AT&T Technical Journal
Vol. 64, No. 5, May–June 1985
Printed in U.S.A.

Sojourn Time Distribution in a Multiprogrammed Computer System

By K. M. REGE and B. SENGUPTA*

(Manuscript received July 5, 1984)

We present a method for calculating the moments and the distribution of sojourn time in a multiprogrammed computer system. We assume that the CPU and I/O subsystem can be represented by a general state-dependent server who works according to the processor sharing discipline. Further, at most m jobs may be simultaneously receiving service. Thus, m is the multiprogramming level of the system. The arrival of jobs occurs according to a Poisson process, and the arrivals must wait in a waiting area if m jobs are already receiving service. The method presented may be useful in designing the multiprogramming level needed to meet certain objectives on the characteristics of the sojourn time.

I. INTRODUCTION

This paper is concerned with finding the moments and the distribution of sojourn time in a multiprogrammed computer system. We assume that jobs arrive into a computer system according to a Poisson process at a rate of λ. The computer system is divided into two areas, a waiting area and a service area. The service area can hold at most m jobs, where m is the multiprogramming level of the system. The jobs in the service area receive service from a server whose service rate is assumed to be state dependent. The server operates at a rate μ_j [using the Processor Sharing (PS) discipline] whenever there are j jobs in the service area with $\mu_m = \mu$. We assume that each job's service

* Authors are employees of AT&T Bell Laboratories.

time is exponentially distributed, so that on every service completion, each customer in the service area is equally likely to leave the system. An arrival into the system goes directly into the service area if it is not full. An arriving job goes to the waiting area of unlimited size if there are already m jobs in the service area. The customers are drawn from the waiting area into the service area according to the First-In First-Out (FIFO) discipline. Figure 1 depicts this situation schematically. The model may be useful in determining the multiprogramming level needed to meet a certain response-time characteristic.

The model is an obvious generalization of the M/M/1 First-Come First-Served (FCFS) queue ($m = 1$) and of the M/M/1 PS queue ($\mu_j = \mu$ and $m = \infty$). A little thought should convince the reader that the M/M/m FCFS queue is also a special case of this model with $\mu_i = i\sigma$ for $i = 1, \cdots, m$, where σ is the service rate of each server. Avi-Itzhak and Heyman had proposed the use of a state-dependent server to approximate the CPU and I/O subsystem.[1] We depict a multiple CPU and disk subsystem in Fig. 2, which is approximated by a state-dependent server in our model. The service rate μ_i of our model is obtained by solving for the throughput in the closed queueing network of Fig. 2 with a population size of i. Reiser and Lavenberg describe a

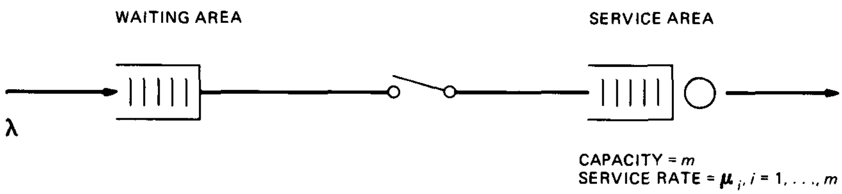

Fig. 1—Model of a multiprogrammed computer.

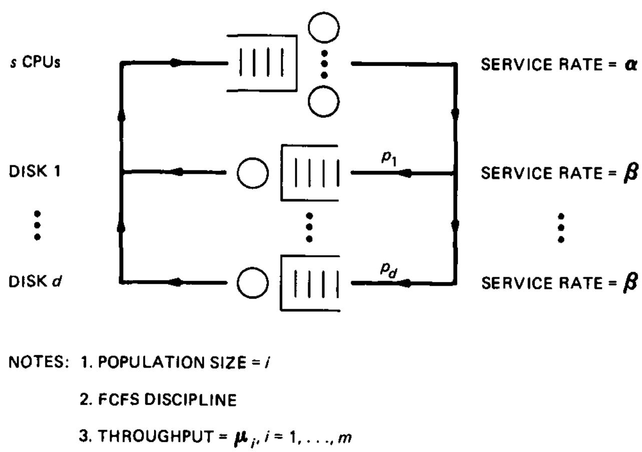

Fig. 2—Model of a multiple CPU and disk subsystem.

