MANAGEMENT SCIENCE
Vol. 30, No. 1, January 1984
Printed in U.S.A.

QUEUES IN WHICH CUSTOMERS RECEIVE SIMULTANEOUS SERVICE FROM A RANDOM NUMBER OF SERVERS: A SYSTEM POINT APPROACH*

PERCY H. BRILL† AND LINDA GREEN‡

We examine a multi-server queueing system with Poisson arrivals in which customers require simultaneous service from a random number of servers. Servers assigned to the same customer begin and end service concurrently. Service times are, in general, assumed to be exponentially distributed. A system point approach is presented as a framework for obtaining the waiting time distribution for each customer type. Explicit solutions are derived for the two-server system.
(MULTIPLE SERVER QUEUES; SYSTEM POINT METHOD; LEVEL CROSSINGS; DISTINGUISHABLE CUSTOMERS)

0. Introduction

In many queueing systems, different types of customers require concurrent service from different numbers of servers. Service doesn't start until the required number of servers is available and servers associated with the same customer begin and end service simultaneously. Examples are prevalent, particularly in computer and communications systems. In computer systems, programs which contend for space on storage devices have varying size and residency time requirements (see, e.g., Arthurs and Kaufman 1979 and Omahen 1977). The programs can therefore be viewed as customers which require simultaneous service from a random number of the storage units, which are the servers of the system. An almost identical situation occurs in the area of satellite communications (Nigam 1975). Each satellite has a fixed total number of channels (servers) available for transmissions of various types. Each of these customer types, e.g. television programs, telephone conversations, and data transmissions, requires a different number of channels for a random period of time. A transmission of any type will be delayed if an insufficient number of channels is available at the time its request for service is made.

Though there are many examples of queueing systems of this type, there is relatively little literature on their analysis. This is probably due to the complexity involved as a result of the large state space required to model such systems by traditional means, even under the usual assumptions of Poisson arrivals and exponential service times. Gimpelson (1965) examined a system in which a single wide-band facility is used to carry communication traffic of two types: wide-band and narrow-band, and in which either no queueing is allowed or a finite queueing capacity is provided for wide-band calls only. Using numerical methods, he obtained the blocking probabilities for each traffic type. Wolman (1972) studied a problem in which data traffic is directed to a random number of destinations and cannot be transmitted until the required number of receivers are free. He obtained approximations to the steady-state waiting times by solving a simpler model in which it was assumed that all messages are addressed to exactly m receivers, where m is an integer greater than 1. Arthurs and Kaufman (1979) analyzed the pure loss system version of the problem we examine here. Under the

*Accepted by John P. Lehoczky; received August 26, 1982. This paper has been with the authors 2 months for 2 revisions.
†University of Waterloo.
‡Columbia University.

0025-1909/84/3001/0051$01.25

assumption of Poisson arrivals, they showed that the steady-state distribution of the number of customers of each type in service has a product form and is dependent on the service time distributions only through their means. Related queueing systems in which it is assumed that the service times of servers working on the same customer are independent once service is begun were studied by Green (1978, 1980) and Gillent (1980), and a distinguishable server counterpart with two customer types was analyzed by Green (1981b).

The only previous analysis of the system studied in this paper was done by Kim (1979). Kim modelled the system as a Markov process with state variables defined by n, the number of customers in the system, and the vector $(c_1, c_2, \ldots, c_f)$, where c_i is the number of servers demanded by the ith customer in the system (in order of arrival) and f is the minimum of n and U, the maximum number of customers that can be in service simultaneously. Using a matrix-geometric approach (see e.g. Neuts 1980), he derived an algorithm for computing the steady-state probabilities of the number in system. The major limitation of this approach is that only the overall expected waiting time can be computed. The definition of the state space does not allow for either the mean waiting time by customer type or any distributional results. In addition, the computational effort required renders this approach impractical for all but very small systems (size is a function of K^U where K is the number of servers).

This paper presents a framework for deriving the waiting time distribution for each customer type in the multi-server queueing system in which customers need simultaneous service from a random number of servers. We will assume Poisson arrivals and, with certain exceptions, exponential service times. We use a System Point (SP) approach which is an extension of the System Point Method developed by Brill (1975) and further elucidated and applied in e.g. Brill and Posner (1981a, 1981b) and Brill (1979). The method is illustrated by explicitly solving the two server system with exponential service times.

1. Definitions and System Point Theory

The general system consists of s servers and k customer types ($k \leqslant s$). Type i, $i = 1, \ldots, k$, customers arrive according to a Poisson process at rate λ_i and require an exponentially distributed amount of service time with mean $1/\mu_i$ simultaneously from $c(i)$ servers ($c(i) \leqslant s$). (A general service time distribution can be assumed for Type j customers if $c(j) = s$.) Customers are served in the order of their arrival. (See Green (1981a) on the relative effectiveness of other service order disciplines.)

Define $W^{(i)}(t)$ to be the virtual waiting time (in queue before service) at time t for a Type i customer. We define a random variable $M(t)$, called the *system configuration* at time t, such that the process $\{\langle W^{(i)}(t), M(t)\rangle\}$ is Markovian for $i = 1, \ldots, k$. Let the k customer types be numbered from 1 to k such that $c(i) < c(j)$ for $i < j$. Then $M(t) = (n_1, \ldots, n_k)$ if a Type 1 customer arriving at t would "see" n_i Type i customers, $i = 1, \ldots, k$, in service at his service-starting epoch *exclusive* of customers who arrive at or subsequent to time t. We will denote the set of possible configurations for a given system by $\mathscr{M}$. For example, in the two-server system where Type 1 customers require one server and Type 2 require two, $\mathscr{M} = \{(0,0),(1,0)\}$ and $M(t) = (1,0)$ if the system is not empty at time t and the last arrival was a Type 1, while $M(t) = (0,0)$ otherwise.

The stochastic process $\{\langle W^{(i)}(t), M(t)\rangle\}$ is called the System Point process for customer type i. Let $f_t^{(i)}(w,m)$ denote its density. It is assumed that

$$\lim_{t\to\infty} f_t^{(i)}(w,m) = f^{(i)}(w,m), \qquad i = 1, \ldots, k.$$

When all service times are exponential and $\mu_i = \mu$ for all i, a necessary and sufficient condition for the existence of a limiting distribution can be obtained from matrix-geometric theory as in Kim (1979). (The general equilibrium condition for matrix-geometric queueing systems, which can be found in Neuts (1980), basically requires that the total probability of an upward jump in the number in system be strictly less than that of a downward jump.) In §3 this condition is given explicitly for the two server system.

For each SP process, if the state of the system is $\langle w, m\rangle$ at time t, the state may be pictured as a point with coordinates (t, w) in a coordinate system corresponding to configuration m. This point, denoted by SP(i) for the ith process, traces out the sample function as t increases and is called the System Point. For a system with N configurations (see §5 for a characterization of N) there will be N such coordinate systems, called "pages" in SP theory, for each customer type. The N pages may be thought of as being one behind the other in a "book" with the projection being the "cover." For a full description of the concepts of SP, "pages," and "book," see Brill and Posner (1981a). So if there are k customer types, their virtual waiting time sample paths will be traced out simultaneously in k "books."

Figures 1 and 2 illustrate this for the two-server system in which both customer types require an exponentially distributed amount of service time with mean $1/\mu$ (exp(μ)). The configuration set was described above. Figure 1 depicts the pages and a possible sample function traced out by SP(1) over time and Figure 2 depicts the sample function for SP(2) corresponding to the same set of events. In each figure, if

