Opl Res. Q., Pergamon Press 1977. Vol. 28, 2, i, pp. 325 to 337. Printed in Great Britain

On the Relative Waiting Times in the GI/M/s and the GI/M/1 Queueing Systems

URI YECHIALI

Department of Statistics, Tel-Aviv University, Israel

The expected steady-state waiting time, $W_q(s)$, in a GI/M/s system with interarrival-time distribution $H(\cdot)$ is compared with the mean waiting time, W_q, in an "equivalent" system comprised of s separate GI/M/1 queues each fed by an interarrival-time distribution $G(\cdot)$ with mean arrival rate equal to $1/s$ times that of $H(\cdot)$. For $H(\cdot)$ assumed to be Exponential, Gamma or Deterministic three possible relationships between $H(\cdot)$ and $G(\cdot)$ are considered: $G(\cdot)$ can be of the "same type" as $H(\cdot)$; $G(\cdot)$ can be derived from $H(\cdot)$ by assigning new arrivals to the individual channels in a cyclic order; and $G(\cdot)$ may be obtained from $H(\cdot)$ by assigning customers probabilistically to the different queues. The limiting behaviour of the ratio $R = W_q/W_q(s)$ is studied for the extreme values (1 and 0) of the common traffic intensity, ρ. Closed form results, which depend on the forms of $H(\cdot)$ and $G(\cdot)$ and on the relationships between them, are derived. It is shown that W_q is greater than $W_q(s)$ by a factor of at least $(s+1)/2$ when ρ approaches one, and that R is at least $s(s!)$ when ρ tends to zero. In the latter case, however, R goes to infinity (!) in most cases treated. The results may be used to evaluate the effect on the waiting times when, for certain (non-queueing) reasons, it is needed to partition a group of s servers into several small groups.

INTRODUCTION

IN DESIGNING a service system (of, for example, elevators, telephones or check-out counters) comprising several servers, it is well known that grouping the servers to give service in parallel is always "better" than partitioning them into several small groups. In certain situations, however, it is preferable to partition the servers instead of grouping all of them together. For example, a department store puts several cashiers in various locations so that customers will not have to walk long distances to pay for their purchases—although, from a strict queueing point of view, it is more efficient to put all the cashiers in one place. Another example is a telephone service (such as directory assistance) which handles many incoming calls. In various instances, there are limitations on the number of operators who may be hooked up into a single group; therefore, it may be technologically and economically better to partition the operators into several groups.

Here we compare a GI/M/s queueing system with its partition into s identical GI/M/1 queues, specifically considering a service system comprising $s \geqslant 2$ identical servers, where the interarrival times have common distribution function $H(\cdot)$ with mean $1/\lambda$, and the service times at each channel are i.i.d. Exponential random variables with mean $1/\mu$. The comparative system has s

separate but identical GI/M/1 queues such that, for each individual channel (= queue), the interarrival time distribution is $G(\cdot)$ with mean s/λ.

Denote by W_q the expected steady-state waiting time (excluding service) of an arbitrary customer in each of the GI/M/1 queues, and by $W_q(s)$ its corresponding value in the GI/M/s system. We are concerned with evaluating the ratio $R = W_q/W_q(s)$ of the mean waiting times in the two different configurations. In what follows we mainly deal with the Gamma family of interarrival time distributions and study the limiting properties of the ratio R when the traffic intensity $\rho = \lambda/(s\mu)$ approaches its extreme values, 0 or 1. It will be shown that these limits depend on the forms of the distributions $H(\cdot)$ and $G(\cdot)$ and the various possible relationships between them—i.e., on the way in which the general stream with intensity λ is separated into s individual streams, each with a mean rate of λ/s arrivals per unit time.