method of solving the throughput of such a closed queueing network.[2] Fredericks obtains approximations for mean delays in a multiprogrammed computer system and discusses the question of accuracy of the state-dependent server model.[3] Konheim and Reiser examine a computer system with one disk and one CPU, subject to a bound on the number of jobs present.[4] Mitra obtains the waiting time distribution for a computer system fed by jobs from a finite number of terminals.[5] Salza and Lavenberg investigate hierarchical decomposition methods for approximating response-time distributions in certain closed queueing network models of computer performance.[6] Coffman, Muntz, and Trotter obtain the sojourn time distribution of the M/M/1 queue with processor sharing discipline.[7]

In Section II of this paper, we discuss some preliminary results. Sections III and IV are concerned with finding the moments of sojourn time. In Section V, we provide a method of obtaining the distribution of sojourn time. In Section VI, we present some numerical examples.

II. DEFINITIONS AND PRELIMINARIES

Let us define the following quantities:

N = Number in system seen by an arrival, excluding itself.

$P_n = P(N = n)$.

I = Number in system seen by a customer entering the service area, including itself.

$q_i = P(I = i)$.

U = Time spent in the waiting area by an arrival.

$F(u) = P(U \leq u)$.

$z_n = E(U^n)$.

T = Time spent in the service area by a tagged customer.

$B_i(s)$ = Laplace-Stieltjes transform of the conditional distribution of T given $I = i$, i.e.,

$$B_i(s) = \int_0^\infty e^{-st} dP(T \leq t \mid I = i).$$

$$x_{in} = E(T^n \mid I = i) = (-1)^n \frac{d^n}{ds^n} B_i(s) \Big|_{s=0} .$$

$$V = \text{sojourn time, i.e., } V = U + T.$$

By solving the equation of a birth and death process, one obtains the following results directly.

Let

$$w_n = \prod_{j=1}^{n} \frac{\lambda}{\mu_j} \quad \text{for} \quad n = 1, \cdots, m.$$

Then

$$P_o = \left[1 + \sum_{n=1}^{m-1} w_n + \frac{w_m}{1-\rho}\right]^{-1},$$

where

$$\rho = \lambda/\mu_m = \lambda/\mu$$

and

$$P_n = \begin{cases} w_n P_o & \text{if } \ n = 1, \cdots, m \\ \rho^{n-m} P_m & \text{if } \ n \geq m + 1. \end{cases}$$

Let L be the mean number in the system. Then

$$L = P_o \sum_{n=1}^{m-1} n w_n + \frac{P_m \rho}{(1-\rho)^2} + \frac{m P_m}{1-\rho}.$$

Let W be the mean time spent in the system. Then, from Little's law,

$$W = L/\lambda.$$

III. CHARACTERIZATION OF U, I, AND V

In this section, we will derive the distribution and moments of U, and the distribution of I, and we will characterize the moments of V. It is clear that

$$U = \begin{cases} 0 \quad \text{if} \quad N \leq m - 1 \\ \text{sum of } (N - m + 1) \text{ independent exponentials each} \\ \qquad\qquad \text{of rate } \mu \text{ if } N \geq m. \end{cases}$$

Then,

$$F(u) = \sum_{i=0}^{m-1} P_i + \sum_{k=m}^{\infty} P_k \int_0^u \frac{\mu e^{-\mu t}(\mu t)^{k-m}}{(k-m)!}\, dt \quad \text{for} \quad u \geq 0$$

and

$$z_n = EU^n = \int_0^\infty u^n dF(u) = \sum_{k=m}^{\infty} P_k \int_0^\infty u^n \frac{\mu e^{-\mu u}(\mu u)^{k-m}}{(k-m)!}\, du.$$