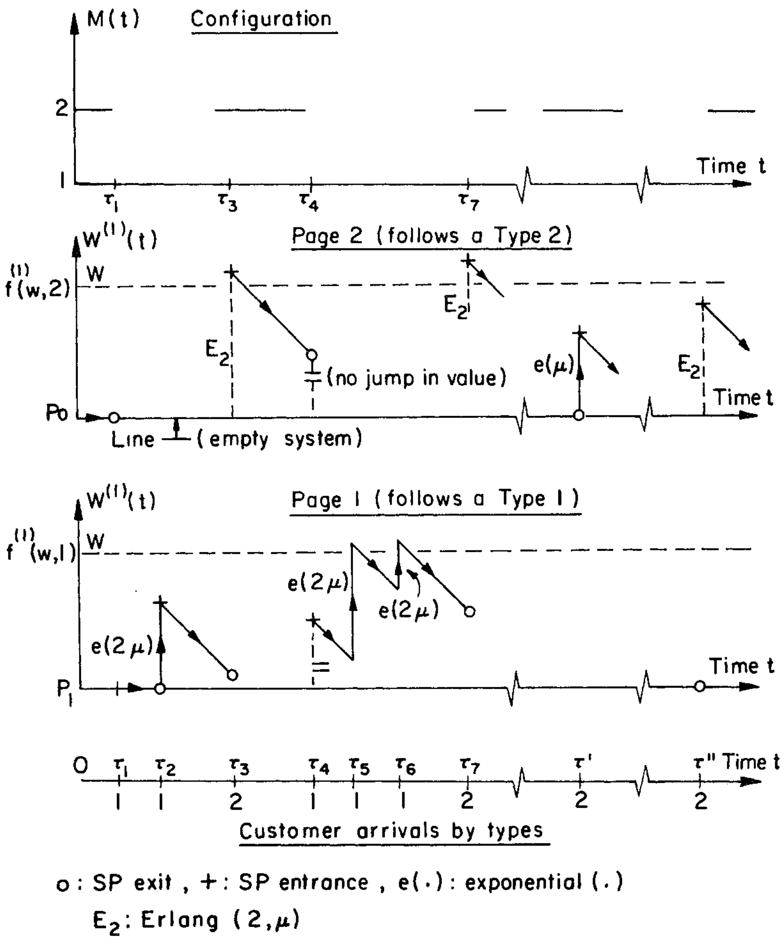

FIGURE 1

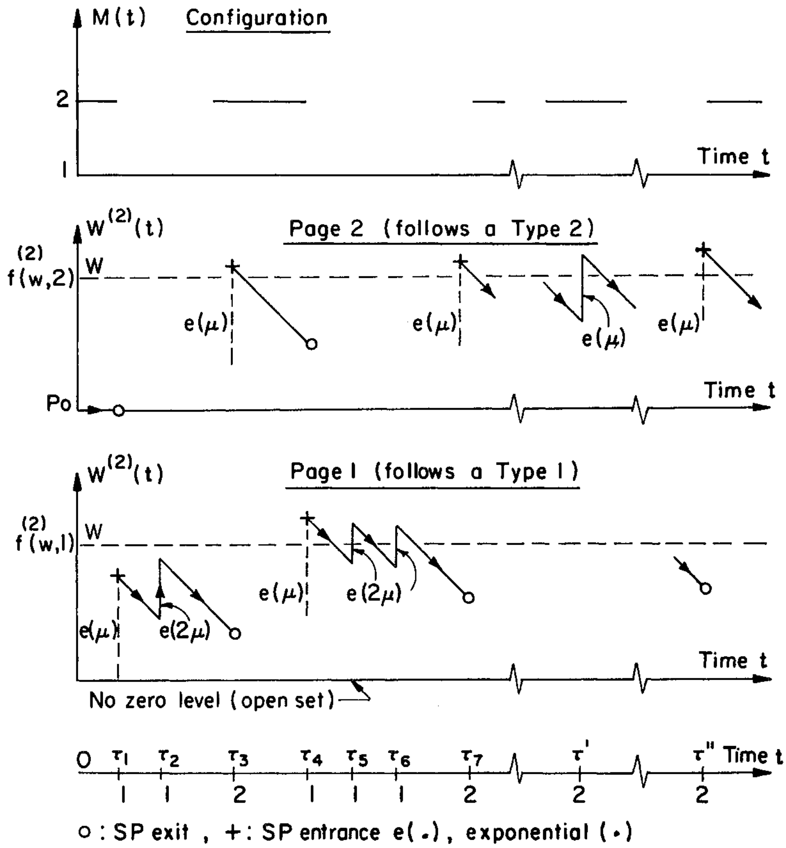

FIGURE 2

the two pages are superimposed, the resulting sample function will be a piecewise continuous graph of the usual virtual waiting time traced out on the "cover". Consider the sample function for a Type 1 customer as shown in Figure 1. For simplicity, the configurations (1,0) and (0,0) are denoted as 1 and 2, respectively. At $t = 0$ the system is empty and so SP(1) is on page 2. The first customer, a Type 1, arrives at τ_1, when the state is $\langle 0,2\rangle$. SP(1) jumps to $(\tau_1^+, 0)$ on page 1 since a Type 1 arrival at τ_1^+ would enter service immediately and see another Type 1 in service. The next arrival at τ_2 is also of Type 1 and occurs before the first customer ends service. So SP(1) stays on page 1 and jumps to $(\tau_2^+, W^{(1)}(\tau_2^+))$ where $W^{(1)}(\tau_2^+)$ is exp(2μ). At τ_3 a Type 2 customer arrives and SP(1) jumps to page 2. Since $M(\tau_3^-) = 1$, a Type 1 arrival at τ_3^- would have entered service while one of the previous Type 1's was still in service. Since the Type 2 arrival at τ_3 cannot enter service with a Type 1, the size of the SP(1) jump is distributed as the sum of a Type 1 and a Type 2 service time, which in this case is Erlang $(2, \mu)$. The pattern becomes clear: If SP(1) is at (t, w), $w \geqslant 0$, on page 1, the arrival of a Type 1 customer causes a jump of size exp(2μ) on the same page, while a Type 2 arrival results in a jump to page 2 of size Erlang $(2, \mu)$. If SP(1) is on page 2, a Type 1 arrival causes a jump to page 1 with no increase in waiting time while a Type 2 arrival results in a jump of exp(μ) on the same page. Similarly, a pattern can be established for SP(2): A jump from page 1 is exp(2μ) and stays on page 1 if due to a Type 1 arrival, and is exp(μ) to page 2 if due to a Type 2. A jump from page 2 is to page 1 and of size exp(μ) if caused by a Type 1 arrival, and is exp(μ) and stays on page 2 if due to a Type 2.

System Point theory relates the joint limiting densities $f^{(i)}(w, m)$, $w > 0$, $m \in \mathscr{M}$ to the long-run average rate at which SP(i) crosses level w on page m in the state space.

The basic theorem given below is true in any system whose virtual wait has sample functions for each configuration as shown in Figures 1 and 2, i.e. piecewise continuous and decreasing at a 45° angle for positive values. The proof depends only on the geometric properties of the sample functions and the existence of the limiting probability density functions. Therefore this proof is essentially identical to that of the corresponding theorem in Brill and Posner (1981a) and is omitted.

Let $\mathscr{D}_t^{(i)}(w,m)$ denote the number of SP(i) downcrossings of level $w > 0$ on page m in $(0,t]$. For each SP, an m-downcrossing of level $w > 0$ occurs at time t_0 if $W(t_0) = w$ and there exists $\epsilon > 0$ such that $W(t) < w$ for $t \in (t_0, t_0 + \epsilon)$ and the configuration is $M(t) = m$ for $t \in (t_0, t_0 + \epsilon)$. Let $\mathscr{I}_t^{(i)}(m)$ be the number of SP(i) impacts of level zero on page m in $(0,t]$. An SP m-impacts with level zero at t_0 if there exists $\epsilon > 0$ such that $W(t) > 0$ for $t \in (t_0 - \epsilon, t_0]$, $W(t_0) = 0$ and $M(t) = m$ for $t \in (t_0 - \epsilon, t_0]$.

THEOREM 1.

$$\lim_{t\to\infty} E\left[\mathscr{D}_t^{(i)}(w,m)\right]/t = f^{(i)}(w,m), \qquad w > 0, \quad i = 1, \ldots, k, \quad m \in \mathscr{M},$$

$$\lim_{t\to\infty} E\left[\mathscr{I}_t^{(i)}(m)\right]/t = f^{(i)}(0^+,m), \qquad i = 1, \ldots, k, \quad m \in \mathscr{M}, \quad \text{where}$$

$$f^{(i)}(0^+,m) = \lim_{w\downarrow 0} f^{(i)}(w,m).$$

In the next section we present expressions ((2.3)–(2.6)) which equate entrance and exit rates of sets in the state space by the system point. It is important to note that these equations are valid because this system has the characteristics of Poisson arrivals, no multiple events, and the Markov property of $\{\langle W^{(i)}(t), M(t)\rangle, t \geqslant 0\}$. It is precisely these properties that are invoked in the proofs of the corresponding results in Brill and Posner (1981a).