Three types of relationship between the distributions $H(\cdot)$ and $G(\cdot)$ will be treated. We will first assume that H and G are of the "*same type*". That is, if $H(\cdot)$ is Exponential, Gamma or Deterministic, so is $G(\cdot)$. Next we will assume that G is derived from H by *cyclic* assignment which systematically

TABLE 1. $\lim_{\rho \to 1} \left[R(\rho) = \dfrac{W_q}{W_q(s)} \right]$

Distribution of T \ The procedure in which G is derived from H	Same type	Cyclic: $(ks+i)$th arrival is directed to the ith channel	Probabilistic: A new arrival is directed to the ith channel with probability $P_i = 1/s$
Exponential	s	$\dfrac{s+1}{2}$	s
Gamma$(n, \lambda n)$	s	$\dfrac{sn+1}{n+1}$	$\dfrac{2sn-n+1}{n+1}$
Deterministic	s	s	$2s-1$

TABLE 2. $\lim_{\rho \to 0} \left[R(\rho) = \dfrac{W_q}{W_q(s)} \right]$

Distribution of T \ The procedure in which G is derived from H	Same type	Cyclic	Probabilistic
Exponential	∞	$s(s!)$	∞
Gamma	∞	$s(s!)^n$	∞
Deterministic	∞	∞	∞

directs every $(ks + i)$th customer $(k = 0, 1, 2, \ldots; i = 1, 2, \ldots, s)$ to the ith channel. Finally, we will treat the case where G is derived from H by assigning the arrivals *probabilistically* to the individual channels; i.e., a newly arrived customer is assigned to the ith channel with probability $P_i = 1/s$.

The results are summarized in Tables 1 and 2 where the limiting values of R are explicitly given for the Exponential, Gamma and Deterministic distributions. These results show that W_q is greater than $W_q(s)$ by a factor of at least $(s + 1)/2$ when ρ approaches one, and that R is at least $s(s!)$ when ρ tends to zero. In the latter case, however, R goes to infinity (!) in most cases treated.

SOME KNOWN RESULTS AND NOTATION

For the GI/M/s system we suppose that customers arrive at instants $t_0 = 0, t_1, t_2, \ldots, t_n, \ldots$; the interarrival times $T_n = t_n - t_{n-1}$ $(n = 1, 2, \ldots)$ are i.i.d. positive random variables having the common distribution function $P\{T \leqslant t\} = H(t)$ with mean $E(T) = 1/\lambda$. The service times $\{V_i\}$ at each channel are i.i.d. r.v. with d.f. $P(V_i \leqslant v) = 1 - e^{-\mu v}, v \geqslant 0$.

When the system is separated into s identical GI/M/1 queues, the epochs of arrival to an individual channel are $t_{n_0}, t_{n_1}, t_{n_2}, \ldots, t_{n_k}, \ldots$ where the interarrival times $X_k = t_{n_k} - t_{n_{k-1}}$ $(k = 1, 2, \ldots)$ are i.i.d. positive r.v. having the common d.f. $P(X \leqslant x) = G(x)$ with finite mean $E(X) = s/\lambda$. The service times at each separate channel have the same probabilistic characteristics as the service times at each of the channels in the GI/M/s system.

Define the traffic intensity by $\rho = \lambda/(s\mu)$. It is well known[1,2,3] that each of the two systems attains its steady-state regime if and only if $\rho < 1$. The analysis in this work is carried out under this assumption.

For the GI/M/1 model Pollaczek[2] has shown that the average waiting time of an arbitrary customer for his service to start is given by

$$W_q = \frac{\gamma}{\mu(1 - \gamma)} \tag{1}$$

where $z = \gamma$ is the unique solution in $(0, 1)$ of the equation

$$z = \tilde{G}[\mu(1 - z)] \tag{2}$$

and

$$\tilde{G}(\theta) = \int_0^\infty e^{-\theta x}\, \mathrm{d}G(x)$$

is the Laplace–Stieltjes transform of the distribution function $G(\cdot)$.[3]

For the GI/M/s system, the average waiting time is given by

$$W_q(s) = \frac{\alpha}{\mu s(1 - \alpha)D} \tag{3}$$

where $\omega = \alpha$, is the unique solution in (0, 1) of the equation

$$\omega = \tilde{H}[\mu s(1 - \omega)]. \tag{4}$$

D is calculated from the relation

$$D = 1 + (1 - \alpha) \sum_{i=0}^{s-2} \beta_i \tag{5}$$

where

$$\beta_{s-1} = 1; \; \beta_j = \sum_{i=j-1}^{s-2} \beta_i P_{ij} + \sum_{i=s-1}^{\infty} \alpha^{i-s+1} P_{ij} \quad (j = 0, 1, \ldots, s-1);$$

$$\beta_{-1} = 0, \tag{6}$$

and the P_{ij}'s are the one-step transition probabilities of the embedded Markov chain of the GI/M/s queue.

