We-Min Chow

Central Server Model for Multiprogrammed Computer Systems with Different Classes of Jobs

Abstract: A computer system can usually be interpreted as a closed network with two different types of servers. It is then possible to convert the network into a single server system with state-dependent arrivals. This paper investigates the stationary behavior of a single server queue with different classes of jobs. It is assumed that the input process has state-dependent exponential inter-arrival times and preemptions at the server are not allowed. The exact solution is obtained by finding the relationship between the time average probability distribution and the departure average probability distribution. The latter can be derived, based upon an imbedded Markov Chain.

Introduction

To study the stochastic behavior of computer systems, queuing networks are frequently used. In many cases, one can possibly describe these networks by two different types of servers, for example, a computer with a set of terminals, or a CPU with a finite number of I/O processors. It is therefore convenient to consider one type of server as the "input source" and the other as the "service facility." If the number of jobs in the network is fixed, the arrival rate to the service facility is in general a nonincreasing function of the queue length. On the other hand, if the jobs are not identical, distinguishable distributions for the successive service times may be expected, contingent upon the queuing discipline. Consequently, a queuing system with state-dependent arrival and service processes will be very helpful for practical use. Problems with a single class of jobs in this area have previously been studied as a generalized M/G/1 queue by means of renewal equations by Courtois and Georges [1].

In this paper, the stationary behavior of a single server queue is investigated. Based upon the so-called "Renewal Reward Theorem" [2], we show a simple way to treat the single-class-job problems and attempt to solve problems with two classes of jobs. The techniques used in this paper can be extended to the cases of more than two classes. Because of the tedious derivations, however, we do not discuss these problems.

We assume that the inter-arrival times are independently exponentially distributed (not necessary identically), and preemptions at the service facility are not allowed.

For the M/G/1 FIFO queue, it is well known that A_j, the proportion of departures that have j jobs left behind, is equal to the proportion of time there are j jobs in the system, P_j. In the case of state-dependent arrivals, however, this property is no longer true. A frequently used technique is to compute $\{A_j\}$ first, and to evaluate $\{P_j\}$ by expressing each P_j as a function of $\{A_j\}$ [1, 3, 4]. This approach is adopted here.

Notation

M^i = the total number of type i jobs, or the waiting room capacity (including the one in service, if any) for type i jobs [5]

i-arrival = an arrival that belongs to type i

i-departure = a departure that belongs to type i

$\lambda_j^{\,i}$ = the arrival rate of i-arrival when there are j type i jobs in the system

$x_n^{\,i}$ = the number of type i jobs remaining in the system when the nth departure occurs

$X_n = (x_n^{\,1}, x_n^{\,2})$

S^i = the service time of a type i job

S_n = the nth service time

$G^i(t|j, k) = P[S^i \le t | X_n = (j, k)]$

$d^i(X_n) = P[S_{n+1} = S^i | X_n]$

$T^i(u|j)$ = the $(u - j)$th i-arrival time given that there are j type i jobs in the system at $t = 0$

$T(u, v|j, k)$ = the first passage time to the state (u, v), given that $X_n = (j, k)$ at the last departure epoch

$F_u^{\,i}(t|j) = P[T^i(u|j) \le t]$

$F_{uv}(t|j, k) = P[T(u, v|j, k) \le t]$

$A_{jk} = P[X_\infty = (j, k)]$

= the departure average probability of the state (j, k)

P_{jk} = the time average probability that the system is in state (j, k)

$$U(S) = \begin{cases} 1, & \text{if } S \geq 0, \\ 0, & \text{if } S < 0. \end{cases}$$

Renewal reward process

Before our discussions, we state the Renewal Reward Theorem as follows:

Let $\{N(t),\ t \geq 0\}$ be a renewal process with inter-arrival times $T_1, T_2, \cdots$. Suppose that a reward Y_n is earned during the nth inter-arrival time, and the pairs (T_n, Y_n) are independently and identically distributed. Define $Y(t) = \Sigma_{n=1}^{N(t)} Y_n$, if $\mathrm{E}[Y_n]$ and $\mathrm{E}[T_n]$ are both finite, then with probability one, we have

$$(i)\ \frac{Y(t)}{t} \rightarrow \frac{\mathrm{E}[Y]}{\mathrm{E}[T]}, \text{ as } t \rightarrow \infty$$

$$(ii)\ \frac{\mathrm{E}[Y(t)]}{t} \rightarrow \frac{\mathrm{E}[Y]}{\mathrm{E}[T]}, \text{ as } t \rightarrow \infty$$

The proof of the theorem can be found in [2].

