ON A MANY SERVER QUEUE WITH NON-RECURRENT INPUT AND NEGATIVE EXPONENTIAL SERVERS

C. PEARCE

(Received 7 February 1967, revised 7 August 1967)

1. Introduction

We consider a queueing system with k identical servers in parallel, the services being negative exponential with parameter μ. The input is a natural generalisation of the usual general recurrent input. If we denote the sequence of arrival points by $\{A_n, n \geqq 0\}$ then the inter-arrival intervals are given by

$$A_{n+1}-A_n = f_0(U_{n+p})+f_1(U_{n+p-1})+\cdots+f_p(U_n), \qquad n \geqq 0, \tag{1.1}$$

where the f_i are (integrable) non-negative functions and $\{U_n\}$ is a sequence of identically and independently distributed random variables. In the simplest case, $p = 0$, this is just a general recurrent input. We write $U(\cdot)$ for the probability distribution function of the U_n.

The aim of this paper is to determine the equilibrium distribution (when it exists) of the number of customers in the system as found by arrivals. We also consider for a particular instance of (1.1) the case when the waiting room is finite (Section 5) and the case of infinitely many servers (Section 6).

2. The equilibrium queuelength distribution

We denote by $P(u_0, u_1, \cdots, u_{m+p-1})$, or, more compactly, by

$$P_j(u^{(m+p-1)}), \qquad j \geqq 0,$$

the probability, conditional on the values assumed by the first $m+p$ members of the sequence $\{U_n\}$ being $u_0, u_1, \cdots, u_{m+p-1}$, that the number of customers found in the system by the arrival at A_m is j.

From the theory of the simple death process we have that $\Gamma_{ij}(u)$, the probability that the number of customers in the system is initially i and finally j in a time interval of length u during which there are no arrivals, is given by

$$(2.1)\quad \Gamma_{ij}(u) = \begin{cases} \binom{i}{j}[1-\exp(-\mu u)]^{i-j}\exp(j\mu u), & j \leqq i \leqq k, \\ \exp(-k\mu u)(k\mu u)^{i-j}/(i-j)!, & k \leqq j \leqq i, \\ \int_0^u \exp(-k\mu x)\mu k(\mu k x)^{i-k-1}/(i-k-1)!\,\Gamma_{kj}(u-x)dx, & j < k < i, \\ 0, \quad \text{otherwise.} \end{cases}$$

By comparing the numbers of customers in the system at

$$A_{m+1}-0,\ A_m-0,$$

we see that

$$(2.2)\quad P_j(u^{(m+p)}) = \begin{cases} \sum_{i=0}^{\infty} P_i(u^{(m+p-1)})\Gamma_{i+1,0}(f_0(U_{m+p})+\cdots+f_p(U_m)), & j=0, \\ \sum_{i=0}^{\infty} P_{j+i-1}(u^{(m+p-1)})\Gamma_{j+i,j}(f_0(U_{m+p})+\cdots+f_p(U_m)), & j>0. \end{cases}$$

With the traffic intensity condition

$$(2.3)\qquad \sum_{i=0}^{p} Ef_i > (\mu k)^{-1},$$

we know from Loynes [1] that the limits

$$P_j = \lim_{m\to\infty} E[P_j(U^{(m+p)})], \qquad j \geqq 0,$$

exist and constitute a proper probability distribution, and we can justify the definitions

$$(2.4)\quad \begin{aligned} Q_j(\omega_1,\cdots,\omega_p) &= \lim_{m\to\infty} E[P_j(U_0,\cdots,U_m,u_{m+1},u_{m+2},\cdots,u_{m+p})], & j \geqq 0, \\ Q(\omega_1,\cdots,\omega_p;z) &= \sum_{j=0}^{\infty} Q_{j+k-1}(\omega_1,\cdots,\omega_p)z^j/j!, & |z| < \infty, \end{aligned}$$

where the expectation is with respect to $U_0, \cdots, U_{m-1}$, and where $\omega_1, \cdots, \omega_p$ are the particular values $u_{m+1}, \cdots, u_{m+p}$.