From (1) and (3) we readily have

$$R = \frac{W_q}{W_q(s)} = s\left[\frac{\gamma(1 - \alpha)}{\alpha(1 - \gamma)}\right] D. \tag{7}$$

In what follows it will be shown that, for the family of Gamma distributions, α and γ are functions of ρ. Thus, in the sequel, it will be understood that $\alpha = \alpha(\rho)$, $\gamma = \gamma(\rho)$ and $R = R(\rho)$. Also, in order to gain a better understanding of the results, we occasionally treat separately the Exponential and Deterministic distributions, although they are, respectively, a special case and a limiting case of the Gamma distribution.

THE CASE $\rho \to 1$

It is well known that the solutions $\gamma \in (0, 1)$ for (2) and $\alpha \in (0, 1)$ for (4) exist if and only if

$$f'(1) = \frac{1}{\rho} > 1 \quad \text{and} \quad F'(1) = \frac{1}{\rho} > 1,$$

respectively (where $f(\gamma) = \tilde{G}[\mu(1 - \gamma)]$ and $F(\alpha) = \tilde{H}[\mu s(1 - \alpha)]$). Since both $f(\gamma)$ and $F(\alpha)$ are convex increasing functions in (0, 1) and $f(1) = F(1) = 1$, it follows that both $\gamma \to 1$ and $\alpha \to 1$ as $\rho \to 1$. Moreover, using equation (6) we derive that

$$\beta_{s-j} \leqslant \prod_{i=s-1}^{s-j+1} (1/P_{i-1,i}) \quad \text{for } j = 2, 3, \ldots, s,$$

and from equation (34) in the Appendix it follows that for general inter-arrival-time distributions

$$\lim_{\rho \to 1} \sum_{i=0}^{s-2} \beta_i < \infty.$$

Thus, in this case, $D \to 1$ as $\rho \to 1$ and

$$\lim_{\rho \to 1} R(\rho) = s \lim_{\rho \to 1} \left[\frac{1 - \alpha}{1 - \gamma} \right]. \tag{8}$$

We now consider three different types of relation between $H(\cdot)$ and $G(\cdot)$:

A. $H(\cdot)$ and $G(\cdot)$ are of the "*same type*".

B. $G(\cdot)$ is derived from $H(\cdot)$ by *cyclic* assignment of arriving customers.

C. $G(\cdot)$ is derived from $H(\cdot)$ by *probabilistic* assignment of arrivals to the individual channels.

Proposition 1

If $H(\cdot)$ and $G(\cdot)$ are of the "same type" (with one of the above three distributions) then $\lim_{\rho \to 1} R(\rho) = s$.

Proof

(i) *The Gamma distribution.* Let $T \sim \text{Gamma}(n, \lambda n)$. That is, T has the Gamma distribution with shape parameter n and scale parameter λn. (The Exponential distribution is the special case when $n = 1$). In other words, the density function of $H(\cdot)$ is

$$h(t) = e^{-\lambda n t} \frac{(\lambda n)^n t^{n-1}}{(n-1)!}.$$

Similarly, we assume that $X \sim \text{Gamma}(n, \lambda n/s)$. Now from (2) and (4) we calculate

$$\alpha = \frac{(\rho n)^n}{[(1 - \alpha) + \rho n]^n} \tag{9}$$

and

$$\gamma = \frac{(\rho n)^n}{[(1 - \gamma) + \rho n]^n} \tag{10}$$

It follows immediately that $\alpha = \gamma$ and hence $\lim_{\rho \to 1} R(\rho) = s$.