According to the theorem, it is clear that the time average probability of a certain state can be obtained by taking the ratio of the expected length of time that the system is in the state during a recurrent cycle to the mean length of the cycle.

Customer averaged probability

For an M/G/1 FIFO queue with state-dependent exponential inter-arrival times, a Markov chain is usually defined by looking at the departure epochs. Let X_n be the number of jobs remaining in the system just after the nth departure epoch. Then the stationary distribution $A_j = P[X_\infty = j]$ can be obtained by using

$$X_{n+1} = X_n - \delta_n + Y_{n+1}, \tag{1}$$

where Y_{n+1} is the number of arrivals during the $(n+1)$th service time, and $\delta_n = 1$, if $X_n > 0$; $\delta_n = 0$, otherwise.

For two classes of jobs, X_n is expressed by a vector, i.e., $X_n = (x_n^1, x_n^2)$, where x_n^i is the number of type i jobs left behind by the nth departure. Then a Markov chain is given by

$$X_{n+1} = X_n - \Delta_n + Y_{n+1}, \tag{2}$$

where Y_{n+1} = the arrival pattern, and its ith element indicates the number of i arrivals during S_{n+1};

$$\Delta_n = \begin{cases} \text{null vector, if } X_n = (0, 0); \\ \text{the } i\text{th unit vector, if } S_{n+1} = S^i \text{ and } x_n^i > 0. \end{cases}$$

The distribution of the $(n+1)$th service time, S_{n+1}, depends upon the state, X_n, and the queuing discipline. One may define $S_{n+1} = S^i$, with probability $d^i(X_n)$ and $d^1(X_n) + d^2(X_n) = 1$. If $X_n = (0, 0)$, then $d^i(X_n)$ is interpreted as the probability that an i-arrival follows the idle period immediately. For non-preemptive priority queues, then

$d^1(X_n) = 1$, if $x_n^1 > 0$;

$d^2(X_n) = 1$, if $x_n^2 > 0$, $x_n^1 = 0$.

Some other examples can be set by letting $d^i(X_n)$ be constant (i.e., independent of the state X_n), or having $d^i(X_n) = x_n^i/(x_n^1 + x_n^2)$, $i = 1, 2$.

The transition probabilities of the Markov chain defined by (2) is then given by

$$\begin{aligned} P[X_{n+1} &= (u, v) | X_n = (j, k)] \\ &= P[Y_{n+1} = (u - j + 1, v - k) | S_{n+1} = S^1, X_n \\ &\qquad = (j, k)] d^1(j, k) \\ &\quad + P[Y_{n+1} = (u - j, v - k + 1) | S_{n+1} = S^2, X_n \\ &\qquad = (j, k)] d^2(j, k). \end{aligned} \tag{3}$$

Although we have implicity assumed that S^i is not dependent upon X_n, it is clear that one can easily generalize this assumption.

The conditional distribution of Y_n can be obtained by using the convolutions of a set of independent exponential distributions. If the arrival processes of different types of jobs are independent, i.e., if the inter-arrival time of type i jobs depends upon the queue size of their own kind only, then

$$\begin{aligned} P[Y_{n+1} &= (u, v) | S_{n+1} = S^i, X_n = (j, k)], \\ &= \int_0^\infty P[T^1(u|j) \leq s \leq T^1(u+1|j), T^2(v|k) \\ &\qquad \leq s \leq T^2(v+1|k)] dG^i(s|j, k), \\ &= \int_0^\infty [F_u^1(s|j) - F_{u+1}^1(s|j)][F_v^2(s|k) \\ &\qquad - F_{v+1}^2(s|k)] dG^i(s|j, k), \end{aligned} \tag{4}$$

where

$$\begin{aligned} F_u^i(s|j) &= P[T^i(u|j) \leq s], \\ &= \begin{cases} \int_0^t [1 - e^{-\lambda_{u-1}^i (t-s)}] dF_{u-1}^i(s|j), & \text{if } u > j \\ U(s), & \text{if } u = j. \end{cases} \end{aligned} \tag{5}$$

