv [v + min(0,\ U_0,\ \ldots,\ U_{n-1})]^+ \sim [v - \hat{w}_n]^+,$$

where $\{\hat{w}_n\}$ is the waiting time process of the dual queueing system generated by interchanging the service time distribution and the interarrival time distribution of the original system each other, in which the process is assumed to start from the state $\hat{w}_0 = 0$. Of course $w_n(0)$ and $M(v)$ are highly correlated, but as for the expectation we can write

$$(2.5)\qquad \mathrm{E}w_n(v) = \mathrm{E}w_n(0) + \mathrm{E}[v - \hat{w}_n]^+ .$$

Hereafter we assume that the original system satisfies the equilibrium condition $-\infty < \mathrm{E}(u_0) < 0$. So the dual queueing system turns out to be a divergent queueing system.

2.2 On the transient patterns.

At first we will sum up the results obtained by Blomqvist [1] on the transient pattern of $\{w_n(0)\}$ process when $v = 0$. Let us denote $w_n(0)$ by w_n for simplicity.

Proposition 2.1 (Blomqvist)

(i) $\{w_n\}$ *is stochastically increasing in* n, *i.e.* $w_0 \overset{st.}{\leqq} w_1 \overset{st.}{\leqq} w_2 \overset{st.}{\leqq} \dots$. *Further there exists at least one* x *such that the strict inequality* $P\{w_n \geqq x\} < P\{w_{n+1} \geqq x\}$ *holds.*

(ii) $\{Ew_n\}$ *is strictly increasing and concave in* n.

(iii) $\{var(w_n)\}$ *is also strictly increasing in* n.

Next we will state on the transient patterns of the $\{w_n(v)\}$ process. First we get the following rather trivial results.

Proposition 2.2

(i) $w_n(v)$ *is stochastically increasing in* v *for all* n.

(ii) $Ew_n(v)$ *is strictly increasing in* v *for all* n.

(iii) $var(w_n(v))$ *is also increasing in* v *for all* n.

Proof. (i) and (ii) are trivial from (2.3), and for more general queueing systems such as G/G/k with Markovian inputs analogous results may be derived by using stochastic monotonicity of the Markov chain (see Daley [4]). Now we will prove (iii). From (2.3) we get

$$(2.6) \qquad var(w_n(v)) = var(w_n) + var(M(v)) + 2cov(w_n, M(v)).$$

From (2.4) we have $var(M(v)) = var([v - \hat{w}_n]^+)$ and

$$(2.7) \qquad \frac{\partial}{\partial v} var([v - \hat{w}_n]^+) = \frac{\partial}{\partial v} \{\int_0^v (v-x)^2 dF_n(x) - (\int_0^v (v-x)dF_n(x))^2\}$$

$$= 2(1 - F_n(v)) \cdot \int_0^v (v-x)dF_n(x) \geqq 0$$

where $F_n(x)$ is the d.f. of $\hat{w}_n$, hence $var(M(v))$ is increasing in v. Thus it is enough to show that

$$(2.8) \qquad cov(w_n, M(v+\Delta)) - cov(w_n, M(v)) = cov(w_n, M(v+\Delta) - M(v)) \geqq 0$$

for any $\Delta > 0$. Now from the definitions of w_n and $M(v)$, they can be regarded as functions of the associated random variables $(u_0, u_1, u_2, \dots, u_{n-1})$. Clearly, w_n is monotone non-decreasing in each u_i. And $M(v+\Delta) - M(v)$ is also monotone non-decreasing in each u_i, since $m = min(0, U_0, \dots, U_{n-1})$ is monotone non-decreasing in each u_i and $M(v+\Delta) - M(v) = (m+v+\Delta)^+ - (m+v)^+$ is so in m. Thus, (2.8) is assured from the definition of association.

Proposition 2.3

(i) *The difference* $(w_n(v) - w_n)$ *is stochastically decreasing in* n *for all* v.

(ii) *And the difference* $(\mathrm{E}w_n(v) - \mathrm{E}\hat{w}_n)$ *is strictly decreasing in* n *for all* v.

Proof. (i) is immediately deduced from (2.3). The "strict" decreasingness in (ii) is derived from the later assertion in Proposition 2.1 - (i), which is also true for $\hat{w}_n$.

The above propositions are shown by using the ordinary stochastic order, but we need the Stoyan's order in the following further discussion.

