SINGLE SERVER QUEUES WITH MODIFIED SERVICE MECHANISMS

G. F. YEO

(received 17 July, 1961, revised 12 April 1962)

Summary

This paper considers a generalisation of the queueing system $M/G/1$, where customers arriving at empty and non-empty queues have different service time distributions. The characteristic function (c.f.) of the stationary waiting time distribution and the probability generating function (p.g.f.) of the queue size are obtained. The busy period distribution is found; the results are generalised to an Erlangian inter-arrival distribution; the time-dependent problem is considered, and finally a special case of server absenteeism is discussed.

1. Introduction

Let us consider the following generalisation of the queueing system $GI/G/1$. In a single server system customers arrive at the instants $\tau_1, \tau_2, \cdots$, such that $t_n = \tau_n - \tau_{n-1} (n \geqq 1)$, $t_0 = 0$, are independently and identically distributed random variables with common distribution functions (d.f.'s) $A(x)$, finite expectations $a = E(t_n) = \int_0^\infty x dA(x)$ and c.f.'s $\alpha(\theta) = \int_{0-}^\infty e^{i\theta x} dA(x) (\mathscr{I}(\theta) \geqq 0)$. If the n-th arrival joins a non-empty queue let his service time be s_n, while if he joins an empty queue let it be r_n; s_n and $r_n (n \geqq 1)$ are sequences of independently and identically distributed random variables with common d.f.'s $B(x)$ and $D(x)$ respectively, finite but non-zero expectations $b = E(s_n) = \int_0^\infty x dB(x)$ and $d = E(r_n) = \int_0^\infty x dD(x)$, and c.f.'s $\psi(\theta) = \int_{0-}^\infty e^{i\theta x} dB(x)$ and $\zeta(\theta) = \int_{0-}^\infty e^{i\theta x} dD(x)$. Let w_n be the time the n-th arrival waits before commencing service, then

$$(1) \qquad w_{n+1} = \begin{cases} w_n + u_n & w_n + u_n > 0, \quad w_n > 0 \\ c_n & c_n > 0, \qquad\quad\; w_n = 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $u_n = s_n - t_{n+1}$ and $c_n = r_n - t_{n+1}$; $\{u_n\}$ and $\{c_n\}$ are independently and identically distributed random variables with the d.f.'s

$$(2) \qquad U(x) = \int_{0-}^\infty B(x+y) dA(y), \quad C(x) = \int_{0-}^\infty D(x+y) dA(y),$$

https://doi.org/10.1017/S144678870002749X Published online by Cambridge University Press

with expectations $E(u_n) = b - a$, $E(c_n) = d - a$ respectively, $E(|u_n|) < \infty$, $E(|c_n|) < \infty$ and c.f.'s

$$\int_{-\infty}^{\infty} e^{i\theta x} dU(x) = \alpha(-\theta)\psi(\theta), \quad \int_{-\infty}^{\infty} e^{i\theta x} dC(x) = \alpha(-\theta)\zeta(\theta).$$

This queueing system may arise in several ways, such as a machine shutting down when no items remain to be serviced, and the service mechanism is different for the first item of a group. It is useful in obtaining some results in priority queueing, and in the traffic model considered by Yeo and Weesakul [7] (see also § 5).

Finch [2] has considered a process which differs from ours only when the n-th arrival joins an empty queue; the customer then waits for a time v_n before commencing a service of length s_n, instead of receiving an immediate service of length r_n. If we put $D(x) = \int_0^x B(x-y)dV(y)$, where $V(x)$ is the d.f. of the $\{v_n\}$, then we can compare results obtained by the two processes [see § 6].

For the $GI/G/1$ system it is known [5] that a unique stationary waiting time distribution exists if and only if $E(u) = b - a < 0$. Finch [2] has shown for his process that this is the stationarity condition, and it is also a necessary and sufficient condition for the existence of a stationary distribution for our process. The proof extends Lindley's [5] argument in a similar manner to that in [2], and will not be given here. The condition is independent of the service time distribution for a customer joining a non-empty queue; we may intuitively expect this, for regardless of the non-zero finite size of the service time for a customer finding the queue empty, the waiting time must eventually reduce to zero again with probability unity if $E(u) < 0$, and may build up indefinitely if $E(u) \geqq 0$.