For the machine-repairman problem (e.g., a computer serves $M^1 + M^2$ terminals), λ_u^i is a linear function of u, i.e., $\lambda_u^i = \lambda^i(M^i - u)$, $u = 0, 1, \cdots, M^i$. Thus,

$$F_u^{\,i}(s|j) = \sum_{k=u-j}^{M^i} \binom{M^i}{k} (1 - e^{-\lambda^i s})^k (e^{-\lambda^i s})^{M^i - k}.$$

For another example, we may have a single CPU with $(N^1 + N^2)$ I/O processors servicing $M^1 + M^2$ jobs (each type of job forms a single I/O queue), then $\lambda_u^{\,i}$ is a piecewise linear function of u, that is, for $N^i \ge M^i$, $i = 1, 2$;

$$\lambda_u^{\,i} = \begin{cases} \lambda^i M^i & , \text{if } N^i - u \ge M^i; \\ \lambda^i (N^i - u), & \text{if } N^i - u \le M^i. \end{cases}$$

It can be seen that if $N^i - u \le M^i$, the number of type i jobs in I/O processors is less than the number of I/O processors; therefore, the arrival rate to the CPU is linearly decreasing in u. The distribution of the number of i-arrivals during each CPU service time can be computed by using the standard techniques derived for GI/M/R queuing systems [3].

Once the transition probabilities are obtained, we use the equilibrium equations to evaluate the stationary probability distribution. Thus, we have

$$A_{uv} = \sum_{(j,k)} P[X_{n+1} = (u, v) | X_n = (j, k)] A_{jk},$$

and

$$\sum_{(j,\, k)} A_{jk} = 1, \tag{6}$$

$j = 0, 1, \cdots, M^1$; $k = 0, 1, \cdots, M^2$, and $j + \mathrm{k} < M^1 + M^2$.

In the case that $\lambda_j^{\,1} = \lambda^1$, $\lambda_k^{\,2} = \lambda^2$ $\forall j, k$, then

$$A_{uv} = a_{uv}^{\;1} A_{00} \frac{\lambda^1}{\lambda^1 + \lambda^2} + a_{uv}^{\;2} A_{00} \frac{\lambda^2}{\lambda^1 + \lambda^2}$$
$$+ \sum_{j=1}^{u+1} \sum_{k=0}^{v} a^1_{u-j+1, v-k} A_{jk} d^1(j, k) +$$
$$+ \sum_{j=0}^{u} \sum_{k=1}^{v+1} a^2_{u-j, v-k+1} A_{jk} d^2(j, k),$$

where

$$a_{uv}^{\;i} = \int_0^\infty \frac{(\lambda_1 t)^u}{u!} e^{-\lambda_1 t} \frac{(\lambda_2 t)^v}{v!} e^{-\lambda_2 t} dG^i(t).$$

Define $\tilde{A}(z_1, z_2) = \sum_u \sum_v z_1^{\,u} z_2^{\,v} A_{uv}$,

$$\tilde{D}^i(z_1, z_2) = \sum_u \sum_v z_1^{\,u} z_2^{\,v} A_{uv} d^i(u, v), \text{ and}$$

$$\tilde{G}^i(s) = \int_0^\infty s^{-st} dG^i(t).$$

It can be shown that

$$\tilde{A}(z_1, z_2) = \Big[\frac{\lambda^1}{\lambda^1 + \lambda^2} \tilde{G}^1[\lambda^1(1 - z_1) + \lambda^2(1 - z_2)]$$
$$+ \frac{\lambda^2}{\lambda^1 + \lambda^2} \tilde{G}^2[\lambda^1(1 - z_1) + \lambda^2(1 - z_2)]\Big](1 - \rho)$$
$$+ \tilde{G}^1[\lambda^1(1 - z_1) + \lambda^2(1 - z_2)] \frac{1}{z_1} [\tilde{D}^1(z_1, z_2)$$
$$- \tilde{D}^1(0, z_2)] + \tilde{G}^2[\lambda^1(1 - z_1) + \lambda^2(1 - z_2)]$$
$$\times \frac{1}{z_2} [\tilde{D}^2(z_1, z_2) - \tilde{D}^2(z_1, 0)],$$