Proposition 2.4

If $v \leqq \mathrm{E}w$, *then there exists at least a stochastically increasing sequence of random variables* $\{w'_n\}$ *such that (A)* $\mathrm{E}w'_0 = v$ *and* $w_n(v) \quad w'_n$ *for all* n *and that (B) it has the same limiting distribution* $W(x)$ *as* $\{w_n(v)\}$. *And* $\mathrm{E}w'_n$ *is strictly increasing in* n *in this case.*

If $v > \mathrm{E}w$, *then there exists at least a stochastically decreasing* $\{w'_n\}$ *such that it has the same properties (A) and (B) as above, but* $\mathrm{E}w'_n$ *is strictly decreasing in* n *in this case.*

Proof. Let us define $\{w'_n\}$ recursively by $w'_{n+1} = (w'_n + u_n)^+$, where w'_0 is any non-negative random variables satisfying $\mathrm{E}w'_0 = v$ and $\{u_n\}$ is the same sequence as in generating $\{w_n(v)\}$. Then we can recursively show that $w'_{n+1} = (w'_n + u_n)^+ \quad (w_n(v) + u_n)^+ = w_{n+1}(v)$ holds for each n from the properties *(a)*, *(b)*, *(c)* and *(e)*. Thus *(A)* is ascertained for any w'_0 such that $\mathrm{E}w'_0 = v$. And *(B)* is also true from the definitions of $\{w'_n\}$.

Now the remaining subject is how to determine w'_0 so as to make $\{w'\}$ stochastically increasing or not subject to $\mathrm{E}w'_0 = v$. Let w'_0 be the random variable such that $w'_0 \sim (w - c)^+$ if $v \leqq \mathrm{E}w$, or such that $w' \sim (w + c')$ if $v > \mathrm{E}w$, where w is a typical random variables with the d.f. $W(x)$ and the constants c and c' are the values such that $\mathrm{E}w'_0 = v$ may hold for each case. Of course, it is possible to choose such c and c' satisfying $\mathrm{E}w'_0 = v$, for $\mathrm{E}(w - c)^+$ and $\mathrm{E}(w + c')$ are continuous in c and c' respectively. Then, it is easily seen after a little calculation that w'_n decreases stochastically in n if $v > \mathrm{E}w$ and increases if $v < \mathrm{E}w$. And choosing w'_0 as above, the strict increasingness or decreasingness of $\mathrm{E}w'_n$ are easily shown.

However, we do not know yet what conditions should be putted upon w'_0 in order to make a sequence $\{w'_n\}$ stochastically increasing in n or, further, increasing in the Stoyan's sense. Let $W_n(x)$ be the d.f. w'_n, and we have

$$(2.9) \qquad W_{n+1}(x) = \int_{-\infty}^{x} W_n(x - y)dG(y),$$

which is the Wiener-Hopf equation first derived by D.V. Lindley. First, by using (2.9) iteratively it is soon noticed that, $\{w_n'\}$ is stochastically increasing in n, i.e. $\{W_n(x)\}$ is decreasing in n for each x, if and only if $W_0(x) \geqq W_1(x)$ for any x. Analogously from (b) and (c) $\{w_n'\}$ is increasing in the Stoyan's sense if and only if $w_0' \subset w_1'$. Thus the problem is reduced to finding the condition $W_0(x) \geqq W_1(x)$ or $w_0' \subset w_1'$ holds. Of course, similar results are also deduced concerning the decreasingness of $\{w_n'\}$.

Here some sufficient conditions are easily found: if $W_0(x) \geqq G(x)$ for all x, we have $W_0(x) \geqq W_1(x)$ for all x. And if $W_0(x)$ is represented as $W_0(x) = \alpha G(x) + (1-\alpha)W(x)$ for some α $(0 \leqq \alpha \leqq 1)$, we have also $W_0(x) \geqq W_1(x)$ In these cases, $w_0 \subset w_1$ also holds. However, the condition $W_0(x) \geqq W(x)$ for all x does not necessarily imply that $W_0(x) \geqq W_1(x)$ for each x. For example, for any fixed $c > 0$, let $W_0(x) = W(x)$ for $x \geqq c$ and $W_0(x) = W(c)$ for $0 \leqq x < c$. Then, after a little calculation, it is seen that $W_0(x) \leqq W_1(x)$ for each x such that $x \geqq c$, and in this case $w_0 \subset w_1$ does not hold. But, if we define $W_0(x) = W(x)$ $(x < c)$; $= 1$ $(x \geqq c)$, $w_0 \subset w_1$ holds but $W_0(x) \geqq W_1(x)$ does not for any c. And from Proposition 2.6 stated later, if $W_0(x) = 0$ $(x < v)$; $= 1$ $(x \geqq v)$ and v is smaller than some value l, $w_0 \subset w_1$ holds. So it may be expected that the transient behaviour of the waiting time in the global sense greatly depends upon whether $W_0(x) \geqq W(x)$ for large value of x or not.

2.3 Transient patterns of $\{Ew_n(v)\}$.

Hereafter, we will concentrate only upon the transient patterns of the expectation value of the $\{w_n(v)\}$ process and investigate them in rather detail.

Let l be a solution of the equation $\int_{-x}^{\infty}(1 - G(y))dy = x$. It is known that l is the unique solution of this equation and $l \leqq Ew$ ([11]). And from the following propositions it will be seen that l is also a critical point for initial values of the waiting time process.