This can be shown to be equal to

$$\frac{P_m n!}{(1-\rho)(\mu-\lambda)^n}.$$

This result could also have been obtained as follows:

$$z_n = EU^n = E(U^n \mid N \geq m)P(N \geq m) + E(U^n \mid N < m)P(N < m).$$

Given that a customer waits (i.e., $N \geq m$), the waiting room behaves like an M/M/1 FCFS queue. Thus, the wait time corresponds to the sojourn time of an M/M/1 queue and is exponential with parameter $(\mu - \lambda)$. So,

$$z_n = \frac{n!}{(\mu - \lambda)^n} \cdot \frac{P_m}{1 - \rho}.$$

We will now derive the distribution of I. Consider first the case where $I > m$:

$$\begin{aligned} q_i &= \int_{0+}^{\infty} P(I = i \mid U = u) dF(u) \\ &= \int_0^{\infty} \sum_{k=m}^{\infty} \frac{e^{-\lambda u}(\lambda u)^{i-m}}{(i-m)!} \cdot P_k \cdot \frac{\mu e^{-\mu u}(\mu u)^{k-m}}{(k-m)!} \, du. \end{aligned}$$

After some algebra one can show that

$$q_i = P_i \quad \text{for} \quad i > m.$$

This can be explained by the fact that an arrival to and a departure from the waiting area see the same distribution of the number in the waiting area and that the number in the service area is constant, given that $N > m$. For $I < m$, the number in the system including the arrival is one more than the number seen by the arrival. So

$$q_i = P_{i-1} \quad \text{for} \quad i = 1, \cdots, m-1.$$

For $I = m$, it is possible for an arrival into the service area to see m customers including itself in one of two possible ways. First, there were $m - 1$ customers in the system and an arrival occurred. Second, the arrival saw at least m customers and no one arrived during its wait. So,

$$q_m = P_{m-1} + \int_{0+}^{\infty} P(I = m \mid U = u) dF(u) = P_{m-1} + P_m$$

by the arguments used above. Thus,

$$q_i = \begin{cases} P_{i-1} & \text{if} \quad i = 1, \cdots, m-1 \\ P_{m-1} + P_m & \text{if} \quad i = m \\ P_i & \text{if} \quad i > m. \end{cases}$$

We now characterize the moments of the sojourn time in terms of x_{in}. In the next section, we will show how to calculate x_{in}.

$$EV^n = \sum_{j=0}^{n} \frac{n!}{j!(n-j)!} E(T^j U^{n-j})$$

and

$$E(T^j U^{n-j}) = \int_{0^-}^{\infty} u^{n-j} E(T^j \mid U = u) dF(u)$$

$$= \sum_{i=1}^{\infty} \int_{0^-}^{\infty} u^{n-j} E(T^j \mid I = i) P(I = i \mid U = u) dF(u).$$

The last step follows from the fact that T is independent of U given I. We now consider three cases:

1. For $j = n$,

$$ET^n = \sum_{i=1}^{\infty} x_{in} q_i.$$

2. For $0 < j < n$, the expression for $E(T^j U^{n-j})$ is

$$\sum_{i=m}^{\infty} \int_{0^-}^{\infty} u^{n-j} x_{ij} \frac{e^{-\lambda u}(\lambda u)^{i-m}}{(i-m)!} dF(u),$$

which, after some algebra, reduces to

$$P_m \sum_{i=m}^{\infty} \frac{(i+n-j-m)!}{(i-m)!} \frac{x_{ij}}{\lambda^{n-j}} \rho^{i+n-j-m}.$$

3. For $j = 0$, but $n \neq 0$, the expression is simply

$$EU^n = z_n = \frac{P_m n!}{(1-\rho)(\mu-\lambda)^n}.$$

So,

$$E(T^j U^{n-j}) = \begin{cases} \sum_{i=1}^{\infty} x_{in} q_i & \text{if} \quad j = n \\ \frac{P_m}{\lambda^{n-j}} \sum_{i=m}^{\infty} \frac{(i+n-j-m)!}{(i-m)!} x_{ij} \rho^{i+n-j-m} & \text{if} \quad 0 < j < n \\ \frac{P_m n!}{(1-\rho)(\mu-\lambda)^n} & \text{if} \quad j = 0, \text{ but } n \neq 0. \end{cases}$$