where

$$\rho = 1 - \lambda^1 \mathrm{E}[S^1] - \lambda^2 \mathrm{E}[S^2].$$

A single class of jobs

The Renewal Reward Theorem is now used to obtain the time average probability distribution. First, we study the problem with a single class of jobs. We say that a renewal occurs if j jobs are left behind in the system by a departure. Let

C_j = the time interval between successive renewals or the recurrent time.
R_j = the number of departures during C_j or the number of jobs completed during C_j.
D = the inter-departure time.

Because among R_j departures during the recurrent time, exactly one departure can have j jobs left behind, it must be the case, by using the theorem, that

$$A_j = \frac{1}{\mathrm{E}[R_j]}. \tag{7}$$

For $C_j = \Sigma_{k=1}^{R_j} D_k$, then

$$\mathrm{E}[C_j] = \mathrm{E}[D]\mathrm{E}[R_j]. \tag{8}$$

Note that

$$D = \begin{cases} S_n & \text{if } X_n > 0, \\ S_n + I & \text{if } X_n = 0, \end{cases}$$

where I is the idle period of the service facility.

$$\mathrm{E}[D] = \mathrm{E}[S_n] + \frac{1}{\lambda_0} A_0. \tag{9}$$

Equations (7), (8) and (9) imply that

$$\mathrm{E}[C_j] = \left(\mathrm{E}[S_n] + \frac{A_0}{\lambda_0}\right) \frac{1}{A_j}.$$

We still have to evaluate the expected length of time that the system is in the state j during a recurrent cycle, C_j. This can be done by introducing a "dual" system [6]. We say that a pair of queuing systems are dual systems;

if there is an arrival (or a departure) in one system, then a departure (or an arrival) occurs accordingly in the other. Clearly, the inter-arrival time and the service time are interchanged in the dual queue. Moreover, if one system is in the state j, then we let the corresponding state of its dual system be $M-j$ so that a transition from the state $M-j-1$ to the state $M-j$ becomes a renewal (see Fig. 1). As to our problem, it is not difficult to see that the dual system has a state-dependent exponential service time, and it is operated under LIFO rule with preemptive resume discipline. When this dual system is in the state $M-j$ and a job, namely J, is preempted by a new arrival, then correspondingly a departure occurs in the original system (which is now in the state $j-1$). As soon as the number of jobs in the dual system is reduced to $M-j$, job J is resumed and its service time is still exponentially distributed with a rate λ_j (due to the memoryless property). This is then equivalent to saying that the original queue is in the state j and the next inter-arrival time is exponentially distributed with the rate λ_j. Note that the dual queue can be in the state $M-j$ (hence the original system is in the state j) if and only if job J is being serviced and exactly one such job exists during each cycle. Since the work is conserved, the expected length of time that the original system is in the state j is equal to the mean service time of job J in the dual queue, that is, $1/\lambda_j$.

Applying the theorem, we have

$$P_j = \lambda_i^{-1}/\mathrm{E}[C_j]$$

$$= \left(\mathrm{E}[S_n] + \frac{A_0}{\lambda_0}\right)^{-1} \frac{A_j}{\lambda_j} \;\forall j < M. \tag{10}$$

This result has been obtained in [1].

Equation (10) indicates that P_j is proportional to A_j/λ_j $\forall j < M$. For a homogeneous arrival process (i.e., $\lambda_j = \lambda$, $\forall j$) and $M = \infty$, $A_0 = 1 - \lambda\mathrm{E}[S_n]$. Consequently, (10) becomes $P_j = A_j \;\forall j$, the well known property of M/G/1 FIFO queues.