2. The Waiting time

Let $W_n(x) = Pr\{w_n \leqq x | w_0 = w_0(y)(0 \leqq y < \infty)\}$ be the waiting time d.f. for the n-th arrival; $W_n(x) = 0$ for $x < 0$ by definition, and when $x \geqq 0$ we have for $E(u) < 0$, which we assume to be true throughout unless otherwise stated, that $W(x) = \lim_{n\to\infty} W_n(x)$ exists as a d.f. independently of the initial distribution $W_0(y)$. We shall now prove the following theorem by two methods; the first is similar to Theorem 2 of [2], while the second is an extension of Takács' work [6].

THEOREM 1. If $A(x) = 1 - e^{-\lambda x} (x \geqq 0)$, then the c.f. $\varphi(\theta)$ for the stationary waiting time distribution is given by

$$\varphi(\theta) = \frac{W(0)\{i\theta - \lambda\psi(\theta) + \lambda\zeta(\theta)\}}{i\theta + \lambda - \lambda\psi(\theta)}, \tag{3}$$

where the probability $W(0)$ that a customer arrives to find the queue empty is

$$W(0) = (1-\lambda b)(1-\lambda b+\lambda d)^{-1}. \tag{4}$$

Proof (a). From equation (1) we can write

$$Pr\{w_{n+1} \leqq x\} = Pr\{w_n > 0, w_n + u_n \leqq x\} + Pr\{w_n = 0, c_n \leqq x\} \quad x \geqq 0,$$

from which we immediately obtain

$$W_{n+1}(x) = \int_{-\infty}^{x} W_n(x-y)dU(y) - W_n(0)\{U(x)-C(x)\} \qquad x \geqq 0. \tag{5}$$

As $W(x)$ exists as a d.f. we have

$$W(x) = \int_{-\infty}^{x} W(x-y)dU(y) - W(0)\{U(x)-C(x)\} \qquad x \geqq 0. \tag{6}$$

Following [5] and [2] we define a function $W^*(x)$ by

$$W^*(x) = \int_{-\infty}^{x} W(x-y)dU(y) \qquad -\infty < x < \infty. \tag{7}$$

When $x < 0$ we have

$$\begin{aligned} W^*(x) &= \int_{y=-\infty}^{x} W(x-y)\int_{z=0}^{\infty} dB(y+z)\lambda e^{-\lambda z}dz \\ &= \int_{y=0}^{\infty}\int_{z=0}^{\infty} W(y)\lambda e^{-\lambda(z+y-x)}dy\,dB(z) \\ &= Ce^{\lambda x}, \end{aligned} \tag{8}$$

where $C = W^*(0) = \int_{-\infty}^{0} W(-y)dU(y)$.

Now for $x \geqq 0$ we have from (6) and (7) that

$$W^*(x) = W(x) + W(0)\{U(x)-C(x)\} \qquad x \geqq 0, \tag{9}$$

so that

$$W^*(0) = W(0)\{1+U(0)-C(0)\} = W(0)\{1+\psi(i\lambda)-\zeta(i\lambda)\}.$$

We take Fourier transforms in (6) and (7), and using (8), obtain

$$\begin{aligned} \int_{-\infty}^{\infty} e^{i\theta x}dW^*(x) &= \int_{x=-\infty}^{\infty} e^{i\theta x}\int_{y=-\infty}^{x} dW(x-y)dU(y) \\ &= \lambda(\lambda+i\theta)^{-1}\psi(\theta)\varphi(\theta) \\ &= \int_{-\infty}^{0} e^{i\theta x}dW^*(x) + \int_{0-}^{\infty} e^{i\theta x}dW(x) - W(0) \\ &\quad + W(0)\int_{0-}^{\infty} e^{i\theta x}dU(x) - W(0)\int_{0-}^{\infty} e^{i\theta x}dC(x) \\ &= \lambda(\lambda+i\theta)^{-1}W^*(0) + \varphi(\theta) - W(0) \\ &\quad + \lambda(\lambda+i\theta)^{-1}W(0)\{\psi(\theta)-\zeta(\theta)-\psi(i\lambda)+\zeta(i\lambda)\}; \end{aligned}$$

https://doi.org/10.1017/S144678870002749X Published online by Cambridge University Press

equation (3) follows from these relations and (4) is obtained from (3) by the limiting process $\theta \to 0$.