IV. CALCULATION OF x_{in}

In this section, we show how to calculate x_{in}. The transforms $B_i(s)$ satisfy the following equations:

$$B_{i+1}(s) = \left[\frac{\lambda + \mu_{i+1}}{\lambda + \mu_{i+1} + s}\right] \cdot \left[\frac{\lambda}{\lambda + \mu_{i+1}} B_{i+2}(s) + \frac{\mu_{i+1}}{\lambda + \mu_{i+1}} (c_{i+1} B_i(s) + (1 - c_{i+1}))\right] \quad (1)$$

for $i = 1, 2, \cdots$ and

$$B_1(s) = \left[\frac{\lambda + \mu_1}{\lambda + \mu_1 + s}\right]\left[\frac{\lambda}{\lambda + \mu_1} B_2(s) + \frac{\mu_1}{\lambda + \mu_1}\right], \qquad (2)$$

where

$$c_i = \begin{cases} \dfrac{i-1}{i} & \text{if} \quad i < m \\ \dfrac{m-1}{m} & \text{if} \quad i \geq m \end{cases} \quad \text{and} \quad \mu_i = \begin{cases} \mu & \text{if} \quad i < m \\ \mu & \text{if} \quad i \geq m. \end{cases}$$

These equations are obtained by assuming that the system is in state $i + 1$ in steady state and conditioning on the time of the next event. An event is defined to be a service completion or an arrival, whichever occurs first. By rearrangement,

$$B_{i+2}(s) - \frac{\lambda + \mu_{i+1} + s}{\lambda} B_{i+1}(s) + \frac{\mu_{i+1}c_{i+1}}{\lambda} B_i(s) = -\frac{\mu_{i+1}}{\lambda}(1 - c_{i+1}).$$

Taking the nth derivative, multiplying by $(-1)^n$, and setting $s = 0$, we get

$$x_{i+2,n} - \frac{\lambda + \mu_{i+1}}{\lambda} x_{i+1,n} + \frac{\mu_{i+1}c_{i+1}}{\lambda} x_{i,n} = -\frac{n}{\lambda} x_{i+1,n-1} \qquad (3)$$

for $i \geq 1$, $n = 1, 2, \cdots$, and

$$x_{2,n} - \frac{\lambda + \mu_1}{\lambda} x_{1,n} = -\frac{n}{\lambda} x_{1,n-1}, \qquad (4)$$

where

$$x_{i,0} = 1, \qquad i \geq 1. \qquad (5)$$

Equation (3) is a second-order partial difference equation with variable coefficients. However, we note that for $i \geq m$, the coefficients do not vary with i. So we will first show a method of solving an equivalent system with coefficients that do not vary with i and then present a procedure for the solution of eq. (3). The interested reader is referred to Boole[8] or Jagerman[9] for details of the techniques used to solve this difference equation. Consider the equation

$$y_{i+2,n} - \frac{\lambda + \mu}{\lambda} y_{i+1,n} + \frac{\mu c}{\lambda} y_{in} = -\frac{n}{\lambda} y_{i+1,n-1} \qquad (6)$$

for positive integer valued i and n, where $c = c_m$. For the homogeneous version of this equation, the solution is

$$y_{in} = A_n\sigma_1^i + B_n\sigma_2^i,$$

where σ_1 and σ_2 are the roots of

$$\sigma^2 - \frac{\lambda + \mu}{\lambda}\sigma + \frac{\mu c}{\lambda} = 0,$$

i.e.,

$$\sigma_1 = \frac{[\lambda + \mu + \sqrt{\lambda^2 + \mu^2 + 2\lambda\mu(1 - 2c)}]}{2\lambda}$$

and

$$\sigma_2 = \frac{[\lambda + \mu - \sqrt{\lambda^2 + \mu^2 + 2\lambda\mu(1 - 2c)}]}{2\lambda}.$$