2. Model Equations for the Two-Server System

We assume a two-server system with two customer types: Type 1 customers arrive according to a Poisson process at rate λ_1, and require an exponentially distributed amount of service time with mean $1/\mu$ from one server; Type 2 customers arrive according to a Poisson process at rate λ_2 and require service from both servers simultaneously with service time distribution B. We define $\lambda = \lambda_1 + \lambda_2$. Let p_0 be the equilibrium probability that the system is empty. This is a nonwaiting state for both customer types. For SP(1) define p_1 as the probability that exactly one Type 1 customer is in service and there is no queue. This is a nonwaiting state for a Type 1 customer. As noted before, there are two configurations for this system: $M(t) = 1$ if the system is not empty at time t and the last customer to arrive was a Type 1; $M(t) = 2$ otherwise. This is equivalent to the definition of configuration given for the general model in the last section.

For each customer type, we will derive a set of balance equations which equate the rate of entrance of the SP into the set $\{\langle u,m\rangle \mid 0 < u < w\}$, $w > 0$, $m = 1,2$, with its rate of exit from this set. These equations plus the balance equations for the nonwaiting states and the normalizing condition that all probabilities sum to one will be solved for the case where B is exp(μ) in the next section to yield the waiting time densities for each customer type. Note, however, that it is sufficient to solve for the functions relating to SP(1) only, namely $f^{(1)}(w,1)$, $f^{(1)}(w,2)$ and the probabilities p_0 and p_1. The SP(2) results can be computed from the following relationships:

$$\begin{aligned} W^{(2)}(t) &= W^{(1)}(t) + S, \qquad t \geqslant 0, \quad M(t) = 1, \\ W^{(2)}(t) &= W^{(1)}(t), \qquad t \geqslant 0, \quad M(t) = 2, \end{aligned} \tag{2.1}$$

where S is distributed as the service time of a Type 1 customer, i.e., $\exp(\mu)$. The first relation holds because an arriving Type 2 who follows a Type 1 customer into service must wait in queue an amount of time equal to that of a Type 1 who would have arrived at the same instant as the Type 2, plus a residual Type 1 service time. The second holds because the virtual waits of Type 1 and Type 2 customers who follow a Type 2 into service are identical for $t \geqslant 0$. Relationship (2.1) leads to the following computation for the equilibrium pdf's in SP(2), once the pdf's in SP(1) are known:

$$\begin{aligned} f^{(2)}(w,1) &= \int_{x=0}^{w} f^{(1)}(x,1)\mu e^{-\mu(w-x)}\,dx + p_1\mu e^{-\mu w}, \qquad && w>0,\\ f^{(2)}(w,2) &= f^{(1)}(w,2), && w>0. \end{aligned} \tag{2.2}$$

The following equations are based on the principle of stationary set balance. Rigorous derivations of (2.3) and (2.4) can be obtained as in the proofs of Theorems 5 and 6 of Brill and Posner (1981a). Intuitively, these results are obtained by balancing the entrance and exit rates of SP(1) into the set $(0,w)$ on each page. For page 1 we get (see Figure 1):

$$\begin{aligned} &f^{(1)}(w,1) + \lambda_1(1-e^{-2\mu w})p_1 + \lambda_1\int_0^w f^{(1)}(\alpha,2)\,d\alpha\\ &\quad = \lambda_2\int_0^w f^{(1)}(\alpha,1)\,d\alpha + \lambda_1\int_0^w e^{-2\mu(w-\alpha)}f^{(1)}(\alpha,1)\,d\alpha + f^{(1)}(0^+,1). \end{aligned} \tag{2.3}$$

This equation is explained as follows. The left-hand side (LHS) is the total rate at which SP(1) enters $(0,w)$ on page 1. This can occur in three ways:

(i) it can be at level $\hat{w} > w$ on page 1 and downcross level w;

(ii) it can be at level $w=0$ on page 1 (one Type 1 in service with remaining service time R, say, and no queue), a new Type 1 customer arrives with service time S, say, and $\min(R,S) < w$;

(iii) it can be at level $\hat{w}$ in the set $(0,w)$ on page 2 and a Type 1 customer arrives.

From Theorem 1, the rate of the first of these occurrences is given by the first term of the LHS of (2.3). The other two terms of the LHS are clearly the rates of the second and third occurrences, respectively. Similarly, there are three ways in which SP(1) can exit the set $(0,w)$ on page 1 and the rates of these events are given by the right-hand side (RHS) of (2.3). The first term is the rate at which SP(1) will exit due to Type 2 arrivals. The second term is derived by noting that if the Type 1 virtual wait is $\alpha < w$, then a Type 1 arrival will cause the virtual wait to exceed w with probability $e^{-2u(w-\alpha)}$. Finally, the third term is, by Theorem 1, the rate at which SP(1) impacts with level $w=0$.

Note that Type 2 customers use both servers simultaneously as if they were a single server. Hence the virtual wait process on page 2, as in $M/G/1$, remains Markovian when Type 2 customers have a general service time distribution. Therefore the SP(1) set balance equation for page 2 is

$$\begin{aligned} &f^{(1)}(w,2) + \lambda_2 B(w)p_0 + \lambda_2 p_1\int_0^w \mu e^{-\mu\alpha}B(w-\alpha)\,d\alpha\\ &\quad + \lambda_2\int_0^w\int_0^{w-\alpha}\mu e^{-\mu((w-\alpha)-t)}B(t)\,dt\, f^{(1)}(\alpha,1)\,d\alpha\\ &\quad = \lambda_1\int_0^w f^{(1)}(\alpha,2)\,d\alpha + \lambda_2\int_0^w (1-B(w-\alpha))f^{(1)}(\alpha,2)\,d\alpha + f^{(1)}(0^+,2), \end{aligned} \tag{2.4}$$

where $B(x) = Pr$ (Type 2 service time $\leqslant x$). Again, the LHS represents the total entrance rate of SP(1) into $(0,w)$ on page 2 and the RHS, the total exit rate. On this page, there are four ways in which the SP can enter $(0,w)$ with rates given by the four terms of the LHS of (2.4). Respectively, these terms are (i) the downcrossing rate into $(0, w)$ on page 2; (ii) the rate at which Type 2 customers with service time less than w arrive to an empty system; (iii) the rate at which Type 2 customers arrive when a Type 1 is in service, there is no queue and the sum of the Type 1's remaining service time and the Type 2's service time is less than w; (iv) the analogous case when a Type 2 arrives when the system state is $\langle \alpha, 1 \rangle$, $0 < \alpha < w$. The exit rates are given by the three terms on the RHS and are due to: a Type 1 arrival, a Type 2 arrival that increments the virtual wait above w, and the rate at which SP(1) impacts with $w = 0$. For the nonwaiting states corresponding to exactly one Type 1 in service and no queue, and the empty system, respectively, we can write

$$(\lambda + \mu) p_1 = \lambda_1 p_0 + f^{(1)}(0^+, 1), \tag{2.5}$$

$$\lambda p_0 = \mu p_1 + f^{(1)}(0^+, 2). \tag{2.6}$$

The pdf of an arbitrary Type 1 customer is given by the total density for SP(1),

$$g^{(1)}(w) = f^{(1)}(w, 1) + f^{(1)}(w, 2), \qquad w > 0. \tag{2.7}$$

The normalizing condition for SP(1) is

$$\int_0^\infty g^{(1)}(w)\, dw + p_0 + p_1 = 1. \tag{2.8}$$

Using the same reasoning, we obtain the stationary set balance equation for SP(2) on page 1 as

$$\begin{aligned} f^{(2)}(w, 1) &+ \lambda_1 (1 - e^{-\mu w}) p_0 + \lambda_1 \int_0^w (1 - e^{-\mu(w-\alpha)}) f^{(2)}(\alpha, 2)\, d\alpha \\ &= \lambda_2 \int_0^w f^{(2)}(\alpha, 1)\, d\alpha + \lambda_1 \int_0^w e^{-2\mu(w-\alpha)} f^{(2)}(\alpha, 1)\, d\alpha + f^{(2)}(0^+, 1), \end{aligned} \tag{2.9}$$

and on page 2 as

$$\begin{aligned} f^{(2)}(w, 2) &+ \lambda_2 B(w) p_0 + \lambda_2 \int_0^w B(w - \alpha) f^{(2)}(\alpha, 1)\, d\alpha \\ &= \lambda_1 \int_0^w f^{(2)}(\alpha, 2)\, d\alpha + \lambda_2 \int_0^w [1 - B(w - \alpha)] f^{(2)}(\alpha, 2)\, d\alpha + f^{(2)}(0^+, 2). \end{aligned} \tag{2.10}$$

The balance equation for the empty state is

$$\lambda p_0 = f^{(2)}(0^+, 1) + f^{(2)}(0^+, 2). \tag{2.11}$$

The pdf of an arbitrary Type 2 customer is given by

$$g^{(2)}(w) = f^{(2)}(w, 1) + f^{(2)}(w, 2) \tag{2.12}$$

and the normalizing condition for SP(2) is

$$\int_0^\infty g^{(2)}(w)\, dw + p_0 = 1. \tag{2.13}$$

3. The Probability Density of the Waiting Time

In this section the SP(1) model equations (2.3)–(2.6) are solved for the functions $f^{(1)}(\cdot, 1)$, $f^{(1)}(\cdot, 2)$, $g^{(1)}(\cdot)$, and constants p_1 and p_0 for the case $B(t) = 1 - e^{-\mu t}$. From matrix-geometric theory (Kim 1979 and Neuts 1980), an equilibrium distribution will exist if $\lambda < 2\mu/(2 - p^2)$ where $p = \lambda_1/\lambda$. The pdf's in SP(2), i.e., $f^{(2)}(\cdot, 1)$, $f^{(2)}(\cdot, 2)$ and $g^{(2)}(\cdot)$, are then calculated by using formula (2.2).