On taking expectations and letting $m \to \infty$, we have from (2.2) and (2.1) the following equation for $Q(\omega_1, \cdots, \omega_p; z)$:

$$(2.5)\quad \begin{aligned} \frac{\partial}{\partial z} Q(\omega_1,\cdots,\omega_p;z) = \int_0^{\infty} & Q(u,\omega_1,\cdots,\omega_{p-1};z+k\mu\{f_0(\omega_p)+f_p(u)\}) \\ & \times \exp[-k\mu\{f_0(\omega_p)+\cdots+f_p(u)\}]dU(u). \end{aligned}$$

If we define

https://doi.org/10.1017/S1446788700006522 Published online by Cambridge University Press

$$R(\omega_1, \cdots, \omega_p; z) = Q(\omega_1, \cdots, \omega_p; z) \exp(-z), \ |z| < \infty, \tag{2.6}$$

(2.5) becomes

$$\left(\frac{\partial}{\partial z}+1\right) R(\omega_1, \cdots, \omega_p; z) = \int_0^\infty R(u, \omega_1, \cdots, \omega_{p-1}; \\ z+\mu\{f_0(\omega_p)+\cdots+f_p(u)\})dU(u). \tag{2.7}$$

Inspection of the right hand side of this equation shows that z and ω_p occur in $((\partial/\partial z)+1)R(\omega_1, \cdots, \omega_p; z)$ only in the combination $z+\mu f_0(\omega_p)$. It follows readily that $R(\omega_1, \cdots, \omega_p; z)$ can be written in the form

$$R(\omega_1, \cdots, \omega_p; z) \\ = g(\omega_1, \cdots, \omega_p) \exp(-z)+R_1(\omega_1, \cdots, \omega_{p-1}, z+\mu f_0(\omega_p)), \ |z| < \infty.$$

If this expression for R is substituted in (2.7) and a second differentiation with respect to z performed, we find that

$$\left(\frac{\partial}{\partial z}+1\right)^2 R_1(\omega_1, \cdots, \omega_{p-1}, z) = \int_0^\infty \left(\frac{\partial}{\partial z}+1\right) R_1(u, \omega_1, \cdots, \omega_{p-2}, \\ z+\mu\{f_0(\omega_{p-1})+f_1(\omega_{p-1})+f_2(\omega_{p-2})+\cdots+f_p(u)\})dU(u).$$

Arguing as before we obtain a decomposition

$$R_1(\omega_1, \cdots, \omega_{p-1}, z) = [g_0(\omega_1, \cdots, \omega_{p-1})+zg_1(\omega_1, \cdots, \omega_{p-1})] \exp(-z) \\ +R_2(\omega_1, \cdots, \omega_{p-2}, z+\mu\{f_0(\omega_{p-1})+f_1(\omega_{p-1})\}),$$

and we can continue recursively to eventually obtain, analogously to (2.7), an equation

$$\left(\frac{\partial}{\partial z}+1\right)^{p+1} R_p(z) \\ = \int_0^\infty \left(\frac{\partial}{\partial z}+1\right)^p R_p(z+\mu\{f_0(u)+\cdots+f_p(u)\}) dU(u), \ |z| < \infty. \tag{2.8}$$

Considered as an equation for $((\partial/\partial z)+1)^p R_p$, (2.8) has a unique entire solution (up to a scale factor), which is easily seen to be

$$\left(\frac{\partial}{\partial z}+1\right)^p R_p(z) = \text{const} \times \exp[-z(1-T)], \tag{2.9}$$

where T is the (unique) solution of (2.8)

$$T = \int_0^\infty \exp[-\mu(1-T)(f_0(u)+\cdots+f_p(u))] dU(u)$$

inside the unit circle in the complex plane. That such a T is well-defined

https://doi.org/10.1017/S1446788700006522 Published online by Cambridge University Press

follows from (2.3) by a standard argument making use of Rouché's theorem. (2.9) finally gives us R_p in the form

$$R_p(z) = [a_0+a_1z+\cdots+a_{p-1}z^{p-1}]\exp(-z)+a_p\exp[-z(1-T)],$$

where the a_i are constants.