It can be shown that $\sigma_1 > 1 > \sigma_2$. The constants A_n and B_n will be evaluated later. To find a particular solution to (6), we define the translation operator E and the difference operator Δ such that

$$Eu_i = u_{i+1}$$

and

$$\Delta u_i = u_{i+1} - u_i = (E - 1)\mu_i.$$

We can then rewrite (6) as

$$(E - \sigma_1)(E - \sigma_2)y_{in} = -\frac{n}{\lambda}y_{i+1,n-1}.$$

The solution is

$$y_{in} = \frac{1}{(\sigma_1 - \sigma_2)}\left(\frac{1}{E - \sigma_1} - \frac{1}{E - \sigma_2}\right)\left(-\frac{n}{\lambda}y_{i+1,n-1}\right).$$

In order to solve this, we must evaluate $(E - \sigma)^{-1}\left(-\frac{n}{\lambda}y_{i+1,n-1}\right)$.

Now,

$$(E - \sigma)^{-1}\left(-\frac{n}{\lambda}y_{i+1,n-1}\right) = (E - \sigma)^{-1}\sigma^{i+1}\sigma^{-(i+1)}\left(-\frac{n}{\lambda}y_{i+1,n-1}\right)$$

$$= \sigma^{i+1}(\sigma E - \sigma)^{-1}\sigma^{-(i+1)}\left(-\frac{n}{\lambda}y_{i+1,n-1}\right).$$

The last step is obtained from the shift formula (see footnote on page 73 of Boole[8]):

$$f(E)(a^i u_i) = a^i f(aE)u_i.$$

The expression now is

$$\sigma^{i+1}[\sigma(E-1)]^{-1}\sigma^{-(i+1)}\left(-\frac{n}{\lambda}y_{i+1,n-1}\right)$$

$$=\sigma^{i}\Delta^{-1}\left[\sigma^{-(i+1)}\left(-\frac{n}{\lambda}y_{i+1,n-1}\right)\right]$$

$$=-\frac{n}{\lambda}\sum_{j=1}^{i}\sigma^{i-j}y_{j,n-1}.$$

So a particular solution is

$$y_{in}=-\left(\frac{n}{\lambda(\sigma_1-\sigma_2)}\right)\sum_{j=1}^{i}(\sigma_1^{i-j}-\sigma_2^{i-j})y_{j,n-1}.$$

The general solution is

$$y_{in}=-\left(\frac{n}{\lambda(\sigma_1-\sigma_2)}\right)\sum_{j=1}^{i}(\sigma_1^{i-j}-\sigma_2^{i-j})y_{j,n-1}+A_n\sigma_1^{i}+B_n\sigma_2^{i} \qquad (7)$$

for positive integer valued i and n. It is easy to verify that this satisfies eq. (6).

4.1 Calculation of A_n

We first state that y_{in} has a probabilistic interpretation. We replace the original system by a new one in which the service rate of the server is μ, regardless of the number of customers in the service area. Further, consider a tagged customer whose probability of leaving the system is $1/m$ at each service completion, whenever there are at least two customers in the service area. Then eq. (6) describes the behavior of the conditional moments of the time spent in the service area by the tagged customer in this new system for $i \geq 2$. Clearly, as i tends to infinity, the conditional moments of the original and new systems are identical. Further, as i tends to infinity, the server is going to operate at rate μ for a long time in the original system. If we now tag a customer in the service area, it will leave the system with probability $1/m$ and, therefore, T is exponential with rate μ/m as i tends to infinity in the original system. So

$$\lim_{i\to\infty} y_{in}=\lim_{i\to\infty} x_{in}=n!(m/\mu)^{n}.$$

If we divide eq. (7) by σ_1^{i} and let i tend to infinity, we have

$$A_n=\frac{n}{\lambda(\sigma_1-\sigma_2)}\sum_{j=1}^{\infty}\sigma_1^{-j}y_{j,n-1}.$$

Thus,

$$y_{in} = \frac{n}{\lambda(\sigma_1 - \sigma_2)} \left(\sum_{j=i+1}^{\infty} \sigma_1^{i-j} y_{j,n-1} + \sum_{j=1}^{i} \sigma_2^{i-j} y_{j,n-1} \right) + B_n \sigma_2^i. \qquad (8)$$

At this point, it is possible to verify that this result indeed yields the familiar answer for the M/M/1 FCFS queue. To do this, one has to set $m = 1$. This yields $c = 0$ and $\sigma_2 = 0$. After some algebra, one obtains $y_{in} = n!/\mu^n$ for $i \geq 1$ as expected.