3.1. SP(1) pdf's

The system of integral equations (2.3) and (2.4) was transformed into a system of ordinary differential equations (ODE's) for the functions $f^{(1)}(w,1)$ and $f^{(1)}(w,2)$, $w > 0$. The resulting system is equivalent to the same fourth order homogeneous ODE both for $f^{(1)}(w,1)$ and $f^{(1)}(w,2)$, and is given by (3.1) which was obtained using standard methods. (For details see Brill and Green 1982.) The densities will be those solutions of (3.1) which are bounded, nonnegative, real valued, tend to zero as w tends to infinity, and satisfy certain initial conditions, and the normalizing condition (2.8).

$$\langle \gamma_4 D^4 + \gamma_3 D^3 + \gamma_2 D^2 + \gamma_1 D + \gamma_0 \rangle f^{(1)}(w,j) = 0, \qquad j = 1,2, \tag{3.1}$$

where
$$\begin{aligned}
\gamma_4 &= 1,\\
\gamma_3 &= 2(2\mu - \lambda),\\
\gamma_2 &= \xi + 4\mu^2 - 6\mu\lambda_2 - 4\mu\lambda_1 + \lambda^2,\\
\gamma_1 &= (2\mu - \lambda)\xi - \mu^2\lambda_1 - 4\mu^2\lambda_2 + 2\mu\lambda_2\lambda,\\
\gamma_0 &= -2\mu\lambda_2\xi - 2\mu^3\lambda_1 + \mu^2\lambda_1^2,\\
\lambda &= \lambda_1 + \lambda_2, \text{ and}\\
\xi &= \mu^2 - 2\mu\lambda + \mu\lambda_2.
\end{aligned}$$

The characteristic polynomial equation of (3.1) is

$$\gamma_4 Z^4 + \gamma_3 Z^3 + \gamma_2 Z^2 + \gamma_1 Z + \gamma_0 = 0 \tag{3.2}$$

where Z may be real or complex.

Any real root x_0 and pair of complex roots $x_0 \pm y_0 i$ will give rise to terms $aw^k e^{x_0 w}$, and $w^k e^{x_0 w}(b \cos y_0 w + c \sin y_0 w)$ respectively in the solution of (3.1), where $w > 0$, $0 \leqslant k \leqslant 3$, and a, b, c are constants (see e.g. Rainville and Bedient 1969). Since necessarily

$$\lim_{w \to \infty} f^{(i)}(w,j) = 0, \qquad i,j = 1,2,$$

these terms will have coefficients of zero in the pdf's when $x_0 > 0$. The product of the four roots is γ_0 which is negative in sign when the criterion given in §2 for existence of the pdf's holds, so that (3.2) possesses an odd number (1 or 3) of negative real roots. Hence we always obtain a solution of (3.2) which, when normalized, satisfies the properties of a density function. Moreover (3.2) always has exactly one real positive root (see Brill and Green 1982 for a proof). The functional forms of the pdf's $f^{(i)}(w,j)$, $w > 0$, $i,j = 1,2$ depend on the specific properties of the roots. Two cases will be considered here to illustrate the solution. Other cases are discussed in the Appendix.

Three Distinct Negative Real Roots. Denote the roots by r_j, $j = 1,2,3$. Then the SP(1) pdf's are given by

$$f^{(1)}(w,i) = \sum_{j=1}^{3} a_{ij} e^{r_j w}, \qquad i = 1,2, \tag{3.3}$$

where the constants a_{ij}, p_0 and p_1 are evaluated by means of eight linearly independent equations. The first two of these equations are obtained by substituting (3.3) into (2.5) and (2.6), and the next five are initial conditions (for details see Brill and Green 1982). The eighth equation is the normalizing condition, obtained by substituting (3.3) into (2.8). These substitutions and operations result in the following system of linear equations for the a_{ij}'s, p_0 and p_1 presented in matrix form:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & \lambda_1 & -(\lambda+\mu) \\ 0 & 0 & 0 & 1 & 1 & 1 & -\lambda & \mu \\ r_1-\lambda & r_2-\lambda & r_3-\lambda & \lambda_1 & \lambda_1 & \lambda_1 & 0 & 2\mu\lambda_1 \\ \kappa_1 & \kappa_2 & \kappa_3 & \lambda_1(r_1+2\mu) & \lambda_1(r_2+2\mu) & \lambda_1(r_3+2\mu) & 0 & 0 \\ 0 & 0 & 0 & r_1-\lambda & r_2-\lambda & r_3-\lambda & \mu\lambda_2 & 0 \\ 0 & 0 & 0 & \nu_1 & \nu_2 & \nu_3 & 0 & \mu^2\lambda_2 \\ \mu^2\lambda_2 & \mu^2\lambda_2 & \mu^2\lambda_2 & \delta_1 & \delta_2 & \delta_3 & 0 & 0 \\ -1/r_1 & -1/r_2 & -1/r_3 & -1/r_1 & -1/r_2 & -1/r_3 & 1 & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \\ p_0 \\ p_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{3.4}$$

where
$$\kappa_j = r_j^2 + (2\mu - \lambda)r_j - 2\mu\lambda_2,$$
$$\nu_j = r_j^2 + (\mu - \lambda)r_j - \mu\lambda_1,$$
$$\delta_j = r_j^3 + (2\mu - \lambda)r_j^2 + (\mu^2 - 2\mu\lambda + \mu\lambda_2)r_j - \mu^2\lambda_1 \qquad \text{for } j = 1,2,3.$$

When (3.4) is solved the SP(1) pdf's are completely determined by (3.3) and the values of p_0 and p_1. Example 3 illustrates this case.

Two Complex Conjugate Roots with Negative Real Part. The parameters may have values such that equation (2.3) has one negative real root r, and 2 complex conjugate roots with negative real part, $\alpha \pm \beta i$, where $i = \sqrt{-1}$, and r, α and β are real. The SP(1) pdf's then involve trigonometric functions and are given by

$$f^{(1)}(w,i) = a_{i1}e^{rw} + e^{\alpha w}(a_{i2}\cos \beta w + a_{i3}\sin \beta w), \qquad w > 0, \quad i = 1,2. \tag{3.5}$$

In order to solve for the a_{ij}'s, p_0, and p_1, we construct a system of linear equations similar to (3.4), resulting in:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & \lambda_1 & -(\lambda+\mu) \\ 0 & 0 & 0 & 1 & 1 & 0 & -\lambda & \mu \\ r-\lambda & \alpha-\lambda & \beta-\lambda & \lambda_1 & \lambda_1 & 0 & 0 & 2\mu\lambda_1 \\ \kappa_1 & \kappa_2 & \kappa_3 & \lambda_1(r+2\mu) & \lambda_1(\alpha+2\mu) & \lambda_1\beta & 0 & 0 \\ 0 & 0 & 0 & r-\lambda & \alpha-\lambda & \beta & \mu\lambda_2 & 0 \\ \mu^2\lambda_2 & \mu^2\lambda_2 & 0 & \nu_1 & \nu_2 & \nu_3 & 0 & 0 \\ 0 & 0 & 0 & \delta_1 & \delta_2 & \delta_3 & 0 & \mu^2\lambda_2 \\ -\dfrac{1}{r} & -\dfrac{\alpha}{\alpha^2+\beta^2} & \dfrac{\beta}{\alpha^2+\beta^2} & -\dfrac{1}{r} & -\dfrac{\alpha}{\alpha^2+\beta^2} & \dfrac{\beta}{\alpha^2+\beta^2} & 1 & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \\ p_0 \\ p_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{3.6}$$

where
$$\kappa_1 = r^2 + (2\mu - \lambda)r - 2\mu\lambda_2,$$
$$\kappa_2 = (\alpha^2 - \beta^2) + (2\mu - \lambda)\alpha - 2\mu\lambda_2,$$
$$\kappa_3 = 2\alpha\beta + (2\mu - \lambda)\beta,$$
$$\nu_1 = r^3 + (2\mu - \lambda)r^2 + (\mu^2 - 2\mu\lambda + \mu\lambda_2)r - \mu^2\lambda_1,$$
$$\nu_2 = \alpha^3 - 3\alpha\beta^2 + (2\mu - \lambda)(\alpha^2 - \beta^2) + (\mu^2 - 2\mu\lambda + \mu\lambda_2)\alpha - \mu^2\lambda_1,$$
$$\nu_3 = 3\alpha^2\beta + (2\mu - \lambda)2\alpha\beta + (\mu^2 - 2\mu\lambda + \mu\lambda_2)\beta,$$
$$\delta_1 = r^2 + (\mu - \lambda)r - \mu\lambda_1,$$
$$\delta_2 = (\alpha^2 - \beta^2) + (\mu - \lambda)\alpha - \mu\lambda_1,$$
$$\delta_3 = 2\alpha\beta + (\mu - \lambda)\beta.$$

This case arises in Example 4.