The successive decompositions of R give, via (2.4) and (2.6), the Q_j, $j \geqq k$, in the form

$$Q_{k+r}(\omega_1,\cdots,\omega_p) = \begin{cases} h_r(\omega_1,\cdots,\omega_{p-1-r})\exp\left(-\mu\sum_{i=p-r}^{p}\sum_{j=0}^{p-i}f_j(\omega_i)\right) & \\ \quad +hT^{k+r}\exp\left[-\mu(1-T)\sum_{i=1}^{p}\sum_{j=0}^{p-i}f_j(\omega_i)\right], & 0 \leqq r \leqq p-2, \\ hT^{k+r}\exp\left[-\mu(1-T)\sum_{i=1}^{p}\sum_{j=0}^{p-i}f_j(\omega_i)\right], & r \geqq p-1. \end{cases}$$

On taking expectations with respect to the ω_i we see that the limiting distribution $\{p_j, j \geqq 0\}$ is purely geometric from P_{k+p-1} onwards with common ratio T. This result has been established for the general recurrent input, i.e., in the case $p = 0$, by Kendall [2]. To complete the solution we form equations for $Q_0,\cdots,Q_{k-1}, h_0,\cdots,h_{p-2}, h$ from the relations (2.2) ($j = 1,\cdots,k+p-1$) together with the normalising relation

$$\sum_{j=0}^{\infty} P_j(u^{n+p}) = 1.$$

The resultant equations can be solved for the unknown functions through simple elimination and recursive substitution. The P_j can then be obtained by taking expectations.

3. Limiting distribution of waiting times

If an arrival finds $k-1$ or fewer customers already in the system he does not have to wait to commence service. If he finds $j \geqq k$ customers already present, he has to wait until $j+1-k$ of these have been served before his service can commence. As the probability of precisely $j+1-k$ such services being completed in a time $\leqq x$ is known as

$$\exp(-k\mu x)(k\mu x)^{j+1-k}/(j+1-k)!,$$

the waiting time for arrivals can readily be determined.

If we write

$$P_j = BT^j, \qquad j \geqq k+p-2,$$

then the limiting probability that an arrival has to wait a time $\leqq x$ before he can begin service is

https://doi.org/10.1017/S1446788700006522 Published online by Cambridge University Press

$$F(x) = \sum_{j=0}^{k-1} P_j + \sum_{j=k}^{\infty} P_j[1 - \sum_{i=0}^{j-k} \exp(-k\mu x)(k\mu x)^i/i!]$$
$$= 1 - \sum_{i=0}^{p-2} \sum_{j=i}^{p-2} P_{k+j} \exp(-k\mu x)(k\mu x)^i/i!$$
$$- T^{k+p-2}(1-T)^{-1} B \exp\left(-k\mu x(1-T)\right), \qquad x \geqq 0, p \geqq 2.$$

When $p = 0$, i.e., when the input is general recurrent, it is readily verified that the waiting time distribution reduces to a weight at the origin combined with a negative exponential distribution, a result first noted by Kendall [2].

4. Imbedded chain method

For simplicity we restrict our attention to the case $p = 1$ with f_0 and f_1 scalar multipliers b_0 and b_1, though these restrictions are not necessary. It is also convenient to take b_0, b_1 positive with sum unity. We shall see although the arrival instants are no longer recurrence points, it is possible to extend the usual imbedded chain technique as used, for example, for $GI/M/1$, to facilitate the analysis of our more general systems.

With the simplifications above, (1.1) reduces to

$$A_{n+1} - A_n = b_0 U_{n+1} + b_1 U_n, \qquad n \geqq 0. \tag{4.1}$$

The point R_n, $n \geqq 0$, occurring a time $b_1 U_n$ after A_n can be intuitively regarded as the instant at which the effect of U_n in the input stream ceases and that of U_{n+1} commences. The points A and R occur alternately in time, R_n segmenting the interval (A_n, A_{n+1}) into positions of length $b_1 U_n$, $b_0 U_{n+1}$, and A_{n+1} segmenting (R_n, R_{n+1}) into portions $b_0 U_{n+1}$, $b_1 U_{n+1}$. With the proviso that arrivals are never multiple the sequences $\{A_n\}$, $\{R_n\}$ are disjoint and interlacing. We shall consider the number of customers in the system at the regenerative time points $\{R_n\}$.

Let P_{ij} represent the probability that j individuals are waiting or being served at a regenerative point, conditional on the corresponding number at the previous regenerative point being i $(i, j \geqq 0)$.