4.2 Calculation of B_n

We have so far been able to find y_{in} up to a constant, and a comparison of (3) and (6) should convince the reader that y_{in} obtained from (8) satisfies (3) for $i \geq m$. Thus, after finding x_{in} for $i \geq m$, we can recursively compute $x_{m-1,n}, \cdots, x_{1n}$ by using (3). It is clear that each of these variables is a linear function of the unknown constant B_n. Finally, we will find B_n to satisfy (4). This is done by choosing two trial values of B_n, evaluating two sets of x_{in}, and using linear interpolation to satisfy (4). The formal procedure to do this is as follows:

1. Set $n = 1$. Set $x_{i0} = y_{i0} = 1$ for all $i \geq 1$.
2. Select two trial values of B_n, say B_n^1 and B_n^2.
3. For each B_n^k ($k = 1, 2$), use (8) to obtain y_{in}^k for $i \geq 1$ and $k = 1, 2$. In this step, the terms $y_{j,n-1}$ on the right-hand side of (8) should not be indexed by k.
4. Set $x_{in}^k = y_{in}^k$ for $i \geq m$ and $k = 1, 2$.
5. Use (3) recursively to compute $x_{m-1,n}^k, \cdots, x_{1n}^k$ for $k = 1, 2$. In this step, the terms $x_{i+1,n-1}$ on the right-hand side of (3) should not be indexed by k.
6. From x_{in}^k ($k = 1, 2$) evaluate the left-hand side of (4). Let L_n^k be the left-hand side corresponding to B_n^k for $k = 1, 2$.
7. Since x_{in} is a linear function of B_n for all i, we have

$$B_n = \gamma B_n^1 + (1 - \gamma) B_n^2,$$

$$x_{in} = \gamma x_{in}^1 + (1 - \gamma) x_{in}^2,$$

and

$$Y_{in} = \gamma Y_{in}^1 + (1 - \gamma) Y_{in}^2 \quad \text{for all positive } i,$$

where

$$\gamma = \left(-\frac{n}{\lambda} x_{1,n-1} - L_n^2 \right) / (L_n^1 - L_n^2).$$

Note that γ does not have to be between zero and one.

8. Set $n = n + 1$ and go to step 2.

V. THE DISTRIBUTION OF SOJOURN TIME

In this section, we provide a direct method of obtaining the distribution of sojourn time. First, we observe that U and T are independent random variables conditional on I. Thus,

$$P(V \leq t) = \sum_{i=1}^{\infty} [P(U \leq t \mid I = i) {*} P(T \leq t \mid I = i)] q_i, \tag{9}$$

where * is the convolution operator. It is easy to see that

$$P(U \leq t \mid I = i) = \frac{\left[P(I = i \mid U = 0) P(U = 0) + \int_{0+}^{t} P(I = i \mid U = u) dF(u) \right]}{q_i},$$

from which we obtain, after some routine algebra (for $t \geq 0$),

$$P(U \leq t \mid I = i) = \begin{cases} 1 & \text{if } \; i \leq m - 1 \\ 1 - P_m e^{-\mu t}/q_m & \text{if } \; i = m \\ \displaystyle\int_0^t \frac{(\mu u)^{i-m} \mu e^{-\mu u}}{(i - m)!} du & \text{if } \; i > m. \end{cases} \tag{10}$$

If we let $V(s)$ and $U_i(s)$ be the Laplace-Stieltjes transforms of $P(V \leq t)$ and $P(U \leq t \mid I = i)$, respectively, then from (9) we have

$$V(s) = \sum_{i=1}^{\infty} B_i(s) U_i(s) q_i. \tag{11}$$

It is possible to use the method of Section IV to solve eq. (1) for $B_i(s)$ directly. In fact, for $i \geq m$, $B_i(s)$ has the form

$$B_i(s) = \frac{\mu/m}{\mu/(m + s)} + B(s) \sigma_2^i(s), \tag{12}$$

where

$$\sigma_2(s) = \frac{[\lambda + \mu + s - \sqrt{(\lambda + \mu + s)^2 - 4\lambda\mu c}]}{2\lambda}.$$