Proposition 2.5

$$(2.10) \qquad Ew_1(v) \underset{(\leqq)}{\geqq} v \qquad \text{if} \quad v \underset{(\geqq)}{\leqq} l .$$

Proof. By using (2.9), we have

$$(2.11) \qquad Ew_1(v) = \int_{-v}^{\infty}(1 - G(x))dx = \left(\int_{-l}^{\infty} + \int_{-v}^{-l}\right)(1 - G(x))dx$$

$$= l + \int_{-v}^{-l} (1 - G(x))dx.$$

Since $(1 - G(x)) \leqq 1$ in the above, we obtain (2.10).

Proposition 2.6

If $v \leqq l$, then $EW_n(v)$ is increasing in n and further $\{w_n(v)\}$ is increasing in the Stoyan's sense.

Proof. From Proposition 2.5 we have $Ew_1(v) \geqq v$. Regard $Ew_1(v)$ as a random variable taking a constant value, and $v \sqsubset Ew_1(v) \subset w_1(v)$ holds from the property *(e)*. Hence from *(b)* and *(c)* the relation $w_1(v) = (v + u_0)^+ \sim (v + u_1)^+ \sqsubset (w_1(v) + u_1)^+ = w_2(v)$ is obtaind. Through the same arguments, we have $w_n(v) \sqsubset w_{n+1}(v)$ for all n, which completes the proof.

Now the following open questions occur to us: whether does $Ew_n(v)$ oscillate in many times or not for $v > l$? ; whether does $Ew_n(v)$ approach to Ew from above or not for large v ? We cannot answer here the above matters at all.

However, from examples stated in the last of the section 2.2 and so on, it is conjectured that $Ew_n(v)$ may once decrease beyond Ew and gradually increase to Ew from below for any large v. Further, it is conjectured that Ew_n' may approach to Ew from below for large n, if there exists $x_0 < \infty$ such that $P\{w_0' \leqq x\} > P\{w \leqq x\}$ for any $x > x_0$.

In the next section, we shall show a way of calculating exact values of $Ew_n(v)$ for M/M/1 and M/D/1 and study the transient behaviours of them more concretely.

3. Transient solutions for M/M/1 and M/D/1.

3.1 The fundamental relations for M/G/1.

In this section we are going to obtain the exact forms of the mean waiting time of the n-th arriving customer for the systems M/M/1 and M/D/1 and to examine their transient behaviour in detail. For later use we prepare at first a few fundamental relations for M/G/1 without proof. These relations are analogously derived as in [2] (pp. 249-255).

Let λ be the arrival rate of customers and $B(x)$ be the d.f. of the service times. As the recurrence relation for the mean waiting times of the successive customers we have

$$(3.1) \qquad \mathrm{E}w_n(v) = \mathrm{E}w_{n-1}(v) - \frac{1}{\lambda}(1-\rho) + \frac{1}{\lambda}\mathrm{P}\{w_n(v) = 0\}$$

and $\mathrm{E}w_0(v) = v$. So we can calculate $\mathrm{E}w_n(v)$, if the exact values of $\mathrm{P}\{w_n(v) = 0\}$ $(n = 1, 2, \ldots)$ are in hand. Now we consider the way of obtaining the value of $a_n(v) \equiv \mathrm{P}\{w_n(v) = 0\}$.

Let $r(v)$ be the number of customers served in the busy period initiated by the 0-th customer whose residual waiting time at the time 0 is v, i.e. $w_0 = v$. And let $R_v(z)$ be the generating function of $r(v)$ and denote $R_0(z)$ by $R(z)$ simply. If $\rho < 1$, then we have

$$(3.2) \qquad R_v(z) = z \cdot e^{-\lambda v(1-R(z))} \cdot \beta[(1 - R(z))]$$

for $|z| \leqq 1$, where $\beta(\theta)$ is the Laplace transform of $B(x)$. And $R(z)$ satisfies the functional equation

$$(3.3) \qquad R(z) = z \cdot \beta[\lambda(1 - R(z))]$$

and we have

$$(3.4) \qquad \mathrm{P}\{r(0) = k\} = \int_0^\infty \frac{(\lambda x)^{k-1}}{k!} e^{-\lambda x} \, dB^{k*}(x)$$

where $B^{k*}(x)$ is the k-fold convolution of $B(x)$.

Since $a_n(v) = \mathrm{P}\{w_n(v) = 0\}$ is considered as the renewal density at the time n of the delayed renewal process generated by random variables $r(v)$ and $r(0)$, thus we have

$$(3.5) \qquad Q(z) \equiv \sum_{n=1}^{\infty} a_n(v) z^n = \frac{R_v(z)}{1 - R(z)} = \sum_{k=1}^{\infty} e^{-\lambda v(1-R(z))} \cdot \{R(z)\}^k$$

for $|z| < 1$. The value of $a_n(v)$ is just the coefficient of z^n in the expansion of this generating function.

3.2 The value of $\mathrm{P}\{w_n(v) = 0\}$ for M/M/1.

Let μ be the service rate. In this case the system size process observed at each arrival and departure time points is regarded as a one dimensional random walk with an reflecting barrier at the origin which are generated by transition probabilities $p_{i,i+1} = p = \frac{\lambda}{\lambda + \mu} = \frac{\rho}{1 + \rho}$, $p_{i,i-1} = 1 - p = \frac{1}{1 + \rho}$ $(i = 1, 2, 3, \ldots)$, $p_{01} = 1$ and $p_{00} = 0$.