(ii) *The Deterministic distribution.* In this case, we compare the D/M/s and the D/M/1 processes. From (4) and (2) we readily get

$$\alpha = e^{-(1-\alpha)/\rho} \tag{11}$$

$$\gamma = e^{-(1-\gamma)/\rho} \tag{12}$$

The obvious conclusion is that $\alpha = \gamma$ and therefore, again, $\lim_{\rho \to 1} R(\rho) = s$.

Q.E.D.

Remark 1

There are two meanings of "ρ approaching 1": It could mean "$\mu \to \lambda/s$", with λ held fixed. There is no problem here. Or it could mean "$\lambda \to s\mu$", with μ held fixed. In this case one must be specific about how the inter-arrival-time distribution changes as its mean changes. Since the Gamma (n-Erlang, to be precise) distributions are parameterized by λ we have no problem here either.

Remark 2

Another approach to the "same-type" case where, in fact, $\alpha = \gamma$ for *general* (and not just Gamma) interarrival–time distributions might be as follows. If we interpret the statement "of the same-type" to mean that T is distributed as Y/λ and X is distributed as $Y/(\lambda/s)$, where Y is a random variable with mean 1, then $\alpha = E\{\exp(-s\mu(1-\alpha)T)\} = E\{\exp(-(1-\alpha)Y/\rho)\}$ and $\gamma = E\{\exp(-\mu(1-\gamma)X)\} = E\{\exp(-(1-\gamma)Y/\rho)\}$. Hence $\alpha = \gamma$.

Giving the above results one may be inclined to conclude that, in general, for any $H(\cdot)$ and $G(\cdot)$ satisfying $E(T) = 1/\lambda$ and $E(X) = s/\lambda$, we will always have that $\alpha = \gamma$ or, at least, that $\lim_{\rho\to 1}[(1-\alpha)/(1-\gamma)] = 1$, and hence, $\lim_{\rho\to 1} R(\rho) = s$.

This, however, turns out to be incorrect and as was indicated above it will be shown that $R(\rho)$ depends on the specific way in which the general arrival stream for the GI/M/s queue is partitioned into the s individual streams.

We now consider *Cyclic Assignment* and suppose that every $(ks+i)$th arrival $(k = 0, 1, 2, \ldots; i = 1, 2, \ldots, s)$ is assigned to the ith queue. Thus if $\{T_m\}$ is the sequence of interarrival times to the GI/M/s system then the interarrival time X to each of the s separate identical channels is the sum of s i.i.d. random variables, i.e.,

$$X = \sum_{m=1}^{s} T_m.$$

Thus,

$$\gamma = \tilde{G}[\mu(1-\gamma)] = (\tilde{H}[\mu(1-\gamma)])^s. \tag{13}$$

If $T \sim \text{Gamma}(n, \lambda n)$ then from the properties of the Gamma distribution it follows that $X \sim \text{Gamma}(sn, \lambda n)$. We thus have, using (2), (4), and (13), that

$$\alpha = \frac{(\rho n)^n}{[(1-\alpha) + \rho n]^n} \tag{14}$$

and

$$\gamma = \frac{(\rho n)^{sn}}{[(1-\gamma)/s + \rho n]^{sn}} \tag{15}$$

In this case $\alpha \neq \gamma$ and the problem is to calculate $\lim_{\rho\to 1}[(1-\alpha)/(1-\gamma)]$ where both α and γ tend to 1 as ρ approaches 1.

Proposition 2

For the Cyclic Assignment in which $T \sim \text{Gamma}(n, \lambda n)$ and $X \sim \text{Gamma}(sn, \lambda n)$, we have $\lim_{\rho\to 1} R(\rho) = (sn + 1)/(n + 1)$.

Proof

Rewriting (14) as

$$\alpha \sum_{k=1}^{n} \binom{n}{k} (1 - \alpha)^k (\rho n)^{n-k} = (1 - \alpha)(\rho n)^n,$$

dividing by $(1 - \alpha)$, differentiating both sides with respect to ρ and taking limits as $\rho \to 1$ we obtain, after algebraic manipulations, that

$$\lim_{\rho\to 1} \alpha'(\rho) = 2n/(n + 1) \tag{16}$$

In a similar way we derive

$$\lim_{\rho\to 1} \gamma'(\rho) = 2sn/(sn + 1) \tag{17}$$

Thus, from (8)

$$\lim_{\rho\to 1} R(\rho) = \frac{sn + 1}{n + 1} \qquad \text{Q.E.D.} \tag{18}$$

The *Exponential* case is derived from (18) by letting $n = 1$, and in that case $R(\rho) \to (s + 1)/2$.