The only arrival during (R_m, R_{m+1}) is at A_{m+1}, and we can write down directly

$$P_{ij} = \begin{cases} \displaystyle\int_0^\infty \sum_{r=0}^{i} \Gamma_{i,i-r}(b_0 u)\Gamma_{i-r+1,0}(b_1 u)dU(u), & j = 0, \\ \displaystyle\int_0^\infty \sum_{r=0}^{i-j+r} \Gamma_{i,i-r}(b_0 u)\Gamma_{i-r+1,j}(b_1 u)dU(u), & 0 < j \leqq i+1, \\ 0, & j > i+1. \end{cases} \tag{4.2}$$

It is readily verified that the number of customers present, observed at the points of the sequence $\{R_n, n \geqq 0\}$, constitutes an irreducible, aperiodic Markov chain. By the standard Feller theory, to establish ergodicity on the imbedded chain it suffices to find a non-zero vector $\boldsymbol{x}$ whose components form an absolutely convergent series and for which

$$x_j = \sum_{i=0}^{\infty} x_i P_{ij}, \qquad j \geqq 0. \tag{4.3}$$

Such a vector, if it exists, must be unique to within a scale factor, and, when the sum of the components is normalized to unity, $\boldsymbol{x}$ becomes the equilibrium distribution of probability in the system. We shall show that under the traffic intensity condition

$$EU > (\mu k)^{-1} \tag{4.4}$$

such a vector $\boldsymbol{x}$ exists.

From (4.2) we have that, with a trial

$$\boldsymbol{x} = \{x_0, \cdots, x_{k-1}, 1, T, T^2, \cdots\},$$

(4.3) reduces to

$$T = \int_0^\infty \exp\left[-\mu uk(1-T)\right] dU(u) \tag{4.5}$$

for $j > k$. By virtue of (4.4) Rouché's theorem yields that (4.5) has a (unique) solution inside the unit circle. We have thus established ergodicity under condition (4.4) provided we can find a set $x_0, \cdots, x_{k-1}$ of values satisfying (4.3) for $0 \leqq j \leqq k$. To prove that this can be done, we note from (4.2) that $P_{ij} = 0$ for $i < j-1$, so that, as $P_{j-1,j} > 0$ for $j > 0$, the equations (4.3) for $j = k, k-1, \cdots, 1$ can be used for the successive determination of $x_{k-1}, x_{k-2}, \cdots, x_0$ in terms of T. That

$$x_0 = \sum_{i=0}^{\infty} x_i P_{i0}$$

is consistent with the values so obtained follows from the fact that the terms of each row of the matrix (P_{ij}) sum to unity.

Under (4.4), therefore, the number of customers present as observed at regenerative points has a proper equilibrium probability distribution, and this is of the form

$$P\text{ (number of customers} = j) = T^{j-k}\pi_k, \qquad j \geqq k.$$

The instants chosen as regenerative points have physical significance only in the input process, and it is of somewhat more interest in the queueing system to know the number of customers present at arrival instants.

The first arrival after a regenerative point R_m is at A_{m+1}, and the length $b_0 U_{m+1}$ of the interval (R_m, A_{m+1}) is independent of the previous history of the system. As the number of services completed during (R_m, A_{m+1})

https://doi.org/10.1017/S1446788700006522 Published online by Cambridge University Press

depends only on the number of customers present at R_m and on the time b_0U_{m+1}, the number of customers present as found by an arrival must possess a proper limiting distribution.

If P_j represents the stationary probability that an arrival finds j customers, then

$$P_j = \int_0^\infty \sum_{r=0}^\infty \pi_{j+r}\Gamma_{j+r,j}(b_0u)dU(u), \qquad j \geqq 0.$$

Use of the known form of $\{\pi_i, i \geqq 0\}$ and of the Γ_{ij} gives

$$P_j = \begin{cases} \int_0^\infty \sum_{r=0}^{k-j} \pi_{j+r}\Gamma_{i+r,j}(b_0u)dU(u) \\ \quad + \int_0^\infty \pi_k T\mu k \left[\int_0^u \exp[-k\mu x(1-T)]\Gamma_{kj}(u-x)dx\right] dU(u), & 0 \leqq j < k, \\ T^{j-k}\pi_k \int_0^\infty \exp\,[-k\mu b_0 u(1-T)]dU(u), & j \geqq k. \end{cases}$$