PROOF (b). Let us for the moment consider the time-dependent problem. For a Poisson process with parameter λ, we have for small time $\delta t > 0$ that during a period of length δt the probability of (i) no arrival is $1 - \lambda\delta t + o(\delta t)$, (ii) one arrival is $\lambda\delta t + o(\delta t)$, and (iii) more than one arrival is $o(\delta t)$. For the waiting time d.f. $W(u, x, t) \equiv W(x, t)$ at time t, given that there is a waiting time $u \geqq 0$ at time zero, we have for $x \geqq 0$,

$$W(x, t+\delta t) = (1-\lambda\delta t)W(x+\delta t, t) + \lambda\delta t \int_0^x W(x-y, t)dB(y) + \lambda\delta t W(0, t)\{D(x) - B(x)\} + o(\delta t). \tag{10}$$

Let $W'(x, t) = (\partial/\partial x)W(x, t)$ be a right-hand derivative of $W(x, t)$; this exists for all $x \geqq 0$. From (10) by letting $\delta t \to 0$ we obtain

$$\frac{\partial}{\partial t} W(x, t) - \frac{\partial}{\partial x} W(x, t) = -\lambda W(x, t) + \lambda \int_0^x W(x-y, t)dB(y) + \lambda W(0, t)\{D(x) - B(x)\} \quad x \geqq 0. \tag{11}$$

As $t \to \infty$ we have, for $\lambda b < 1$, that $(\partial/\partial t)W(x, t) \to 0$ and

$$W'(x) = \lambda W(x) - \lambda \int_0^x W(x-y)dB(y) + \lambda W(0)\{B(x) - D(x)\} \quad x \geqq 0. \tag{12}$$

Taking Fourier transforms in (12) and noting that $\int_0^\infty e^{i\theta x} W'(x)dx = -W(0) + \varphi(\theta)$ we readily obtain (3) and hence (4).

The moments of the waiting time distribution may be found from (3); the mean waiting time is

$$E(w) = -i\varphi'(0) = \frac{\lambda E(r^2) - \lambda^2 bE(r^2) + \lambda^2 dE(s^2)}{2(1-\lambda b)(1-\lambda b+\lambda d)}.$$

We see that proof (b) is simpler than proof (a) as it is not necessary to introduce an artificial function such as $W^*(x)$. We may extend the second method to the time-dependent case, while we may use the first for the case of an Erlangian inter-arrival distribution.

Taking Fourier transforms in (11) yields the further partial differential equation

$$\frac{\partial}{\partial t}\varphi(\theta, t) = (-i\theta - \lambda + \lambda\psi(\theta))\varphi(\theta, t) + i\theta W(0, t) - \lambda\psi(\theta) + \lambda\zeta(\theta),$$

which has the solution

$$\varphi(\theta, t) = e^{(-\lambda - i\theta + \lambda\psi(\theta))t + i\theta u} + \int_0^t \{i\theta W(0, t-\tau) - \lambda\psi(\theta) + \lambda\zeta(\theta)\} - e^{(\lambda + i\theta - \lambda\psi(\theta))\tau} d\tau.$$

This may be inverted to give (see Gani and Prabhu [3])

$$(13)\quad \begin{aligned} W(x, t) &= K(x+t-u, t) - \int_{\tau=0}^{t} W(0, t-\tau)dK(x+\tau, \tau) \\ &\quad - \lambda \int_{\tau=0}^{t}\int_{y=0}^{x} \{dB(x-y) - dD(x-y)\}dK(y+\tau, \tau), \end{aligned}$$

where $K(x, t) = \sum_{n=0}^{\infty} e^{-\lambda t}(\lambda t)^n (n!)^{-1} B_n(x)$, $B_n(x)$ being the n-fold convolution of $B(x)$; I have been unable to obtain $W(0, t)$ explicitly, although an expression for its c.f. may be found.