The *Deterministic* case can be obtained either directly or by letting n approach infinity. In the latter case (18) readily implies that $R(1) \to s$ as $n \to \infty$. In the former case, results (11) and (12) apply and the immediate consequence is that $\lim_{\rho\to 1} R(\rho) = s$. Note that (11) and (12) may be obtained from (14) and (15) by letting n approach infinity.

Comparing the results of the "same type" and the "cyclic" assignments it is clear that for $n < \infty$, $(sn + 1)/(n + 1) < s$. That is, by the cyclic assignment we reduce the variability of the interarrival times and therefore reduce the waiting times in the s separate channels. It is also interesting to note that $\lim_{\rho\to 1} R(\rho) = (sn + 1)/(n + 1)$ is non-decreasing in n.

Finally, we consider *Probabilistic Assignment* where the rule of assignment is the following:

Every new arrival is assigned to the ith channel with probability $P_i = 1/s$. If $\{T_m\}$ is the sequence of interarrival times to the GI/M/s system then the interarrival times $\{X_k^{(i)}\}$ to the ith channel $(i = 1, 2, \ldots, s)$ are i.i.d. random variables distributed as

$$X^{(i)} = \sum_{m=1}^{l} T_m, \qquad W.P.\ (1 - P_i)^{l-1} P_i \quad (l = 1, 2, \ldots) \tag{19}$$

That is

$$G_i(x) = P[X^{(i)} \leqslant x] = \sum_{l=1}^{\infty} P\left[\sum_{m=1}^{l} T_m \leqslant x\right](1 - P_i)^{l-1}P_i. \qquad (20)$$

The Laplace–Stieltjes transform of $G_i(x)$ is given by

$$\tilde{G}_i(\theta) = \frac{P_i\tilde{H}(\theta)}{1 - q_i\tilde{H}(\theta)} \qquad (21)$$

where $P_i = 1 - q_i$. Clearly, $E(X^{(i)}) = -\tilde{H}'(0) = E(T)/P_i$, and if $P_i = 1/s$ then $E(X^{(i)}) = s/\lambda$.

Now for each channel we have

$$\gamma_i = \tilde{G}_i[\mu(1 - \gamma_i)] = \frac{P_i\tilde{H}[\mu(1 - \gamma_i)]}{1 - q_i\tilde{H}[\mu(1 - \gamma_i)]}, \quad (i = 1, 2, \ldots, s) \qquad (22)$$

and if $P_i = 1/s$ we have $\gamma_i = \gamma$ for all i and

$$\gamma = \frac{\tilde{H}[\mu(1 - \gamma)]}{s - (s - 1)\tilde{H}[\mu(1 - \gamma)]} \qquad (23)$$

Assuming now that $T \sim \text{Gamma}(n, \lambda n)$ then

$$\tilde{H}(\theta) = [\lambda n/(\theta + \lambda n)]^n \qquad (24)$$

and hence

$$\gamma = \frac{(\rho n)^n}{s\left[\dfrac{1 - \gamma}{s} + \rho n\right]^n - (s - 1)(\rho n)^n}. \qquad (25)$$

Proposition 3

Under Probabilistic Assignment, if $T \sim \text{Gamma}(n, \lambda n)$ and γ is given by (23) then

$$\lim_{\rho \to 1} R(\rho) = \frac{2sn - n + 1}{n + 1}.$$

Proof

As before, from (16), $\alpha'(1) = 2n/(n + 1)$.