A similar argument enables us to determine the stationary probability distribution $\{\rho_j, j \geqq 0\}$ in continuous time. In particular, we can easily show that $\{\rho_j\}$ is geometric from ρ_{k+p} onwards with common ratio T. For, by considering the mean probability of there being j customers in the system over an interval of fixed length separating a pair of consecutive regenerative points, and then allowing this length to vary, we obtain

$$\begin{aligned} \rho_j &= \int_0^\infty \Bigg[\int_0^{b_0u} \sum_{r=0}^\infty \pi_{j+r} \exp\,(-k\mu x)(k\mu x^r)/r!\,dx \\ &\qquad + \int_0^{b_1u} \sum_{r=0}^\infty \pi_{j-1+r} \exp\,(-k\mu(b_0u+x))\,(k\mu(b_0\mu+x))^r/r!\,dx\Bigg] dU(u) \\ &= \pi_{j-1}(\mu k)^{-1} \int_0^\infty \exp\,[-\mu k b_0(1-T)]\,dU(u), \qquad j \geqq k+p. \end{aligned}$$

As stated earlier, the imbedded chain technique can be used with rather more involved problems than those of this section, although the algebra rapidly becomes prohibitive. Consider, for instance, an input for which

$$A_{n+1} - A_n = b_0U_{n+p} + \cdots + b_pU_n, \qquad n \geqq 0,$$

where the b_i are positive constants with sum unity. Here we can select points R_n later than the corresponding arrival points A_n by amounts

$$\sum_{r=0}^{p-1} U_{n+r} \sum_{s=0}^{r} b_{p-s}.$$

The interval (R_n, R_{n+1}) will then have length U_{n+p}, so that the lengths of the intervals (R_n, R_{n+1}) form an identically and independently distributed sequence. For $p > 1$, R_n will not necessarily lie between A_n and A_{n+1} although it will lie between A_n and A_{n+p-1}.

Similar calculations to ours have been performed by Winsten [3], who also uses an imbedded chain whose time points do not correspond to instants of physical discontinuity of the system.

For the idea of extending our method to general p we are indebted to Dr R. M. Loynes.

5. Finite waiting room

The methods of the previous section provide a very easy approach to the consideration of a queue with an input satisfying (4.1) when the size of the waiting room is restricted.

Suppose that the number of customers waiting and in service is restricted to $m \geqq k$, and that any arrivals finding m customers in the system are lost.

(2.1) is still valid, and corresponding to (4.2) we have

$$P_{ij} = \begin{cases} \int_0^\infty \sum_{r=0}^{i} \Gamma_{i,i-r}(b_0 u)\Gamma_{i-r+1,0}(b_1 u)dU(u), & j=0, \quad i \neq m, \\ \int_0^\infty \sum_{r=0}^{i-j+1} \Gamma_{i,i-r}(b_0 u)\Gamma_{i-r+1,j}(b_1 u)dU(u), & 0 < j \leqq i+1, \quad i \neq m, \\ \int_0^\infty \sum_{r=0}^{m-j} \Gamma_{m,m-r}(b_0 u)\Gamma_{m-r+1,j}(b_1 u)dU(u) & \\ \qquad + \int_0^\infty \Gamma_{m,m}(b_0 u)\Gamma_{m,j}(b_1 u)dU(u), & 0 \leqq j < m, \quad i = m, \\ \int_0^\infty \Gamma_{m,m}(b_0 u)\Gamma_{m,m}(b_1 u)\, dU(u), & i = j = m, \\ 0, & j > i+1. \end{cases}$$

As before the Markov chain is clearly aperiodic and irreducible, so since the state space is finite, the chain must be ergodic. The practical solution of the equations

$$x_j = \sum_{i=0}^{m} x_i P_{ij}, \qquad 0 \leqq j \leqq m,$$

again proceeds by a simple recursion.

6. Infinitely many servers

We also give an explicit solution for the equilibrium distribution of the number of busy servers for a system with an infinite number of servers and an input of type (4.1), working with the regenerative points $\{R_n\}$.