3.2. SP(2) pdf's

The pdf's of the waiting times of Type 2 customers can now be calculated using (2.2), the computed SP(1) pdf's and either (3.3) or (3.5) depending on the nature of the roots of (3.2). In the SP(2) calculations, the relative value of μ must also be considered, in order to avoid division by zero, and this results in some further cases (cf. Appendix). The pdf of an arbitrary Type 2 arrival is given by (2.12). Note that in all cases $f^{(2)}(w,2) = f^{(1)}(w,2)$ so that we only have to derive $f^{(2)}(w,1)$ once the SP(1) pdf's are known.

Three Distinct Negative Real Roots. In (3.3) if $r_j + \mu \neq 0$ for $j = 1,2,3$, substituting (3.3) into (2.2) we obtain

$$f^{(2)}(w,1) = \mu \sum_{j=1}^{3} \frac{a_{1j}e^{r_j w}}{r_j + \mu} + \mu\left(p_1 - \sum_{j=1}^{3} \frac{a_{1j}}{r_j + \mu}\right)e^{-\mu w}, \qquad w > 0, \quad \text{and} \tag{3.7}$$

$$f^{(2)}(w,2) = f^{(1)}(w,2) = \sum_{j=1}^{3} a_{2j}e^{r_j w}, \qquad w > 0. \tag{3.8}$$

This is the case in Example 3.

The case when $r_j + \mu = 0$ for some $j = 1,2,3$ is discussed in the Appendix.

Two Complex Conjugate Roots with Negative Real Part. In (3.5) if $\mu + r \neq 0$, substituting (3.5) into (2.2) yields

$$\begin{aligned} f^{(2)}(w,1) &= \frac{\mu a_{11}e^{rw}}{r+\mu} \\ &+ \frac{\mu e^{\alpha w}}{\left[(\alpha+\mu)^2+\beta^2\right]}\left[a_{12}((\alpha+\mu)\cos\beta w + \beta\sin\beta w) + a_{13}((\alpha+\mu)\sin\beta w - \beta\cos\beta w)\right] \\ &+ \left[\frac{-\mu a_{11}}{r+\mu} - \frac{\mu a_{12}(\alpha+\mu)}{\left[(\alpha+\mu)^2+\beta^2\right]} + \frac{\mu a_{13}\beta}{\left[(\alpha+\mu)^2+\beta^2\right]} - \mu p_1\right]e^{-\mu w}, \end{aligned} \tag{3.9}$$

$$w > 0, \quad \text{and}$$

$$f^{(2)}(w,2) = f^{(1)}(w,2) = a_{21}e^{rw} + e^{\alpha w}(a_{22}\cos\beta w + a_{23}\sin\beta w), \qquad w > 0. \tag{3.10}$$

This case is illustrated in Example 4. The case when $\mu + r = 0$ is discussed in the Appendix.

The probability p_0 of having zero customers in the system is, due to Poisson arrivals, also the probability of a zero wait in SP(2). It is only one component of the probability of a zero wait in SP(1). However, p_1 describes a waiting effect in SP(2) different from that in SP(1). In SP(2), p_1 is part of the probability of a positive waiting time for Type 2's, and is incorporated in the expression for $f^{(2)}(w,1)$, observable in (2.2), (3.7), and (3.9). In SP(1), p_1 is the other part of the probability of a zero wait for Type 1's.

4. Examples

The pdf's were computed by means of a computer program for varying values of λ_1, λ_2, and μ. Equations (3.2), (3.4), (3.6) or variants (see Appendix) were solved using FORTRAN and IMSL (International Mathematical and Statistical Library) subroutines in double precision, on an IBM 370 computer. Several examples are presented here which illustrate the dynamics of the model. Two examples demonstrate the correctness of the solution for known results, and exemplify the conditional nature of the virtual waiting time process.

EXAMPLE 1. $\lambda_1 = 1$, $\lambda_2 = 0$, $\mu = 2$.

When $\lambda_1 = 1$, $\lambda_2 = 0$, $\mu = 2$, the system specializes to an ordinary $M/M/2$ queue for Type 1 customers. This results in (3.2) having a pair of equal roots (one root is -3, the double root is -2). A modified form of (3.4) is solved as discussed following (A.1) (see Appendix). The solution for the a_{ij}'s, p_0 and p_1 are:

$$\begin{gathered} a_{11} = 0.3, \quad a_{12} = 0, \quad a_{13} = 0, \quad a_{2j} = 0, \quad j = 1,2,3, \\ p_0 = 0.6, \quad p_1 = 0.3, \\ f^{(1)}(w,1) = 0.3e^{-3w}, \quad f^{(1)}(w,2) = 0, \\ g^{(1)}(w) = 0.3e^{-3w}, \\ f^{(2)}(w,1) = 1.2e^{-2w} - 0.6e^{-3w}, \quad f^{(2)}(w,2) = 0, \\ g^{(2)}(w) = 1.2e^{-2w} - 0.6e^{-3w} \quad \text{for} \quad w > 0. \end{gathered} \tag{4.1}$$

Notice that the coefficients of the "Erlang" term and some of the exponentials in (A.1) are zero.

In (4.1) the SP(1) pdf's constitute the solution for an $M/M/2$ queue with arrival rate 1 and service rate 2. Even though the arrival rate of Type 2 customers is zero, the pdf of the *virtual* wait of Type 2 customers who follow Type 1's exists and is positive, due to the conditional character of the process. Both $f^{(1)}(\cdot,2)$ and $f^{(2)}(\cdot,2)$ are identically zero, as expected, since no customer can follow a Type 2 into service. The form of $g^{(2)}(\cdot)$ is the difference between exponentials. Functions $g^{(1)}(\cdot)$ and $g^{(2)}(\cdot)$ are sketched in Figure 3.

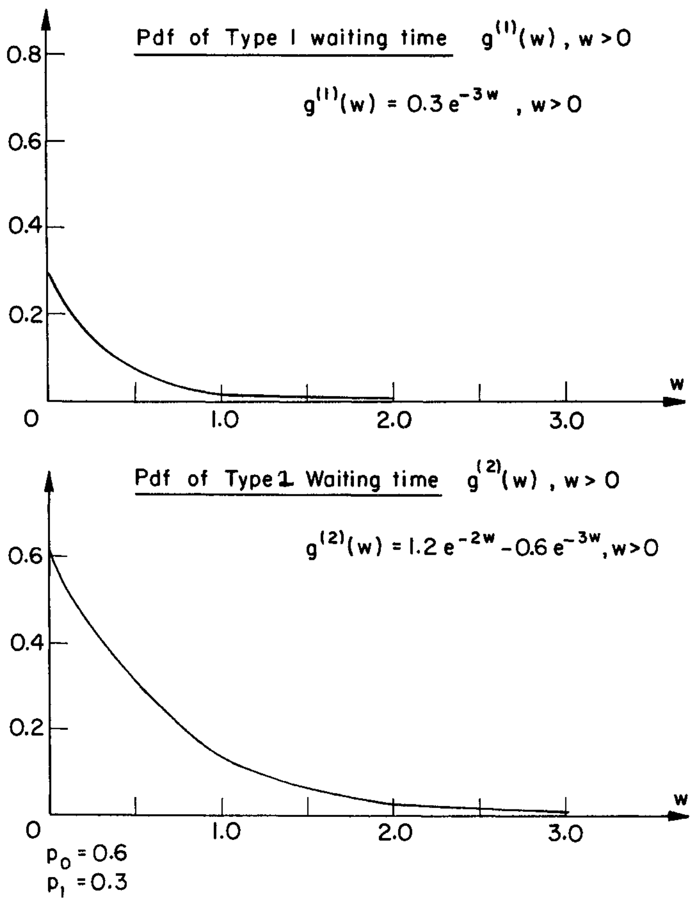

FIGURE 3

Example 1 also illustrates that for every $x > 0$,

$$P(W^{(2)} > x) > P(W^{(1)} > x), \qquad x > 0.$$

This can be seen directly from the solution, since

$$P(W^{(2)} > x) = \int_x^\infty g^{(2)}(w)\,dw = 0.6e^{-2x} - 0.2e^{-3x},$$

$$P(W^{(1)} > x) = \int_x^\infty g^{(1)}(w)\,dw = 0.1e^{-3x}.$$

EXAMPLE 2. $\lambda_1 = 0$, $\lambda_2 = 1$, $\mu = 2$.