To derive $\gamma'(1)$ we apply on (25) the same method used to obtain (16) and we get

$$\gamma'(1) = \frac{2sn}{2sn - n + 1} \qquad (26)$$

Thus, finally,

$$\lim_{\rho \to 1} R(\rho) = \frac{2sn - n + 1}{n + 1} \qquad \text{Q.E.D.} \qquad (27)$$

When $n = 1$, T is *Exponential* (i.e., the arrival stream is Poisson), $\tilde{H}(\theta) = \lambda/(\theta + \lambda)$ and from (21),

$$\tilde{G}_i(\theta) = \frac{P_i\lambda}{\theta + P_i\lambda}$$

that is, the arrival stream to each of the individual channels is a Poisson one with intensity $P_i\lambda$. If we let $P_i = 1/s$ then the current case is identical with part (i) of the "same type" case and $\lim_{\rho\to 1} R(\rho) = s$. Clearly we may obtain this result simply by substituting $n = 1$ in (27).

The *Deterministic* case is derived from (27) by sending n to infinity, which yields, $\lim_{\rho\to 1} R(\rho) = 2s - 1$.

We summarize the above results in Table 1.

Investigation of Table 1 reveals that the cyclic assignment is the best, while the probabilistic assignment is the worst. Another conclusion is that for the family of Gamma distributions the ratio $\lim_{\rho\to 1} R(\rho)$ increases monotonically with n.

THE CASE $\rho \to 0$

The analysis in this case is somewhat more involved than for the case where $\rho \to 1$. The detailed proofs are given in the Appendix, and here we only state the results. These results hold for $T \sim \text{Gamma}(n, \lambda n)$.

Proposition 4

If $H(\cdot)$ and $G(\cdot)$ are of the "same type" then $\lim_{\rho\to 0} R(\rho) = \infty$. This result is somewhat surprising. The explanation is that, when $\rho \to 0$, the waiting time in the system with s servers goes faster to zero than the waiting time in the single server queue, although both waiting times approach zero as $\rho \to 0$.

Proposition 5

For the "Cyclic Assignment" case, $\lim_{\rho\to 0} R(\rho) = s(s!)^n$.

Proposition 6

For "Probabilistic Assignment", $\lim_{\rho\to 0} R(\rho) = \infty$.

We summarize these results in Table 2.

CONCLUSION

The average waiting time in a GI/M/s queue was compared with the average waiting time in an "equivalent" system comprised of s separate GI/M/1 queues. The results may be used to find the optimal partition of servers into several groups when designing a service system where certain restrictions eliminate the possibility of assigning all the servers in a single parallel group. The results may be qualitatively summarized as follows. If the inter-arrival times for both systems is of the "same type" then one would wait, on the average, at least s times as long in the separate system than in the combined one, when ρ goes to 1. When ρ becomes small the ratio of waiting times goes to infinity. By assigning arrivals probabilistically to the various individual channels we even worsen the situation. The best one can do is to assign the new arrivals systematically in a cyclic order such that the $(ks + i)$th customer $(k = 0, 1, 2, \dots; i = 1, 2, \dots, s)$ is always directed to the ith queue. Even in this case one waits in the system comprised of s individual GI/M/1 queues at least $(s + 1)/2$ times as long as in the GI/M/s queue.

APPENDIX

We first show that for the family of Gamma distributions and for the various relationships between $H(\cdot)$ and $G(\cdot)$, both α and γ approach zero as $\rho \to 0$.

Indeed, when $\rho \to 0$, $F'(1) \to \infty$ and $F(0) = \int_0^\infty e^{-\mu st}\,\mathrm{d}H(t)$. If $T \sim$ Gamma $(n, \lambda n)$ then

$$F(0) = \left(\frac{\rho n}{1 + \rho n}\right)^n \to 0.$$

Thus using (9), $\alpha \to 0$ as $\rho \to 0$. Similarly, $f'(1) \to \infty$ as $\rho \to 0$ and $f(0) = \tilde{G}(\mu)$. For the "same type" case it was shown that for $n = 1$, $1 < n < \infty$, and $n = \infty$, we always have $\alpha = \gamma$. Hence, $\gamma \to 0$ as $\rho \to 0$. For the cyclic case, $G \sim \text{Gamma}(sn, \lambda n)$ and hence

$$f(0) = \left(\frac{\lambda n}{\mu + \lambda n}\right)^{sn} = \left(\frac{\rho n}{\frac{1}{s} + \rho n}\right)^{sn} \to 0$$

as $\rho \to 0$. Thus, from (15), $\gamma \to 0$ as $\rho \to 0$. For the probabilistic case, using (23) and (24), one obtains

$$f(0) = [\rho n/(1/s + \rho n)]^n / \{s - (s - 1)[\rho n/(1/s + \rho n)]^n\} \to 0.$$

Hence from (25), $\gamma \to 0$ as $\rho \to 0$. Q.E.D.