It is easily shown that the equilibrium distribution $\{P_i, i \geqq 0\}$ always exists. We denote its generating function by

$$P(z) = \sum_{i=0}^{\infty} P_i z^i, \qquad |z| \leqq 1.$$

We now have

$$\Gamma_{ij}(u) = \begin{cases} \binom{i}{j} [1-\exp(-\mu u)]^{i-j} \exp(-j\mu u), & 0 \leqq j \leqq i, \\ 0 & j > i, \end{cases}$$

and (P_{ij}) maintains the form

$$P_{ij} = \begin{cases} \int_0^\infty \sum_{r=0}^{i} \Gamma_{i,i-r}(b_0 u)\Gamma_{i-r+1,0}(b_1 u)dU(u), & j = 0, \\ \int_0^\infty \sum_{r=0}^{i-j+1} \Gamma_{i,i-r}(b_0 u)\Gamma_{i-r+1,j}(b_1 u)dU(u), & 0 < j \leqq i+1, \\ 0 & j > i+1. \end{cases}$$

Substitution of these values in the equations

$$P_j = \sum_{j=0}^{\infty} P_i P_{ij}, \qquad j \geqq 0,$$

and formation of the generating function $P(z)$ from the left hand sides of these equations gives

$$P(z) = \int_0^\infty [1+(z-1)\exp(-\mu b_0 u)]P[1+(z-1)\exp(-\mu u)]dU(u). \tag{6.1}$$

We now presume the legitimacy of the definitions

$$q_j = (1/j!)[d^j/dz^j P(z)]_{z=1}, \qquad j \geqq 0,$$

our a posteriori justification residing in the fact that we shall solve for the q_j and obtain finite values.

Differentiation of (6.1) yields immediately

$$q_j = q_j \int_0^\infty \exp(-j\mu u)dU(u) + q_{j-1}\int_0^\infty \exp(-\mu u(j-b_1))dU(u), \quad j \geqq 1,$$

whence

$$q_j = q_0 \prod_{i=1}^{j} \left[\int_0^\infty \exp(-\mu u(i-b_1))dU(u) \times \left\{1 - \int_0^\infty \exp(-\mu u i)dU(u)\right\}^{-1} \right], \qquad j \geqq 1.$$

We note that, since

https://doi.org/10.1017/S1446788700006522 Published online by Cambridge University Press

$$\int_0^\infty \exp(-\mu ux)dU(u) \to 0, \text{ as } x \to \infty,$$

the expression inside the square brackets in (6.2) is less than $\frac{1}{2}$ for all sufficiently large i, so that the q_j are bounded. Our assumption is clearly justified.

The solution (6.2) for the q_j is completed by observing that

$$q_0 = P(1) = 1.$$

The equilibrium probabilities $\{P_j\}$ can now be determined from

$$P_j = (1/j!)[d^j/dz^j \sum_{i=0}^{\infty} q_i(z-1)^i]_{z=0}$$
$$= \sum_{i=j}^{\infty} (-)^{i-j} \binom{i}{j} q_i.$$

As before we can deduce the equilibrium distribution $\{\rho_j, j \geqq 0\}$ of the number of busy servers as found by arrivals.

An arrival will find j servers occupied if and only if there were $j+m$ $(m \geqq 0)$ customers in the system at the last regenerative point and there have subsequently been m departures. Thus

$$\rho_j = \sum_{m=0}^{\infty} P_{j+m} \int_0^\infty \Gamma_{j+m,j}(b_0 u)dU(u), \qquad j \geqq 0.$$

References

[1] R. M. Loynes, 'The stability of a queue with non-independent arrival and service times', *Proc. Camb. Phil. Soc.* 58 (1962), 497—520
[2] D. G. Kendall, 'Stochastic processes occurring in the theory of queues and their analysis by means of the imbedded Markov chain', *Ann. Math. Statist.* 24 (1953), 338—54.
[3] C. B. Winsten, 'Geometric distributions in the theory of queues', *J. Roy. Statist. Soc. B* 21 (1959), 1—35.

Department of Probability and Statistics
Sheffield University

https://doi.org/10.1017/S1446788700006522 Published online by Cambridge University Press