Two classes of jobs

It is a little complicated to evaluate the probability distribution of $\{X_n = (j, k)\}$. Although the mean recurrent time can be obtained in a similar way by letting a departure that finds the state (j, k) be a renewal, it is not easy to compute the mean length of time that the system is in the state (j, k) during this recurrent cycle. In the following, we consider the interdeparture time as a cycle. (This idea was mentioned in [1].) Let

$T(u, v|j, k)$ = the first passage time from the event $\{X_n = (j, k)\}$ to the state (u, v);

$T^i(k|j)$ = the $(k-j)$th arrival time of an i-arrival given that there are j type i jobs in the system initially.

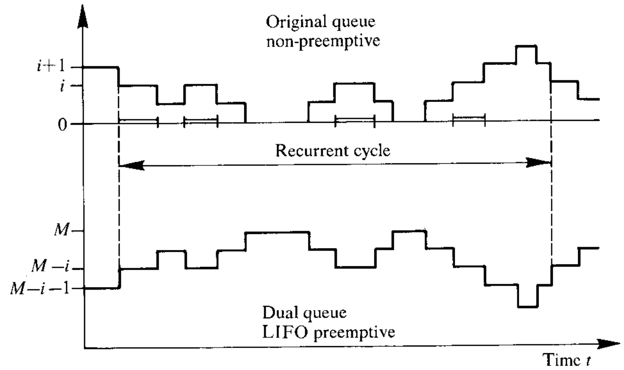

Figure 1 Comparison of states, plotted vs time.

The necessary and sufficient condition that the system will reach the state (u, v) from the state (j, k) is

$$Z = \{T^1(u|j) \le T^2(v+1|k),\; T^2(v|k) \le T^1(u+1|j)\}.$$

Thus,

$$F_{uv}(t|j, k) = P[T(u, v|j, k) \le t]$$

$$= P[\max\,(T^1(u|j), T^2(v|k)) \le t, Z].$$

Define

$$B = \{\max\,(T^1(u|j), T^2(v|k)) \le t\},$$

$$B_1 = B \cap Z,$$

$B_2 = B \cap \{T^1(u+1|j) < T^2(v|k)\}$, and

$$B_3 = B \cap \{T^2(v+1|k) < T^2(u|j)\},$$

Since $T^2(v+1|k) > T^2(v|k)$ and $T^1(u+1|j) > T^1(u|j)$, it is clear that

$B = B_1 \cup B_2 \cup B_3$, and

$B_r \cap B_m = \varnothing$, for $r \ne m$, $\quad r, m = 1, 2, 3.$

Consequently,

$$F_{uv}(t|j, k) = P[B_1]$$

$$= P[B] - P[B_2] - P[B_3]$$

$$= P[T^1(u|j) \le t,\; T^2(v|k) \le t]$$

$$- P[T^1(u+1|j) < T^2(v|k) \le t]$$

$$- P[T^2(v+1|k) < T^1(u|j) \le t];$$

$$F_{uv}(t|j, k) = F_u^{\,1}(t|j)\, F_v^{\,2}(t|k) - \int_0^t F^1_{u+1}(s|j)\, dF_v^{\,2}(s|k)$$

$$- \int_0^t F^2_{v+1}(s|k)\, dF_u^{\,1}(s|j). \tag{12}$$

The above equation is reduced to the case of a single class if we let the type-2 arrival rate $\lambda_j^2 = 0$. Then $T^2(j|0) = \infty$, $\forall j > 0$, and

$$F_k^{\,2}(s|0) = \begin{cases} 0 & \text{if } s > 0 \text{ or } k > 0; \\ 1 & \text{if } k = 0. \end{cases}$$

In the case that $\lambda_j^{\,1} = \lambda^1$, $\lambda_k^{\,2} = \lambda^2$, the events Z and B are mutually independent. This can be shown as follows.

$$\int_0^\infty e^{-st} F_u^{\,1}(t|j)\ dF_v^{\,2}(t|k)$$

$$= \int_0^\infty e^{-st}\left\{\sum_{m=u-j}^{\infty} \frac{(\lambda^1 t)^m}{m!} e^{-\lambda^1 t}\right\} \frac{(\lambda^2 t)^{v-k-1}}{(v-k-1)!} e^{-\lambda^2 t}\lambda^2 dt$$

$$= \sum_{m=u-j}^{\infty} \binom{m+v-k-1}{m} \left(\frac{\lambda^1}{\lambda^1+\lambda^2+s}\right)^m \left(\frac{\lambda^2}{\lambda^1+\lambda^2+s}\right)^{v-k}.$$