Thus we have

$$(3.6) \qquad \mathrm{P}\{r(0) = n\} = \frac{1}{n} \binom{2n-2}{n-1} \frac{\rho^{n-1}}{(1+\rho)^{2n-1}}$$

which is the probability of the first return to the origin at the time $2n$ through the positive axis in the random walk. And let R_{kn} be the coefficient of z^n in the expansion of $\{R(z)\}^k$, then we have

$$(3.7) \qquad R_{kn} = \frac{k}{2n-k}\binom{2n-k}{n}\frac{\rho^{n-k}}{(1+\rho)^{2n-k}}$$

which is the probability of the k-th return to the origin at $2n$ (see Feller [7] p. 77.) Of course, above probabilities can be also obtained by expanding $\{R(z)\}$.

Rewriting (3.5) as

$$(3.8) \qquad Q(z) = e^{-\lambda v}\sum_{k=1}^{\infty} e^{\lambda v R(z)}\{R(z)\}^k = e^{-\lambda v}\sum_{k=1}^{\infty}\{R(z)\}^k \cdot \sum_{j=0}^{\infty}\frac{(\lambda v)^j}{j!}\{R(z)\}^j$$

$$= \sum_{k=1}^{\infty}\{R(z)\}^k \cdot \sum_{j=0}^{k-1}\frac{(\lambda v)^j}{j!}e^{-\lambda v},$$

then we have

$$(3.9) \qquad a_n(v) = \sum_{k=1}^{n} R_{kn}\sum_{j=0}^{k-1}\frac{(\ v)^j}{j!}\ e^{-\lambda v}$$

Proposition 3.1

For the system M/M/1, if $\rho < 1$ *we have*

$$(3.10) \qquad \mathrm{P}\{w_n(v) = 0\} = \sum_{k=1}^{n}\frac{k}{2n-k}\binom{2n-k}{n}\frac{\rho^k}{(1+\rho)^{2n-k}}\sum_{j=0}^{k-1}\frac{(\lambda v)^j}{j!}e^{-\lambda v}\ .$$

Especially in the case of $v = 0$, after a little calculation we can rewrite (3.10) as

$$(3.11) \qquad \mathrm{P}\{w_n = 0\} = \frac{1}{(1+\rho)^n}\left\{\sum_{i=0}^{n-1}\left(\frac{\rho}{1+\rho}\right)^i\binom{-n}{i} - \frac{\rho}{1+\rho}\sum_{i=0}^{n-2}\left(\frac{\rho}{1+\rho}\right)^i\binom{-n-1}{i}\right\},$$

and we can see $\mathrm{P}\{w_n = 0\} \longrightarrow (1-\rho)$ as $n \longrightarrow \infty$.

The exact value of $\mathrm{E}w_n(v)$ can be calculated recursively by using (3.10) and (3.1).

3.3 The value of $\mathrm{P}\{w_n(v) = 0\}$ for M/D/1.

Let b be the service time. Since $\beta(\theta) = e^{-\theta b}$, from (3.2) $\sim$ (3.5) we have

$$(3.12) \qquad R_v(z) = ze^{-(\lambda v+\rho)(1-R(z))},$$

$$(3.13) \qquad R(z) = ze^{-\rho(1-R(z))} = \sum_{k=1}^{\infty} z^k e^{-k\rho}\frac{(k\rho)^{k-1}}{k!}$$

and

$$(3.14)\qquad Q(z) = \sum_{k=1}^{\infty} R_v(z)\{R(z)\}^{k-1} = \sum_{k=1}^{\infty} z^k e^{-(\lambda v+k\rho)(1-R(z))}$$

$$= \sum_{k=1}^{\infty} z^k e^{-(\lambda v+k\rho)} \cdot \sum_{j=0}^{\infty} \frac{(\lambda v+k\rho)^j}{j!} \{R(z)\}^j$$

By expanding (3.14), we get

$$a_n(v) = e^{-(\lambda v+n\rho)} \{1 + \sum_{k=1}^{n-1} e^{(n-k)\rho} \sum_{j=1}^{n-k} \frac{(\lambda v+k\rho)^j}{j!} R_{j,n-k} \},$$

and, replacing $n-k$ by k, we have

$$(3.15)\qquad a_n(v) = e^{-(\lambda v+n\rho)} \{1 + \sum_{k=1}^{n-1} e^{k\rho} \sum_{j=1}^{k} \frac{[\lambda v+(n-k)\rho]^j}{j!} R_{jk} \},$$

where R_{jk} $(1 \leqq j \leqq k)$ is given by

$$(3.16)\qquad R_{jk} = \sum_{\substack{\nu_1+\dots+\nu_j=k \\ \nu_i \geqq 1}} \prod_{i=1}^{j} \frac{(\nu_i\rho)^{\nu_i-1}}{\nu_i!} e^{-\nu_i\rho}$$

from (3.13). Putting $k_i = \nu_i - 1$, we can rewrite (3.16) as

$$(3.17)\qquad R_{jk} = e^{-k\rho} \cdot \rho^{k-j} \cdot \sum_{\substack{k_1+\dots+k_j=k-j \\ k_i \geqq 0}} \prod_{j=1}^{j} \frac{(1+k_i)^{k_i-1}}{k_i!} .$$