For an Erlangian inter-arrival time distribution we make use of Lemma 1 of [1] concerning the roots of $(\lambda + i\theta)^k - \lambda^k \psi(\theta)$ and obtain

THEOREM 2. When $A(x) = 1 - \sum_{r=0}^{k-1} \lambda^r x^r e^{-\lambda x}(r!)^{-1}$ $(k = 1, 2, \cdots)$ we have, for $\lambda b < k$,

$$(14)\quad \begin{aligned} \varphi(\theta) = W(0) \Bigg[\sum_{n=0}^{k-2} \Big\{ i\theta L_n \Big(\prod_{m \neq n} (i\theta - i\theta_m) \Big) - ((\lambda + i\theta)^{k-1} - (\lambda + i\theta)^n)\beta_n \Big\} \\ + i\theta(\lambda + i\theta)^{k-1} - \lambda^k \psi(\theta) + \lambda^k \zeta(\theta) \Bigg] [(\lambda + i\theta)^k - \lambda^k \psi(\theta)]^{-1}, \end{aligned}$$

where

$$(15)\quad \begin{aligned} W(0) = \lambda^{k-1}(k - \lambda b) \Bigg[\sum_{n=0}^{k-2} L_n \prod_{m \neq n} (-i\theta_m) + \lambda^{k-1} - (k-1)\gamma\lambda^{k-2} - \lambda^k(b-d) \\ + \sum_{n=1}^{k-1} n\beta_n \lambda^{n-1} \Bigg]^{-1}, \end{aligned}$$

and

$$\chi_n = -\lambda^k(\psi(\theta_n) - \zeta(\theta_n)) - \sum_{m=0}^{k-2} \{(\lambda + i\theta_n)^{k-1} - (\lambda + i\theta_n)^m\}\beta_m + i\theta_n(\lambda + i\theta_n)^{k-1}$$

$$\beta_n = \frac{(-1)^n \lambda^k}{n!} \left(\frac{d}{d\lambda}\right)^n \{\psi(i\lambda) - \zeta(i\lambda)\}$$

$$\gamma = \frac{(-1)^{k-1}\lambda^{k+1}}{k-1!} \left(\frac{d}{d\lambda}\right)^{k-1} \left\{\frac{\psi(i\lambda) - \zeta(i\lambda)}{\lambda}\right\} = \sum_{n=0}^{k-1} \beta_n$$

$$L_n = (i\theta)^{-1} \chi_n \prod_{m \neq n} (i\theta_n - i\theta_m)^{-1},$$

$\theta_n (0 \leqq n \leqq k-1)$ being the k roots of $(\lambda + i\theta)^k - \lambda^k \psi(\theta)$ in the upper half plane $\mathscr{I}(\theta) \geqq 0$.

The proof generalises that of Theorem 1 and parallels that in [1] much as our proof (a) of Theorem 1 parallels that of Theorem 2 of [2]. The algebra in our case is more unwieldy, mainly because we need to integrate over $[0, \infty)$ a function, i.e. $(U(x) - C(x))e^{i\theta x}$, which extends over the whole range $(-\infty, \infty)$, whereas in [1] there is a function which extends over $[0, \infty)$.

3. The Queue size

Let $Q_n(m)(n = 0, 1, 2, \cdots)$ be the probability that the m-th arrival finds n customers in the queue, and $R_n(m)(n = 0, 1, 2, \cdots)$ be the probability that when he departs there are n customers in the queue. As in [2] we may show that the limiting distributions $Q_n = \lim_{m\to\infty} Q_n(m)$ and $R_n = \lim_{m\to\infty} R_n(m)$ both exist and are equal. We shall now prove

THEOREM 3. *If* $A(x) = 1 - \sum_{r=0}^{k-1} \lambda^r x^r (r!)^{-1} e^{-\lambda x} (k = 1, 2, \cdots)$ *then the p.g.f.* $R(z) = \sum_{n=0}^{\infty} R_n z^n (|z| \leqq 1)$ *is given, for* $\lambda b < k$, *by*

$$R(z) = R_0\{k(z) - zk^*(z)\}\{k(z) - z\}^{-1}, \tag{16}$$

$$R_0 = \{1 - k'(1)\}\{1 - k'(1) + k^{*\prime}(1)\}^{-1}, \tag{17}$$

where

$$k(z) = \sum_{n=0}^{\infty} k_n z^n, \quad k_n = \int_0^{\infty} e^{-\lambda x} \sum_{r=nk}^{(n+1)k-1} \frac{(\lambda x)^r}{r!} dB(x)$$

$$k^*(z) = \sum_{n=0}^{\infty} k_n^* z^n, \quad k_n^* = \int_0^{\infty} e^{-\lambda x} \sum_{r=nk}^{(n+1)k-1} \frac{(\lambda x)^r}{r!} dD(x).$$