Using the above equation and (12), we have

$$\int_0^\infty e^{-st} dF_{uv}$$

$$= \binom{u-j+v-k}{u-j} \left(\frac{\lambda^1}{\lambda^1+\lambda^2+s}\right)^{u-j} \left(\frac{\lambda^2}{\lambda^1+\lambda^2+s}\right)^{v-k},$$

which implies that

$$F_{uv}(t|j,k)$$

$$= \binom{u-j+v-k}{u-j} \left(\frac{\lambda^1}{\lambda^1+\lambda^2}\right)^{u-j} \left(\frac{\lambda^2}{\lambda^1+\lambda^2}\right)^{v-k}$$

$$\times \Gamma(t|u-j+v-k, \lambda^1+\lambda^2), \tag{13}$$

where $\Gamma(t|n, \lambda)$ is the distribution function of a gamma random variable with a mean λ^{-n}.

We now come back to our problem. Let

θ_{uv} = the duration of time that the system is in the state (u, v) during an inter-departure time.

$R_{n+1}(r)$ = the remaining service time of the $(n+1)$th job given that the job has received r units of service time.

$\tau_k^{\,i}$ = the inter-arrival time of a type i job, given that there are k such jobs in the system.

$$\mathrm{E}[\theta_{uv}|X_n = (j,k)] = \int_0^\infty \mathrm{E}[\theta_{uv}|T(u,v|j,k) = r]$$

$$\times\, dF_{uv}(r|j,k) \tag{14}$$

$$\mathrm{E}[\theta_{uv}|T(u,v|j,k) = r]$$

$$= \mathrm{E}[\min\,(R_{n+1}(r), \tau_u^{\,1}, \tau_v^{\,2})|r \le S_{n+1}]\ P[r \le S_{n+1}],$$

because for $r > S_{n+1}$, the system will never reach the state (u, v) and $\theta_{uv} = 0$.

Define $G(t|j,k) = \Sigma_{i=1}^2\ G^i(t|j,k)\ d^i(j,k)$, and $\tau = \min\,(\tau_u^{\,1}, \tau_v^{\,2})$; then

$$P[r \le S_{n+1}] = 1 - G(r|j,k),$$

$$P[\tau > t] = e^{-(\lambda_u^{\,1}+\lambda_v^{\,2})t}, \text{ and}$$

$$P[R_{n+1}(r) > t] = \frac{1-G(t+r|j,k)}{1-G(r|j,k)}.$$

Thus,

$$\mathrm{E}[\theta_{ur}|T(u,v|j,k) = r] = \int_0^\infty e^{-\lambda_{uv}t}\,[1-G(t+r|j,k)]\,dt,$$

where

$$\lambda_{uv} = \lambda_u^{\,1} + \lambda_v^{\,2}.$$

Substituting this equations into (14) and interchanging the integrations, we have

$$\mathrm{E}[\theta_{uv}|X_n = (j,k)]$$

$$= \int_0^\infty \int_0^\infty e^{-\lambda_{uv}t}\,[1-G(t+r|j,k)]$$

$$\times\, dF_{uv}(r|j,k)\ dt$$

$$= \frac{1}{\lambda_{uv}} \int_0^\infty [1-G(y|j,k)] \int_0^y \lambda_{uv} e^{-\lambda_{uv}(y-r)}$$

$$\times\, dF_{uv}(r|j,k)\ dy$$

$$= \frac{1}{\lambda_{uv}} \int_0^\infty [1-G(y|j,k)]\ d\left\{\int_0^y [1-e^{-\lambda_{uv}(y-r)}]\right.$$

$$\left.\times\, dF_{uv}(r|j,k)\right\}$$

$$= \frac{1}{\lambda_{uv}} \int_0^\infty \left\{\int_0^y [1-e^{-\lambda_{uv}(y-r)}]\ dF_{ur}(r|j,k)\right\}$$

$$\times\, dG(y|j,k). \tag{15}$$

$$j-1 \le u,\ k-1 \le v, \text{ and } j+l-1 \le u+v.$$

The above equation shows that $\mathrm{E}[\theta_{uv}|X_n = (j,k)]$ is equal to the product of the expectation of inter-arrival time for the state (j,k) and $P[T(u,v|j,k) + \tau \le S_{n+1}|X_n = (j,k)]$.