After finding $B_i(s)$ for $i \geq m$, we can recursively compute $B_{m-1}(s)$, $\cdots$, $B_1(s)$ by using (1). It is clear that all of these quantities depend on the unknown function $B(s)$. Finally, one can find $B(s)$ to satisfy (2). This would have to be done by choosing two trial values of $B(s)$, evaluating two sets of $B_i(s)$, and using linear interpolation to satisfy (2). It is possible to shown that $V(s)$ simplifies [by using eqs. (10), (11), and (12)] to

$$\sum_{i=1}^{m-1} B_i(s)q_i + B_m(s)\,\frac{(\mu q_m + sP_{m-1})}{(\mu + s)} + \frac{\mu^2}{\mu + s}\left(\frac{\mu}{s(\mu + sm)} + \frac{B(s)\sigma_2^{m+1}(s)}{s + \mu[1 - \sigma_2(s)]}\right).$$

To find $P(V \le t)$, one would have to invert this transform numerically.

In the remainder of this section, we will show a method of calculating $P(T \le t \,|\, I = i)$ directly. Let us assume that at all times, events occur in this system according to a Poisson process at a rate of $\lambda + \text{Max}(\mu_1, \mu_2, \cdots, \mu_m)$. Let α be the maximum value of the arguments. Whenever the system is in state i, arrivals occur with probability $\lambda/(\lambda + \alpha)$, departures occur with probability $\mu_i/(\lambda + \alpha)$, and with probability $(\alpha - \mu_i)/(\lambda + \alpha)$ the system does not change its state. We assume that $\mu_0 = 0$. Let $\nu_{i,k}$ be the probability that a tagged customer departs on the kth event given that $I = i$. Then, $\nu_{i,k}$ can be recursively computed from

$$\nu_{i,k} = \frac{\lambda}{\lambda + \alpha}\,\nu_{i+1,k-1} + \frac{c_i\mu_i}{\lambda + \alpha}\,\nu_{i-1,k-1} + \frac{\alpha - \mu_i}{\lambda + \alpha}\,\nu_{i,k-1} \quad \text{for} \quad i \ge 1,\ k \ge 2$$

$$\nu_{i,1} = \frac{(1 - c_i)\mu_i}{(\lambda + \alpha)} \quad \text{for} \quad i \ge 1$$

and

$$\nu_{0,k} = 0 \quad \text{for} \quad k \ge 1.$$

Since events are occurring at a Poisson rate of $\lambda + \alpha$, departure of the tagged customer at the kth event means that its time in the service area is a Gamma random variable of order k and parameter $\lambda + \alpha$. Thus,

$$P(T \le t \,|\, I = i) = \sum_{k=1}^{\infty} \nu_{i,k} \int_0^t \frac{(\lambda + \alpha)e^{-(\lambda+\alpha)\chi}[(\lambda + \alpha)\chi]^{k-1}}{(k - 1)!}\,d\chi. \quad (13)$$

It is now possible to use (10), (11), and (13) to show that $V(s)$ is given by

$$\sum_{k=1}^{\infty}\left(\frac{\lambda + \alpha}{\lambda + \alpha + s}\right)^k \cdot \left[\sum_{i=1}^{m-1} \nu_{i,k}q_i + \nu_{m,k}\,\frac{(\mu q_m + sP_{m-1})}{(\mu + s)} + \sum_{i=m+1}^{\infty} \nu_{i,k}\left(\frac{\mu}{\mu + s}\right)^{i-m+1}\right].$$

VI. NUMERICAL RESULTS

In Table I we present numerical results for four examples, obtained by using the methods described earlier. In all examples, ρ is 0.9 or 0.95 and μ is 1. The first example is for the M/M/1 FCFS queue and the