With $\lambda_1 = 0$, $\lambda_2 = 1$, $\mu = 2$, (3.2) has three distinct negative real roots, $r_1 = -4$, $r_2 = -2$, $r_3 = -1$, and $\mu + r_2 = 0$ (see Appendix). The system reduces to an $M/M/1$ queue for Type 2 customers with arrival rate 1 and service rate 2.

The solution of system (3.4), and the resulting pdf's are $a_{ij} = 0$ for all i, j, except $a_{23} = 0.5$, $p_0 = 0.5$, $p_1 = 0$, resulting in

$$\begin{aligned} &f^{(1)}(w,1) = 0, \quad f^{(1)}(w,2) = 0.5e^{-w}, \quad g^{(1)}(w) = 0.5e^{-w}, \\ &f^{(2)}(w,1) = 0, \quad f^{(2)}(w,2) = 0.5e^{-w}, \quad g^{(2)}(w) = 0.5e^{-w}. \end{aligned} \tag{4.2}$$

The values of p_0 and $g^{(2)}(w)$, $w > 0$, are necessarily identical to those of an $M/M/1$ queue with arrival rate 1 and service rate 2 for Type 2 customers. Notice again that $g^{(1)}(w)$ is positive, although Type 1's never arrive at the queue, depicting once again the conditional nature of the virtual wait. Also, $g^{(2)}(0^+) = 0.5 = \lambda p_0$, indicating that the impact rate of the "system point" in SP(2) on the zero level is equal to its exit rate from the zero level, as in a regular $M/M/1$ queue (Brill and Posner 1981a).

EXAMPLE 3. $\lambda_1 = \lambda_2 = 1$, μ varies.

Figure 4 illustrates the graphs of $g^{(1)}(\cdot)$ and $g^{(2)}(\cdot)$ for values of $\mu = 1.76$, 2, 2.5 and 3.0, all of which yield three distinct negative real roots for (3.2) and correspond to the case treated in §3. [The existence criterion for the pdf's does not hold for $\mu \leqslant 1.75$.] When $\mu = 1.76$, $g^{(1)}(0^+) = 0.0108$ and $g^{(1)}(w)$ decreases very slowly to 0.0102 at $w = 3.0$. Function $g^{(2)}(w)$ increases very slowly from 0.0099 at 0^+ to a maximum of 0.0104 at approximately $w = 0.8$, and then decreases very slowly to 0.0102 at $w = 3.0$. In fact both $g^{(1)}(\cdot)$ and $g^{(2)}(\cdot)$ are close to uniform pdf's, at least over the interval $0 < w < 10$, while $p_0 = 0.0049$ and $p_1 = 0.0029$. For example, $g^{(1)}(9.9) = 0.0094$ and $g^{(2)}(4.9) = 0.0100$ (not shown in the figure). It is clear from Figure 4 that as μ increases from 1.75 the pdf's shift rather sharply towards the production of rapidly increasing probabilities for smaller waiting times for both types of customers. Pdf's obtained for a large number of other values of $\mu > 1.75$ indicate the same pattern emerging. Figure 4 also illustrates the sensitivity of waiting patterns to changes in μ, when the existence criterion is satisfied by a wide margin, and when it is barely satisfied. For example, when μ changes from 3 to 2.5 (a 16.7% decrease), the pdf's do not change drastically, and the two waiting times' patterns are similar. However, when μ changes from $\mu = 2$ to $\mu = 1.76$ (a 12% decrease) there is a sharp change to very long waits especially for Type 2 customers.

Figure 5 shows that as μ increases from 1.75, p_0 increases first rapidly and then more slowly; p_1 first increases gradually to a maximum in the vicinity of $\mu = 3.3$, and then slowly decreases.

EXAMPLE 4. $\lambda_1 = 2.5$, $\lambda_2 = 1$, $\mu = 4$.

The purpose of this example is to show the pdf's when the parameter values are such that there are two complex conjugate roots with negative real parts for (3.2), corresponding to the first such case in §3. The pdf's contain sine and cosine functions as in (3.6), (3.9) and (3.10), which are reflected by the mildly wavy appearance of $g^{(1)}(\cdot)$