PROOF. We readily see that the limiting distribution satisfies the equations

$$R_n = R_{n+1}k_0 + R_n k_1 + \cdots + R_1 k_n + R_0 k_n^*, \quad n = 0, 1, 2, \cdots \tag{18}$$

When we multiply (18) by z^n and sum we readily obtain (16), and (17) by considering the limit $\lim_{z\to 1-0} R(z)$.

When $k = 1$ the results are considerably simplified with $k(z) = \psi(i\lambda(1-z))$, $k^*(z) = \zeta(i\lambda(1 - z))$, $k'(1) = \lambda b$, $k^{*\prime}(1) = \lambda d$, and R_0 is given by (4).

The average number of customers in the stationary queue, obtained by differentiation from (14), is

$$R'(1) = \frac{R_0\{2k^{*\prime}(1) + k^{*\prime\prime}(1)\}}{2(1 - k'(1))}.$$

4. The busy period

The distribution of the busy period, i.e. the time from the arrival of a customer at an empty queue until the next time the server is free, may be obtained by a slight extension of the work by Takács [6] for the $M/G/1$ system. If there is a waiting time of $x > 0$ in the queue at time zero then the d.f. $F(x, t)$ to the time the queue empties for the first time has the c.f. $\Phi(x, \theta)$ given by Kendall [4] as

$$\Phi(x, \theta) = e^{-x\eta(\theta)},$$

where $\eta(\theta)$ is the unique solution with $\eta(0) = 0$ of the functional equation

$$\eta(\theta) = \lambda - i\theta - \lambda\psi(i\eta(\theta)). \tag{19}$$

As a customer joining an empty queue has d.f. $D(x)$ we have that the c.f. $\gamma(\theta)$ of the busy period distribution is

$$\begin{aligned}\gamma(\theta) &= \int_0^\infty \Phi(x, \theta)dD(x) \\ &= \zeta(i\eta(\theta)).\end{aligned} \tag{20}$$

The moments of this distribution may readily be found by differentiation e.g.

$$-i\gamma'(0) = d(1-\lambda b)^{-1}$$

$$-\gamma''(0) = \frac{E(r^2) - \lambda b E(r^2) + \lambda d E(s^2)}{(1-\lambda b)^3}.$$

5. Cases of server absenteeism

We now consider another generalisation of the $M/G/1$ queueing system which for stationary distributions is a special case of our previous problem. Whenever a departing customer leaves the queue empty the server departs from the counter for a time which has d.f. $G(x)$, expectation $g < \infty$ and c.f. $\xi(\theta)$. When he returns to find at least one customer waiting he commences serving immediately; however, when he returns to find the queue still empty we consider two cases: (i) he departs again for a time with d.f. $G(x)$, and continues to do so until he returns to find a waiting customer, whom he serves at once, and (ii) he remains until a customer arrives, when he begins serving immediately. From the time the server is first occupied customers have a service time d.f. $B(x)$ with expectation $b < \infty$ and c.f. $\psi(\theta)$. If we were to consider service beginning only when the server actually commences to serve a customer then model (i) can be solved by the method of Finch [2]. However, we take the case where a customer commences service immediately he reaches the head of the queue. This is similar to a problem which has arisen in road traffic theory (see Yeo and Weesakul [7]), where we wish to find the delay to vehicles on a minor road when they arrive at an intersection at which they must yield right of way to vehicles on a major road. The vehicles on the minor road are assumed to arrive in a Poisson process, while the distribution of major road vehicles may be more general; the length of time a vehicle waits from reaching the head of a queue to crossing depends on whether there were any other vehicles waiting when it arrived.