Table I—Numerical results

Type of Queue	m	$\mu_i (i = 1, \ldots, m)$	λ	W	EV	EV^2	EV^3	EV^4
M/M/1 FCFS	1	1	0.90	10.0	10.0	200.0	6,000	240,000
			0.95	20.0	20.0	800.0	47,995	3,838,494
M/M/5 FCFS	5	i/m	0.90	12.625	12.625	278.75	8,756	358,121
			0.95	22.556	22.556	927.71	56,023	4,483,284
M/M/1 PS	50	1	0.90	10.0	10.0	358.25	26,817	3,252,699
			0.95	20.0	20.0	1322.0	157,901	27,094,857
General	10	$1 - [(m - i)/m]^2$	0.90	14.388	14.388	380.17	14,260	686,744
			0.95	24.356	24.356	1073.82	67,654	5,538,847

second for the M/M/5 FCFS queue. We approximate the M/M/1 PS queue in the third example by choosing a high value (50) for m. The fourth example is likely to be typical for a multiprogrammed computer system where μ_i is increasing in i but at a diminishing rate. We get a good correspondence between W (obtained from Little's formula) and EV in all cases. It is possible to verify that the higher moments in the first two examples are very accurate.

The computational procedure described earlier works well for up to fairly high values of m (about 100). However, for very large values of m (say 200), the method is prone to numerical difficulties. We feel that this happens because m determines the number of recursions and this, in turn, determines the extent to which errors are compounded. However, in real-life applications, one may not encounter values of m greater than 40, which makes our method suitable for these applications.

VII. ACKNOWLEDGMENTS

The authors are grateful to Dave Jagerman for his help in solving the difference equation of Section IV.

REFERENCES

1. B. Avi-Itzhak and D. P. Heyman, "Approximate Queueing Models for Multiprogramming Computer Systems," Oper. Res., *21*, No. 6 (November 1973), pp. 1212–30.
2. M. Reiser and S. S. Lavenberg, "Mean Value Analysis of Closed Multichain Queueing Networks," IBM Research Report RC7023, 1978.
3. A. A. Fredericks, "Approximations for Customer Viewed Delays in Multiprogrammed, Transaction Oriented Computer Systems," B.S.T.J., *59*, No. 9 (November 1980), pp. 1559–74.
4. A. G. Konheim and M. Reiser, "Finite Capacity Queueing Systems With Applications in Computer Modeling," SIAM J. Computing, 7, No. 2 (May 1978), pp. 210–29.
5. D. Mitra, *Waiting Time Distributions From Closed Queueing Network Models of*

Shared Processor Systems, Performance 81, F. J. Kylstra (Ed.), New York: North Holland, 1981, pp. 113–31.
6. S. Salza and S. S. Lavenberg, *Approximating Response Time Distributions in Closed Queueing Network Models of Computer Performance*, Performance 81, F. J. Kylstra (Ed.), New York: North Holland, 1981, pp. 133–45.
7. E. G. Coffman, R. R. Muntz, and H. Trotter, "Waiting Time Distributions for Processor Sharing Systems," JACM, *17*, No. 1 (January 1970), pp. 123–30.
8. G. Boole, *A Treatise on the Calculus of Finite Differences*, New York: G. E. Stechert, 1931.
9. D. L. Jagerman, unpublished work.

AUTHORS

Kiran M. Rege, B. Tech. (Electrical Engineering), 1977, I.I.T., Bombay; Ph.D. (Electrical Engineering), 1981, University of Hawaii; AT&T Bell Laboratories, 1982—. Mr. Rege spent 1984 on leave of absence, teaching at I.I.T. Bombay in the Department of Electrical Engineering. His technical work at AT&T Bell Laboratories includes modeling and analysis of switching, computer, and communication systems. His research interests include communication theory, queueing theory, and performance analysis of computer and communication systems.

Bhaskar Sengupta, B. Tech. (Electrical Engineering), 1965, I.I.T. Kharagpur, Eng. Sc.D. (Operations Research), 1976, Columbia University; AT&T Bell Laboratories, 1981—. Mr. Sengupta has worked at IBM and the Service Bureau Company and was an Assistant Professor at the State University of New York at Stony Brook. He was also a consultant to Turner Construction Company in New York. At AT&T Bell Laboratories he works on performance problems for communication, computer, and manufacturing systems.