Next we show that

$$\lim_{\rho \to 0} D = \infty. \tag{28}$$

To show this we use equation (5) and observe that, since $\beta_j > 0$ $(j = 0, 1, \dots,$

$s-2$) it suffices to show that $\beta_{s-2} \to \infty$ as $\rho \to 0$. From (6) it follows that

$$\lim_{\rho \to 0} \beta_{s-2} = \lim_{\rho \to 0} [(1 - P_{s-1,s-1})/P_{s-2,s-1}]$$

where, for the GI/M/s queue,

$$P_{s-2,s-1} = \int_0^\infty e^{-\mu t(s-1)}\, \mathrm{d}H(t) = (\rho n)^n \Big/ \left(\rho n + \frac{s-1}{s}\right)^n. \tag{29}$$

Clearly, $P_{s-2,s-1} \to 0$ and, since $P_{s-1,s-1} \to 0$ (see (34)), $\beta_{s-2} \to \infty$ as $\rho \to 0$. Q.E.D.

In view of the above results and equation (7) our problem now is to find

$$\lim_{\rho \to 0} R(\rho) = s \lim_{\rho \to 0} \left(\frac{\gamma}{\alpha} D\right), \tag{30}$$

where both γ and α tend to zero as ρ does.

Proof of proposition 4

When we considered the "Same Type" case for $\rho \to 1$ it was shown that we always have $\alpha = \gamma$. Thus, using (30), it readily follows from (28) that

$$\lim_{\rho \to 0} R(\rho) = s \lim_{\rho \to 0} D = \infty \qquad \text{Q.E.D.} \tag{31}$$

Proof of proposition 5

We first note that

$$\lim_{\rho \to 0} D = 1 + \sum_{i=1}^{s-2} \lim_{\rho \to 0} \beta_i \tag{32}$$

and from (6), for $j = 0, 1, 2, \ldots, s-1$

$$\lim_{\rho \to 0} \beta_j = \lim_{\rho \to 0} \left[\sum_{i=j-1}^{s-2} \beta_i P_{ij} + P_{s-1,j} \right] \tag{33}$$

Now, for $j \leqslant i + 1 \leqslant s$ we have[1,3]

$$\begin{aligned} P_{ij} &= \int_0^\infty \binom{i+1}{j} (1 - e^{-\mu t})^{i+1-j} e^{-\mu t j} e^{-\lambda n t} \frac{(\lambda n)^n t^{n-1}}{(n-1)!} \mathrm{d}t \\ &= \binom{i+1}{j} \sum_{k=0}^{i+1-j} \binom{i+1-j}{k} (-1)^k \frac{(\rho n)^n}{\left[\rho n + \dfrac{k+j}{s}\right]^n} \end{aligned} \tag{34}$$

Using (34) and recursively applying (33) (with $\beta_{s-1} = 1$) we obtain

$$\lim_{\rho \to 0} \beta_j = \frac{\left(\dfrac{s-1}{s}\right)^n \left(\dfrac{s-2}{s}\right)^n \cdots \left(\dfrac{j+1}{s}\right)^n}{\lim\limits_{\rho \to 0} (\rho n)^{(s-j-1)n}} \quad (j = s-2, s-3, \ldots, 1). \tag{35}$$