In order to simplify this, we use the following multinomial Abel identity given in Riordan [5] (p. 26):

$$(3.18)\qquad \sum_{k_1+\dots+k_j=n} \binom{n}{k_1, k_2, \dots, k_j} \prod_{i=1}^{j} (x_i + k_i)^{k_i-1}$$

$$= (x_1 \cdot x_2 \cdot \dots \cdot x_j)^{-1} \cdot x \cdot (x+n)^{n-1}$$

where $x = x_1 + x_2 + \dots + x_j$.

Now applying this identity to the last term in the right side of (3.17), we get

$$(3.19)\qquad R_{jk} = e^{-k\rho} \cdot \rho^{k-j} \cdot \frac{1}{(k-j)!} \cdot j \cdot k^{k-j-1} ,$$

and putting this into (3.15)

$$a_n(v) = e^{-(\lambda v+n\rho)} \{1 + \sum_{k=1}^{n-1} \{ \frac{(\lambda v+n\rho)}{k} - \rho\} \frac{(\lambda v+n\rho)^{k-1}}{(k-1)!} \},$$

we have

Proposition 3.2

For the system M/D/1, if $\rho < 1$ *we have*

(3.20) $$P\{w_n(v) = 0\} = E_{n-1}(\lambda v + n\rho) - \rho E_{n-2}(\lambda v + n\rho)$$

where $E_k(\lambda)$ *is the partial sum of Poisson probabilities with parameter* λ*,*

i.e. $E_k(\lambda) = \sum_{j=0}^{k} \frac{\lambda^j}{j!} e^{-\lambda}$.

Of course $P\{w_n(v) = 0\} \longrightarrow (1 - \rho)$ as $n \longrightarrow \infty$ for each v. And the exact values of $Ew_n(v)$ are also calculated recursively by (3.1).

3.4 Numerical examples for $\{Ew_n(v)\}$.

According to the discussions until now, we can draw graphs of $\{Ew_n(v)\}$ for different values of v. Figure 1 is the graph in the case of M/D/1 with $\rho = 0.6$, which shows typical transient patterns of $\{Ew_n(v)\}$ for every ρ. And the graphs in the case of M/M/1 are also similar to this.

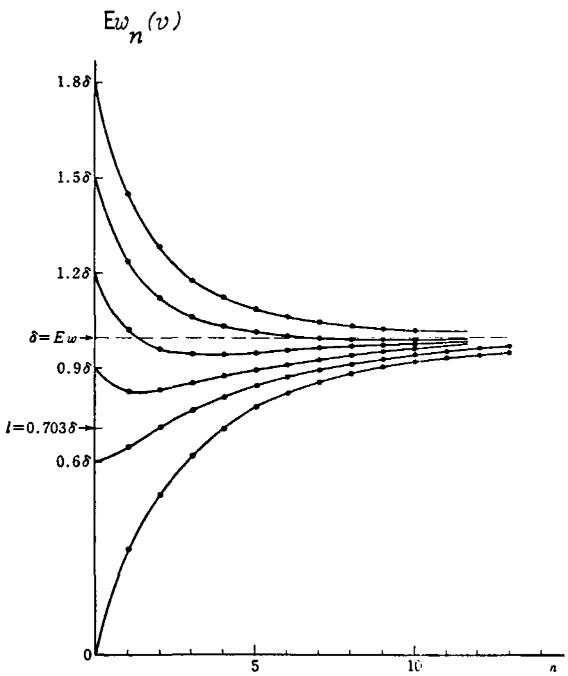

Fig. 1. The graph of $Ew_n(v)$ for M/D/1 ($\rho = 0.6$).

In this case, $Ew_n(1.5\delta)$ attains its minimum 0.9943δ at $n=11$ and $Ew_n(1.8\delta)$ attains its minimum 0.999993δ at $n=45$. Here $\delta \equiv Ew = 0.750$.

It is noteworthy that the value of $P\{w_n(v) = 0\}$ converges more rapidly to its limiting value $1-\rho$ than $Ew_n(v)$ does, which is naturally expected from the recursive relation (3.1) for the mean. This fact suggest to us that when we apply Markov model to a queueing system and estimate some quantities by multiplying the transition matrix with itself by a number of times recursively, percentiles of the distribution may be estimable more exactly than the moments values. On the other hand, when we must judge whether a practical queueing processes are considered to have reached at the equilibrium state or not, it may be more safe to judge from informations concerning the convergence rate of the mean waiting time than to use those of probabilities.