We require to find c.f. $\chi(\theta)$ of the d.f. $L(x)$ of the time from the arrival of a customer at an empty queue to the return of the server. For obtaining stationary distributions we can integrate over the possible times of arrival of a customer after the departure of the last customer of the previous busy period. Considering only the first server absentee period we obtain for the c.f. $\chi_1(\theta)$ of the distribution of the time to the return of the server as

$$\chi_1(\theta) = \int_{x=0}^{\infty}\int_{y=0}^{\infty} \lambda e^{-\lambda y} e^{i\theta(x-y)} dG(x)dy \tag{21}$$
$$= \frac{\lambda}{\lambda + i\theta}\{\xi(\theta) - \xi(-i\lambda)\},$$

which is an improper distribution with $\chi_1(0) = 1 - \xi(-i\lambda) = 1 - g_0$ being the probability that no customers arrive while the server is absent. For the case where the server departs whenever there are no customers waiting we have

$$\chi(\theta) = \chi_1(\theta)\{1 + g_0 + g_0^2 + \cdots\} \tag{22}$$
$$= \frac{\lambda\{\xi(\theta) - g_0\}}{(\lambda + i\theta)(1 - g_0)},$$

while for case (ii) where the server departs only once at the end of a busy period we have

$$\chi(\theta) = \chi_1(\theta) + g_0 = \frac{\lambda\xi(\theta) + i\theta g_0}{\lambda + i\theta}. \tag{23}$$

Once the server actually begins serving the time to completion of service is the same as for customers joining a non-empty queue (although this may be generalised as is necessary for some traffic models) so that

$$\mathfrak{Z}(\theta) = \psi(\theta)\chi(\theta) \tag{24}$$

may be taken as the service time c.f. for customers arriving at an empty queue. The stationary waiting time and queue size distributions may now be obtained from the results of the previous sections.

6. Comparison with Finch [2]

As a special case of our process we put $D(x) = \int_0^x B(x-y)dV(y)$. The distribution of $R_n(Q_n)$ is identical to that in [2], but the waiting time distribution differs, as Finch considers $\{v_n\}$ as part of the waiting time while we have it as part of the service time. However, the delay caused, i.e. waiting time plus service time, is the same in both cases.

As an example suppose $A(x) = 1 - e^{-\lambda x}$, $B(x) = 1 - e^{-\mu x}$, $V(x) = 1 - e^{-\nu x}$ so that $D(x) = 1 - (\mu - \nu)^{-1}(\mu e^{-\mu x} - \nu e^{-\nu x})$. We obtain by inverting (3) that

$$W(x) = 1 - \frac{\lambda\nu}{\mu(\lambda + \nu - \mu)} e^{-(\mu-\lambda)x} + \frac{\lambda(\mu - \lambda)}{(\nu + \lambda)(\nu + \lambda - \mu)} e^{-\nu x} \quad x \geqq 0,$$

with $W(0) = \nu\mu^{-1}(\mu - \lambda)(\nu + \lambda)^{-1}$, while Finch obtains

$$W(x) = 1 - \frac{\lambda\nu}{\mu(\lambda + \nu - \mu)} e^{-(\mu-\lambda)x} - \frac{(\mu - \lambda)(\nu - \mu)}{\mu(\lambda + \nu - \mu)} e^{-\nu x} \quad x \geqq 0.$$

References

[1] Ewens, W. J. and Finch, P. D., "A generalised single server queue with Erlang inputs" (1962) 49, 242—245.

[2] Finch, P. D., "A probability limit theorem with application to a generalisation of queueing theory", Acta. Math. Acad. Sci. Hung., (1959) 10, 317—325.

[3] Gani, J. and Prabhu, N. U., "Remarks on the dam with Poisson type inputs", Aust. J. Appl. Sci., (1959) 10, 113—122.

[4] Kendall, D. G., "Some problems in the theory of dams", J. Roy. Statist. Soc., (1957) B19, 207—212.

[5] Lindley, D. V., "The theory of queues with a single server", Proc. Cambridge Phil. Soc., (1952) 48, 277—289.

[6] Takács, L., "Investigation of waiting time problems by reduction to Markov processes", Acta. Math. Acad. Sci. Hung., (1955) 6, 101—129.

[7] Yeo, G. F., and Weesakul, B., "Distribution of delay to traffic at an intersection, (1962) (submitted for publication).

The Australian National University
Canberra.