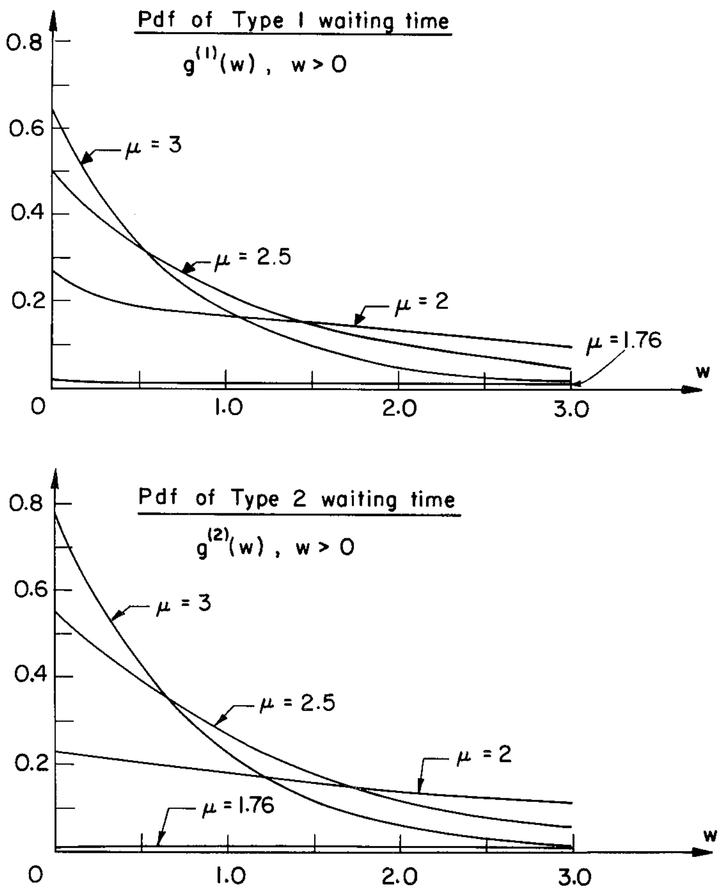

FIGURE 4

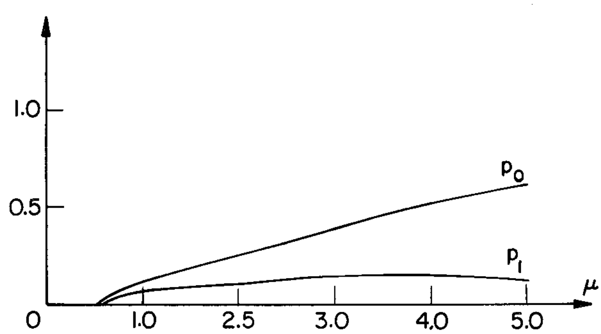

FIGURE 5

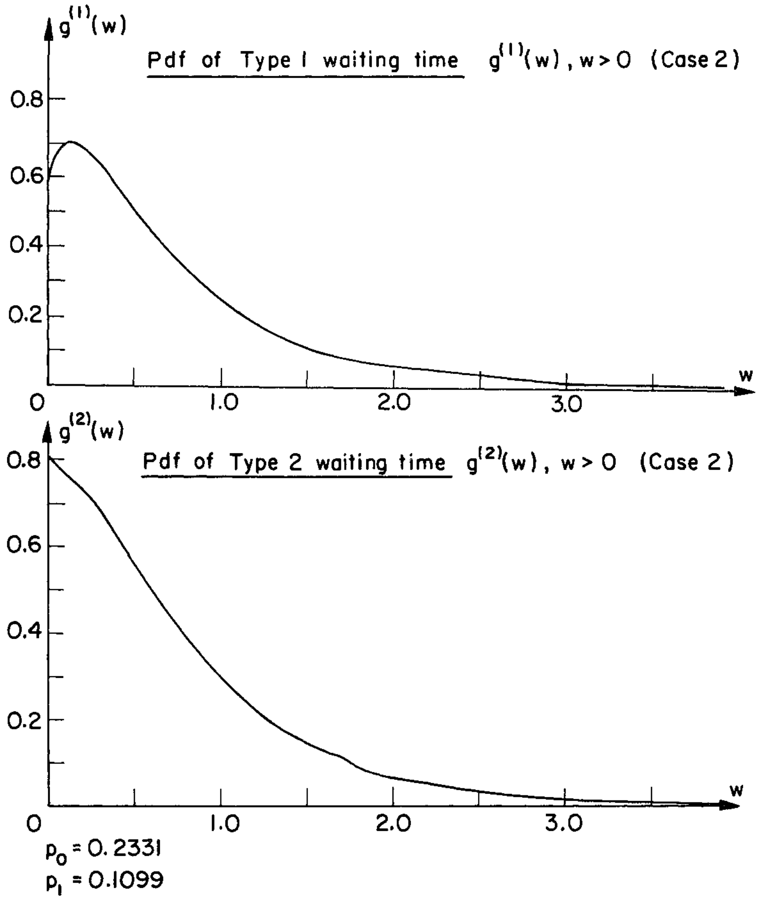

FIGURE 6

and $g^{(2)}(\cdot)$ in Figure 6. The graph of $g^{(2)}(\cdot)$ would be more sinusoidal except for the exponential term $p_1 \mu e^{-\mu w}$ in the convolution (3.9), which smooths it.

5. Generalization for Larger Systems

Some systems with more than two servers are mathematically identical or very similar to the two-server system solved in the preceding sections. For example, if a system has seven servers and there are only two types of customers—those requiring three servers and those requiring all seven, we have the same exact structure as with two servers and the solution in §4 will be valid for this larger system as well.

More generally, the definiton of the system configuration at time t given in §1 can be used in any size system to obtain set balance equations for the virtual waiting time distribution of each customer type. For example, consider the system with three servers and three customer types. Type i customers arrive according to a Poisson process at rate λ_i and require service from i servers. Types 1 and 2 customers have service times which are exponentially distributed with mean $1/\mu$ (this is easily extended to different service rates for each) while Type 3 customers have service time distribution B. By definition, there are four possible configurations for this system: $(0,0,0)$, $(1,0,0)$, $(2,0,0)$, and $(0,1,0)$, which can be simplified notationally to $(0,0)$, $(1,0)$, $(2,0)$ and $(0,1)$. Let p_m be the equilibrium probability of nonwaiting state $m = (0,0)$, $(1,0)$, $(2,0)$ or $(0,1)$ for a Type 1 customer. Then for each SP and each configuration, we can obtain by inspection a set of equations balancing the rates of entrances and exits of the

SP into the set $(0, w)$. The equation corresponding to configuration $(0,0)$ for SP(1) is:

$$f_{00}^{(1)}(w) + \lambda_3\Big[B(w)p_{00} + (B * \exp(\mu))(w)p_{01} + \int_0^w (B * \exp(\mu))(w-\alpha)f_{01}^{(1)}(\alpha)\,d\alpha$$
$$+ (B * \exp(\mu) * \exp(2\mu))(w)p_{20}$$
$$+ \int_0^w (B * \exp(\mu) * \exp(2\mu))(w-\alpha)f_{20}^{(1)}(\alpha)\,d\alpha + (B * \exp(\mu))(w)p_{10}$$
$$+ \int_0^w (B * \exp(\mu))(w-\alpha)f_{10}^{(1)}(\alpha)\,d\alpha\Big]$$
$$= f_{00}^{(1)}(0^+) + (\lambda_1 + \lambda_2)\int_0^w f_{00}^{(1)}(\alpha)\,d\alpha + \lambda_3\int_0^w (1 - B(w-\alpha))f_{00}^{(1)}(\alpha)\,d\alpha, \tag{5.1}$$

where $\exp(\beta)$ is the exponential distribution with rate β and "$*$" denotes convolution. The equations for the remaining three configurations are also easily derived. The zero waiting states balance equations are:

$$f_{00}^{(1)}(0^+) + \mu(p_{10} + p_{01}) = \lambda p_{00}, \tag{5.2}$$

$$f_{10}^{(1)}(0^+) + 2\mu p_{20} + \lambda_1 p_{00} = (\lambda + \mu)p_{10}, \tag{5.3}$$

$$f_{20}^{(1)}(0^+) + \lambda_1 p_{10} = (\lambda + 2\mu)p_{20}, \tag{5.4}$$

$$f_{01}^{(1)}(0^+) + \lambda_2 p_{00} = (\lambda + \mu)p_{01}, \tag{5.5}$$

and the normalizing condition is

$$p_{00} + p_{10} + p_{20} + p_{01} + \int_0^\infty \left[f_{00}^{(1)}(w) + f_{10}^{(1)}(w) + f_{20}^{(1)}(w) + f_{01}^{(1)}(w)\right]dw = 1. \tag{5.6}$$

The general case of R servers and R customer types can be treated in a similar fashion by solving SP(1), and then the other SP's by using the SP(1) results.

The ease of solution will depend upon the number of pages (configurations) for the system which with the zero-waiting states determines the number of equations and unknowns that have to be solved. The number of pages (configurations) is the number $N(R)$ of nonnegative integer solutions $(x_1, \ldots, x_{R-1})$ of the inequality

$$1x_1 + 2x_2 + \cdots + (R-1)x_{R-1} \leqslant R - 1, \qquad R \geqslant 2. \tag{5.7}$$

This inequality is interesting since $N(R)$ is the sum of the number of unrestricted partitions of the integers $0, \ldots, R-1$. Partitions have been treated in L. Euler (1748), G. H. Hardy and S. Ramanujan (1918), S. Ramanujan (1919), C. E. Gupta and J. C. P. Miller (1958), and the works of other well-known mathematicians. Tables of unrestricted partitions for values of R up to 101 are given in Hardy and Ramanujan (1918) along with recurrence relations for computing them for any positive integer. Table 5.1 gives $N(R)$ for several values of R, using Table 1 in Hardy and Ramanujan (1918).

TABLE 5.1

The Number of Configurations $N(R)$ for R Servers and R Customer Types
(*Values Obtained Using Table* 1 *of Gupta and Miller* (1958))

R	2	3	4	5	6	7	8	9	10	15	20	25	30
$N(R)$	2	4	7	12	19	30	45	67	97	502	2087	7338	23025

The solution technique used in this paper will, of course, only be practical for a given application if the configuration space and hence the cost of solution is not unduly large relative to the potential benefits of obtaining an exact solution. However, the methodology can serve as a guide for devising numerical and approximation techniques which would be more appropriate for large values of R (e.g. $R > 6$) since $N(R)$ grows rapidly with R. This dimensionality problem will not be as severe, of course, when the number of customer types is smaller than the total number of servers as is the case in many real applications. For example, in communications systems, there are typically only two or three kinds of transmissions requiring different numbers of channels (see e.g. Gimpelson 1965, Nigam 1975).

The system point model equations, such as (5.1)–(5.4), can always be reduced to algebraic equations for the Laplace transforms of the pdf's, since every integral is either a convolution, or a simple transform of the pdf. Initial conditions are given by (5.5) and (5.6). General numerical methods for solving the system point model equations are presently being developed.

Appendix. Other Cases Generated by Special Values of the Parameters

When the parameters λ_1, λ_2 and μ have values which cause (3.2) to have a pair of equal roots, or make the value of some root equal to $-\mu$, etc. the functional forms of the pdf's differ from those given in (3.3), (3.5) or (3.7)–(3.10). We shall list these cases here and give the corresponding functional forms for the pdf's.

Three Distinct Negative Real Roots for (3.2)

In (3.3) if $r_j + \mu = 0$, only the SP(2) pdf $f^{(2)}(w, 1)$ requires adjustment to avoid division by zero. Thus, if $r_j + \mu = 0$, then

$$\int_0^w a_{1j} e^{r_j x} \mu e^{-\mu(w-x)}\, dx = a_{1j} \mu w e^{-\mu w},$$

and this term would replace the expression

$$\frac{\mu a_{1j}}{r_j + \mu}\left[e^{r_j w} - e^{-\mu w}\right]$$

in (3.7).