Now let us consider the unsolved question mentioned in the last of the section 2: for any value of $v \geqq Ew$, may $Ew_n(v)$ once decrease beyond Ew and increase gradually to Ew afterwards? In the case of M/G/1, from (3.1) this statement is rewritten as follows : for any value of v, may $P\{w_n(v) = 0\}$ ever excess over its limiting value $1-\rho$ and never cross down the value $1-\rho$ afterwards?

However, we cannot make definite theoretically whether this statement is true or not even in the case of M/D/1 and M/M/1. So we are going to examine this numerically. Certainly, $Ew_n(v)$ behaves as in the manner stated in the above statement, if v is smaller than $1.6 \sim 1.8$ times of Ew for $\rho = 0.1 \sim 0.9$ (1) in the both systems. But, $Ew_n(v)$ seems to be monotone decreasing to Ew, if v is larger than these values. For example, in the system M/D/1 with $\rho = 0.4$, for $v = 1.5$ Ew, $Ew_n(v)$ goes accross Ew from above at $n = 4$ and attains its minimum $0.9973 \times Ew$ at $n = 6$, and afterwards it increases to Ew; for $v = 1.8 \times Ew$, $Ew_n(v)$ attains its minimum $0.9999991 \times Ew$ at $n = 21$ and then increases; but, for $v = 2.0 \times Ew$, it is monotone decreasing to Ew and stays at Ew after $n = 30$. To our regret we cannot judge whether this behaviour is owing to computational errors or not.

Anyway, in a practical sense we are allowed to consider that $Ew_n(v)$ may be monotone decreasing in n, if v is larger than above $2 \times Ew$.

3.5 On some time transformations.

In the studies of diffusion approximations of queueing processes, a transform of the time scale and the queue length leads to a simple dimensionless

(1) We have computed the values $Ew_n(v)$ only for $\rho = 0.1 \sim 0.9$ $(+0.1)$ and for $v = 0 \sim 2.0 \times Ew$ $(+0.1 \times Ew)$.

deffusion equation (see Newell [4] and Gaver [8].) Thus the transformed process can be considered to be independent of values of parameters of the system. It is convienient to find such a nice transformation for general queueing processes, which are not neccessarily approximated by diffusion processes, in order to know their transient behaviours. Here we are going to examine such transformations only numerically (never theoretically) for M/M/1 and M/D/1. But the processes $\{w_n\}$, starting from $v = 0$, are only considered from now on.

Let us define as a transformed value of the mean waiting time

$$\gamma_n = \frac{Ew_n}{Ew} \times 100 \quad (\%)$$

after Newell, which may be reasonable, since Ew_n is monotone increasing to Ew and concave in n. Then, γ_n may give a nice information about the closeness to the equilibrium state.

As time transformations, we think of the following scales:

$$\text{(i)} \qquad \tau_1 = \frac{var(u)}{[E(-u)]^2} = \frac{c_a^2 + \rho^2 c_b^2}{(1-\rho)^2},$$

$$\text{(ii)} \qquad \tau_2 = \frac{c_a^2 + \rho c_b^2}{(1-\rho)^2},$$

$$\text{(iii)} \qquad \tau_b = 2\lambda \cdot T_b = 2\lambda\left\{\frac{\lambda b_2}{2(1-\rho)^2} + \frac{b_3}{3(1-\rho)b_2}\right\}$$

$$\text{(iv)} \qquad \tau_g = \lambda \cdot T_g = \lambda \cdot \frac{\lambda b_2}{(1-\rho)^2} = \frac{\rho^2(1+c_b^2)}{(1-\rho)^2}$$

where T_b and T_g are the build up time and the scale given by Gaver [8] respectively and b_j is the j-th moment of the service time. Since T_b and T_g are derived concerning the virtual waiting time process of M/G/1, so taking the dimension into consideration, we adopt λ times of their values, and further let λT_b be doubled for later convenience.

And τ_1 is an adaptation from Newell [14]. We are now considering the waiting time process, whereas the queue length process are dealt with in Newell [14]. τ_2 is a modification of τ_1, which is adopted from the following reason: an essential form of such scales seems to be $1/(1-\rho)^2$ times of some fluctuation quantity caused by the arrival and the service time processes and the contributions of the service time process are made only during busy period, while the proportion that the system is busy is ρ. The reasoning for adopting τ_2 is not so sufficient yet.

Here let us list up each τ for the cases of M/M/1 and M/D/1:

M/M/1	M/D/1
$\tau_1 = \frac{1+\rho^2}{(1-\rho)^2}$	$\tau_1 = \frac{1}{(1-\rho)^2}$
$\tau_2 = \frac{1+\rho}{(1-\rho)^2}$	$\tau_2 = \tau_1$
$\tau_b = \frac{2\rho}{(1-\rho)^2}$	$\tau_b = \frac{(2+\rho)^2}{3(1-\rho)^2}$
$\tau_g = \frac{2\rho^2}{(1-\rho)^2}$	$\tau_g = \frac{\rho^2}{(1-\rho)^2}$

Now we transform the number n of the n-th arriving customer into $n' = n/\tau$ for each τ and compare the values of $\gamma'_{n'}$ $(\equiv \gamma_n)$ numerically for each τ. It is notable that the compound ratio $\tau_1 : \tau_2 : \tau_b : \tau_g$ is nearly equal to $1 : 1 : 1 : 1$ if ρ is near to 1, but the deviation from this becomes much larger for small ρ. Thus, for each time transformation, $\gamma'_{n'}$ behaves almost similarly in one another when ρ is near to 1, i.e. when the diffusion approximations are applicable.