Now, since $\lim_{\rho\to 0} P_{j0} = 1$ for $j = 0, 1, \ldots, s-1$ we have, from (33),

$$\lim_{\rho\to 0} \beta_0 = \frac{\sum_{j=1}^{s-2} \lim_{\rho\to 0} \beta_j + 1}{\lim_{\rho\to 0}(1 - P_{00})} \tag{36}$$

but

$$\lim_{\rho\to 0}(1 - P_{00}) = \lim_{\rho\to 0}\left(\frac{\rho n}{\rho n + \frac{1}{s}}\right)^n$$

thus,

$$\lim_{\rho\to 0} \beta_0 = \lim_{\rho\to 0}\left[\sum_{j=1}^{s-2} \frac{\left(\frac{s-1}{s}\right)^n \left(\frac{s-2}{s}\right)^n \cdots \left(\frac{j+1}{s}\right)^n}{(\rho n)^{(s-j-1)n}} + 1\right]\left(\frac{\rho n + \frac{1}{s}}{\rho n}\right)^n \tag{37}$$

Now from (14), (15) and (35) it follows that for $j = 1, 2, \ldots, s-1$

$$\lim_{\rho\to 0}\left(\frac{\gamma}{\alpha}\beta_j\right) = \lim_{\rho\to 0}\left\{\frac{(\rho n)^{sn}[(1-\alpha) + \rho n]^n}{[(1-\gamma)/s + \rho n]^{sn}(\rho n)^n}\beta_j\right\} = 0 \tag{38}$$

and, therefore, using (32) and (37) we finally have

$$\lim_{\rho\to 0} R(\rho) = s \lim_{\rho\to 0}\left(\frac{\gamma}{\alpha}\beta_0\right) = s\,\frac{\left(\frac{s-1}{s}\right)^n \left(\frac{s-2}{s}\right)^n \cdots \left(\frac{2}{s}\right)^n \left(\frac{1}{s}\right)^n}{(1/s)^{sn}} = s(s!)^n \tag{39}$$

Q.E.D.

Clearly, for the *Deterministic* case (when $n = \infty$) $\lim_{\rho\to 0} R(\rho) = \infty$. On the other hand, a direct derivation would give $\alpha = \gamma$ from which it follows again that $\lim_{\rho\to 0} R(\rho) = \infty$.

The *Exponential* case may be obtained from (39) by substituting $n = 1$ which yields $\lim_{\rho\to 0} R(\rho) = s(s!)$.

Proof of proposition 6

For $T \sim \text{Gamma}(n, \lambda n)$ consider again equation (14), and rewrite it as

$$\alpha y(\alpha, \rho) = n^n \rho^n \tag{40}$$

where

$$y(\alpha, \rho) = \sum_{k=1}^{n} \binom{n}{k} (1-\alpha)^{k-1} (\rho n)^{n-k}. \tag{41}$$

Clearly, $y \to 1$ as $\rho \to 0$.

Taking derivatives of both sides of (40) yields

$$\alpha' y(\alpha, \rho) + \alpha y'(\alpha, \rho) = n^n n \rho^{n-1}. \tag{42}$$

Passing to the limit gives $\lim_{\rho \to 0} \alpha' = 0$. Continuing differentiating in this manner we obtain $\lim_{\rho \to 0} \alpha^{(k)} = 0$ for $k = 1, 2, \ldots, n-1$, where $\alpha^{(k)}$ denotes the kth derivative with respect to ρ. Only for the nth derivative we have

$$\lim_{\rho \to 0} \alpha^{(n)} = n^n n!. \tag{43}$$

Considering (25) and applying the same method as above, we get $\lim_{\rho \to 0} \gamma^{(k)} = 0$ for $k = 1, 2, \ldots, n-1$ and

$$\lim_{\rho \to 0} \gamma^{(n)} = s^{n-1} n^n n!. \tag{44}$$

Hence,

$$\lim_{\rho \to 0} R(\rho) = s^n \lim_{\rho \to 0} D = \infty. \qquad \text{Q.E.D.} \tag{45}$$

ACKNOWLEDGEMENT

The author wishes to thank Mrs. Varda Liberman for very helpful discussions.

REFERENCES

[1] S. Karlin (1966) *A First Course in Stochastic Processes*, Academic Press, New York.

[2] F. Pollaczek (1953) Sur une généralisation de la théorie des attentes, *C.r. Séanc. Acad. Sci. (Paris)* **236**, 578–580.

[3] L. Takács (1962) *Introduction to the Theory of Queues*, Oxford University Press, New York and London.