Two Complex Conjugate Roots with Negative Real Parts for (3.2)

If $\mu + r = 0$, then $f^{(2)}(w, 1)$ has exactly the same form as (3.9), except that $\mu a_{11}(e^{rw} - e^{-\mu w})/(r + \mu)$ is replaced by $a_{11} \mu w e^{-\mu w}$.

Double Negative and One Distinct Negative Real Root for (3.2)

If r_1, r_2, r_3 are the negative real roots let $r_2 = r_3 = s$, so that s is a double root. The SP(1) pdf's then have the functional form

$$f^{(1)}(w, i) = a_{i1} e^{r_1 w} + a_{i2} e^{sw} + a_{i3} w e^{sw}, \qquad w > 0, \quad i = 1, 2. \tag{A.1}$$

The a_{ij} constants in (A.1), p_0 and p_1, are calculated by solving a matrix equation similar to (3.4). The resulting matrix is identical to that of (3.4), except for the following changes in columns three and six. Denoting the entry in row i and column j

by c_{ij}, the new entries are:

$$\begin{aligned}
c_{13} &= 0,\\
c_{33} &= 1,\\
c_{43} &= 2s + 2\mu - \lambda,\\
c_{73} &= 0,\\
c_{83} &= 1/s^2 \qquad \text{in column 3;} \quad \text{and}\\
c_{26} &= 0,\\
c_{36} &= 0,\\
c_{56} &= 1,\\
c_{66} &= 2s + \mu - \lambda,\\
c_{76} &= 3s^2 + 2(2\mu - \lambda)s + \mu^2 - 2\mu\lambda + \mu\lambda_2,\\
c_{86} &= 1/s^2 \qquad \text{in column 6.}
\end{aligned}$$

This case arises if $\lambda_1 > 0$, $\lambda_2 = 0$, $\mu > 0$; the model then reduces to an ordinary $M/M/2$ queue with arrival rate λ_1 and service rate μ for Type 1 customers. This occurs in Example 1.

If $(-\mu)$ is not equal to any root, substitution of (A.1) into (2.2) yields

$$f^{(2)}(w,1) = \frac{\mu a_{11} e^{rw}}{r_1 + \mu} + \left[\frac{\mu a_{12}}{s+\mu} - \frac{\mu a_{13}}{(s+\mu)^2} \right] e^{sw} + \frac{\mu a_{13} w}{s+\mu} e^{sw}$$

$$+ \mu \left[-\frac{a_{11}}{r_1+\mu} - \frac{a_{12}}{s+\mu} + \frac{a_{13}}{(s+\mu)^2} + p_1 \right] e^{-\mu w}, \qquad w > 0. \quad \text{(A.2)}$$

If $(-\mu)$ equals a root, it may equal r_1 or s.

If $\mu + r_1 = 0$, $f^{(2)}(w,1)$ is the same as (A.2) with $\mu a_{11}[e^{rw} - e^{-\mu w}]/(r_1 + \mu)$ replaced by $a_{11}\mu w e^{-\mu w}$.

If $\mu + s = 0$, $f^{(2)}(w,1)$ is given by

$$f^{(2)}(w,1) = \frac{a_{11}\mu}{r+\mu} e^{rw} - \mu\left(\frac{a_{11}}{r+\mu} - p_1 \right) e^{-\mu w} + a_{12}\mu w e^{-\mu w}$$

$$+ \frac{a_{13}\mu}{2} w^2 e^{-\mu w}, \qquad w > 0. \quad \text{(A.3)}$$

Two Other Cases

We have not strictly ruled out that (3.2) may possibly have two complex roots with nonnegative real parts. In that case let r be the one negative real root. Then $f^{(1)}(w,j) = a_j e^{rw}$ $(j = 1, 2)$ and $f^{(2)}(w,2) = f^{(1)}(w,2)$ as it always does. Using (2.2) if $r + \mu \neq 0$, we get

$$f^{(2)}(w,1) = \frac{a_2 \mu e^{rw}}{r+\mu} + \left(\frac{a_2}{r+\mu} - p_1 \right) \mu e^{-\mu w}$$

and if $r + \mu = 0$,

$$f^{(2)}(w,1) = (a_1 + p_1)\,\mu w e^{-\mu w}.$$

If (3.2) should have a triple negative root, s, then

$$f^{(1)}(w,i) = a_{i1}e^{sw} + a_{i2}we^{sw} + a_{i3}e^{sw} \qquad (i = 1,2) \quad \text{and} \quad f^{(2)}(w,2) = f^{(1)}(w,2).$$

The functional form of $f^{(2)}(w, 1)$ is easily calculated using (2.2) and by considering the two cases $s + \mu \neq 0$ and $s + \mu = 0$ separately, in a similar manner to the derivation of (A.2) and (A.3). Neither of the two cases in this subsection has occurred in a multitude of numerical examples.[1]

[1]This research was sponsored by the Faculty Research Fund of the Graduate School of Business, Columbia University and by The National Science and Engineering Research Council of Canada, grant no. A1775.

References

ARTHURS, E. AND J. S. KAUFMAN, "Sizing a Message Store Subject to Blocking Criteria," 4th International Symposium on Modelling and Performance Evaluation of Computer Systems, Vienna, 1979.

BRILL, P. H., "System Point Theory in Exponential Queues," Ph.D. Dissertation, University of Toronto, 1975.

———, "An Embedded Level Crossing Technique for Dams and Queues," *J. Appl. Probab.*, 16 (1979), 174–186.

——— AND L. GREEN, "Queues in Which Customers Receive Simultaneous Service from a Random Number of Servers—A System Point Approach," Research Working Paper No. 446A, Graduate School of Business, Columbia University, 1982.

——— AND M. J. M. POSNER, "The System Point Method in Exponential Queues: A Level Crossing Approach," *Math. Oper. Res.*, 6 (1981a), 31–49.

——— AND ———, "A Two-Server Queue with Non-Waiting Customers Receiving Specialized Service," *Management Sci.*, 27 (1981b), 914–925.

DICKSON, L. E., *New First Course in the Theory of Equations*, John Wiley & Sons, Inc., New York, 1939.

EULER, L., *Introduction in Analysin Infinitorum*, Vol. 1, 1748.

GILLENT, F., "File D'Attente ou le Nombre de Serveurs Demande par le Client est Aléatoire," Thèse de Licenciée en Sciences Mathématiques, Univ. Libre de Bruxelles, 1980.

GIMPELSON, L. A., "Analysis of Mixtures of Wide- and Narrow-Band Traffic," *IEEE Trans. Comm. Technology*, 13 (1965), 258–266.

GREEN, L., "Queues Which Allow a Random Number of Servers per Customer," Ph.D. Dissertation, Yale University, 1978.

———, "A Queueing System in which Customers Require a Random Number of Servers," *Oper. Res.*, 28 (1980), 1335–1346.

———, "Comparing Operating Characteristics of Queues in which Customers Require a Random Number of Servers," *Management Sci.*, 27 (1981a), pp. 65–74.

———, "A Queueing System with Auxiliary Servers," Research Working Paper No. 334A, Graduate School of Business, Columbia University, 1981b.

GUPTA, C. E. AND J. C. P. MILLER, *Royal Society Mathematical Tables, Vol. 4, Tables of Partitions*, Cambridge University Press, London, 1958.

HARDY, G. H. AND S. RAMANUJAN, "Asymptotic Formulae in Combinatory Analysis," *Proc. London Math Soc., Sec.* 2, 17 (1918), 75–115.

KIM, S., "$M/M/s$ Queueing System Where Customers Demand Multiple Server Use," Ph.D. Dissertation, Southern Methodist University, 1979.

NEUTS, M. F., *Matrix-Geometric Solutions in Stochastic Models*, Johns Hopkins University Press, Baltimore, 1980.

NIGAM, A. K., "Analysis for a Satellite Communications System," *Interfaces*, 5 (1975), 37–47.

OMAHEN, K. J., "Capacity Bounds for Multiresource Queues," *Comm. ACM*, 23 (1977), 646–663.

RAINVELLE AND BEDIENT, *Elementary Differential Equations*, MacMillan, New York, 1969.

RAMANUJAN, S., "Some Properties of $p(n)$, the Number of Partitions of n," *Proc. Cambridge Philos. Soc.*, 19 (1919), 207–210.

WOLMAN, E., "The Camp-On Problem for Multiple-Address Traffic," *Bell System Tech. J.*, 51 (1972), 1363–1422.