Table 1 and 2 give the values of $\gamma'_{n'}$ which are transformed by τ_1 and τ_2 respectively for the case of M/M/1. And Table 3 shows the values of $\gamma'_{n'}$ transformed by τ_1 $(= \tau_2)$ for the case of M/D/1. The values $\gamma'_{n'}$ in the tables are interpolated linearly by using the exact values of γ_n for $n = [n'\tau]$ and $n = [n'\tau] + 1$. Thus the values in each table may not be so reliable for small n', for γ_n is concave in n. The blanks in the tables are corresponding to these unreliable values. In order to avoid this fault, in Figure 2 we plot the exact values of γ_n at $n' = n/\tau_2$ for each n. Nevertheless, the corresponding values in each table seem to be near to one another for every ρ and n', especially the values in Table 2 and 3 seem to be very near.

Further, in each table, the values of $\gamma'_{n'}$ for every n', excepting $n' \leqq 1/4$, considered to be almost constant even varying the value of ρ $(\rho = 0.1 \sim 0.9)$. Hence the time scales τ_1 and τ_2 can be considered to give fairly nice transformations, and especially τ_2 may be slightly more favourable than τ_1 for $\rho = 0.1 \sim 0.9$ and $n' \geqq 1/2$. And, when transformed by τ_2, γ'_2 is about 95% for any $\rho = 0.1 \sim 0.9$. Hence, in a practical sense, we can consider that at $n \doteqdot 2\tau_2$ the process has almost reached at the equilibrium state.

On the contrary, since the values of τ_b and τ_g is too small when ρ is smaller than 0.7 or so, comparing with the values of τ_1 and τ_2, the values of $\gamma'_{n'}$ transformed by τ_b and τ_g does not behave so nicely as the values in Table 1, 2, and 3.

From the above discussion, the time transformation $n' = n/\tau_2$ is favourable one both in M/M/1 and in M/D/1, but, of course, we cannot apply this result directly to a general queueing system GI/G/1, even to M/G/1. Thereafter we need exemplify more theoretically on this subject.

Acknowledgement

The author wishes to thank Prof. Hidenori Morimura for his guidance and encouragement.

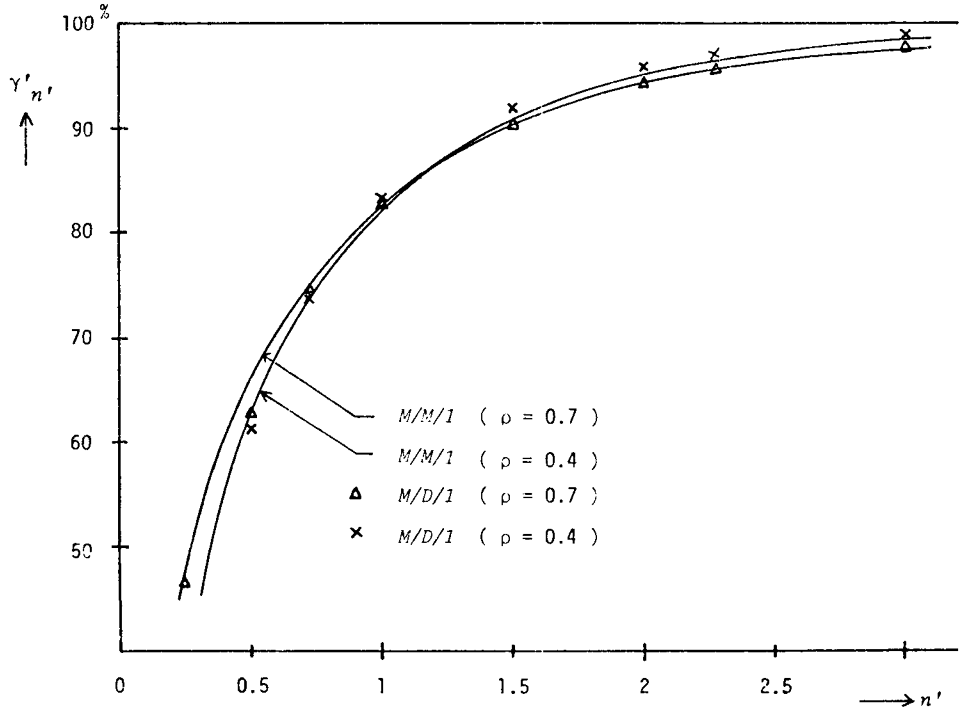

Table 1, 2 and 3, Fig 2. The graph of $\gamma'_{n'}$.

The exact values of $\gamma'_{n'}$ for M/M/1 and M/D/1 are plotted for $\rho = 0.4$ and 0.7.

Table 1. The values of $\gamma'_{n'}$ for τ_1 in the case of M/M/1.

ρ	τ_1	n' 1/4	1/2	1	1.5	2	3
0.1	1.25				94.2	97.5	99.6 %
0.2	1.63			79.4	90.3	95.2	98.7
0.3	2.22			78.2	88.4	93.6	97.8
0.4	3.22		55.6	77.6	87.5	92.7	97.4
0.5	5.00		57.6	78.0	87.3	92.4	96.9
0.6	8.50	41.2	60.4	78.9	87.5	92.4	96.9
0.7	16.6	45.4	63.1	80.1	87.3	92.7	96.9
0.8	41.0	49.4	66.1	81.8	89.1	93.2	97.1
0.9	181	53.7	69.1	83.4	*	*	*

Table 2. The values of $\gamma'_{n'}$ for τ_2 in the case of M/M/1.

ρ	τ_2	n' 1/4	1/2	1	1.5	2	3
0.1	1.36				96.0	98.1	99.8 %
0.2	1.88			84.5	93.1	96.8	99.3
0.3	2.65			82.8	92.1	95.8	98.9
0.4	3.89		63.7	82.6	91.1	95.2	98.5
0.5	6.00		63.7	82.5	90.7	94.8	98.2
0.6	10.0	45.6	65.0	82.7	90.5	94.5	98.0
0.7	18.9	48.6	66.0	83.1	90.4	94.3	97.8
0.8	45.0	51.4	68.3	83.6	90.6	94.3	97.7
0.9	190	54.8	70.2	84.2	*	*	*

Table 3. The values of $\gamma'_{n'}$ for τ_1 $(= \tau_2)$ in the case of M/M/1.

ρ	τ_2	n' 1/4	1/2	1	1.5	2	3
0.1	1.23				96.2	98.5	99.3 %
0.2	1.56			83.7	94.1	97.7	99.4
0.3	2.04			84.4	92.8	96.5	99.1
0.4	2.78		61.2	82.9	91.7	95.6	98.8
0.5	4.00	42.6	63.3	82.8	91.2	95.2	98.4
0.6	6.25	42.5	64.1	82.6	90.7	94.8	98.2
0.7	11.1	46.6	65.3	82.8	90.5	94.4	97.9
0.8	25.0	49.5	67.4	83.3	90.5	94.3	97.7
0.9	100	53.9	69.7	84.1	90.7	94.2	*

(The values for * are not computed.)

References.

[1] Blomqvist, N., "On the transient behaviour of the GI/G/1 waiting times," Skand. Aktur Tidiskr., (1970), 118-129.

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

[3] Daley, D.J., "The serial correlation coefficients of waiting times in a stationary single server queue," J. Austral. Math. Soc., 8 (1968), 683-699.

[4] Daley, D.J., "Stochastically monotone Markov Chain," Z. Wahrscheinlichkeits-theorie verw. Geb., 10 (1968), 305-317.

[5] Davis, H., "The build-up time of waiting lines," Naval Res. Logist. Quart., 2 (1960), 185-193.

[6] Esary, J.D.F., Proschan, F. and Walkup, D.W., "Association of random variables with applications," Ann. Math. Statist., 38 (1967), 1466-1474.

[7] Feller, W., *An introduction to probability theory and its applications*, vol. 1, 2nd ed., John Wiley, 1957.

[8] Gaver, D.P., "Diffusion approximations and models for certain congestion problems," J. Applied Prob., 5 (1968), 607-623.

[9] Hashida, A., "An evaluation for the rate of convergence to equilibrium state for M/M/s," (in Japanese), reported at Japan OR meeting, Autumn (1971).

[10] Heathcote, C.R., "Complete exponential convergence and some related topics," J. Applied Prob., 4 (1967), 217-256.

[11] Marchall, K.T., "Some inequalities in queueing," Opns. Res., 16 (1968), 651-658.

[12] Miyazawa, M., "On the rate of convergence of waiting-time distribution," Research Reports on Information Sciences, No. B-1 (1974), (Dept. of Information Sciences, Tokyo Institute of Technology.)

[13] Morimura, H., "The build-up time of equilibrium waiting time," J. Opns. Res. Soc. Japan, 4 (1962), 76-86.

[14] Newell, G.F., *Applications of Queueing Theory*, Chapman and Hall, 1971.

[15] Riordan, J., *Combinatorial Identities*, John Wiley, 1968.

[16] Stoyan, H. and Stoyan, D., "Monotonieeigenschaften der Kundenwartezeiten in Modell GI/G/1," Zeit. Angewandte Mech., 49 (1969), Heft 12, 729-734.

[17] Vere-Jones, D., "A rate of convergence problem in the theory of queues," Theory Prob. Applications, 9 (1964), 96-103.

Masao MORI

Faculty of Engineering
Ibaraki University
Nakanarusawa-machi 4-12-1
Hitachi, Ibaraki