If $\lambda_j^{\,1} = \lambda^1$ and $\lambda_k^{\,2} = \lambda^2$, it can be shown that

$$\int_0^y [1-e^{-\lambda_{uv}(y-r)}]\ dF_{uv}(r|j,k)$$

$$= \left\{\binom{u+v-j-k}{u-j} \left(\frac{\lambda^1}{\lambda^1+\lambda^2}\right)^{u-j} \left(\frac{\lambda^2}{\lambda^1+\lambda^2}\right)^{v-k}\right\}$$

$$\times\, \Gamma\,(y|u+v-j-k+1, \lambda^1+\lambda^2). \tag{16}$$

For an exponential server with a rate μ, i.e., $G(t|j,k) = 1 - e^{-\mu t}$, Eqs. (15) and (16) give

$$\mathrm{E}[\theta_{uv}|X_n = (j,k)] = \binom{u+v-j-k}{u-j} \left(\frac{\lambda^1}{\lambda^1+\lambda^2+\mu}\right)^{u-j}$$

$$\times \left(\frac{\lambda^2}{\lambda^1+\lambda^2+\mu}\right)^{v-k} \frac{1}{\lambda^1+\lambda^2+\mu}.$$

Remove the condition that $\{X_n = (j, k)\}$ from (15); then

$$E[\theta_{uv}] = \sum_{j=0}^{u} \sum_{k=0}^{v} E[\theta_{uv} | X_n = (j, k)] A_{jk} \tag{17}$$

can be evaluated.

The mean inter-departure time is obtained by

$$\begin{aligned} E[D] &= \sum_{(j,k)} E[D | X_n = (j, k)] A_{jk} \\ &= \left(\frac{1}{\lambda_{00}} + \frac{\lambda_0^{\,1}}{\lambda_{00}} E[S^1] + \frac{\lambda_0^{\,2}}{\lambda_{00}} E[S^2]\right) A_{00} \\ &\quad + \sum_{(j,k)} \{E[S^1] d^1(j, k) + E[S^2] d^2(j, k)\} A_{jk}. \end{aligned} \tag{18}$$

Finally, the time average stationary probability,

$$P_{uv} = E[\theta_{uv}] / E[D], \ \forall\ (u, v) \ni u + v < M^1 + M^2. \tag{19}$$

The probability that the service facility is idle can also be obtained directly by letting the starting point of each busy period be a renewal. Define

R = the number of jobs completed during each recurrent cycle,
B = the length of a busy period, and
I = the length of an idle period.

Since

$$A_{00} = \frac{1}{E[R]},$$

$$E[I] + E[B] = E[R] E[D], \text{ and}$$

$$E[I] = \frac{1}{\lambda_{00}},$$

it follows that,

$$\begin{aligned} P_{00} &= \frac{E[I]}{E[I] + E[B]} \\ &= \frac{A_{00}}{\lambda_{00} E[D]}, \end{aligned}$$

where $E[D]$ is obtained in (18).

Therefore, the utilization of the service facility is given by

$$1 - P_{00} = \frac{\lambda_{00} E[D] - A_{00}}{\lambda_{00} E[D]}.$$

If only the marginal distribution is of interest, the arguement in the last section is valid. Let the event that an i-departure has u type i jobs left behind be a renewal, and define

$A_u^{\,i} = \lim_{n \to \infty} P[X_n^{\,i} = u, S_n = S^i]$, and

$P_u^{\,i}$ = the time average probability that there are u type i jobs in the system, for $i = 1, 2$.

Then

$$\begin{aligned} A_u^{\,1} &= \sum_v P[X_{n+1} = (u, v), S_{n+1} = S^1] \\ &= \sum_{(j,k)} \sum_v P[X_{n+1} = (u, v) | S_{n+1} = S^1, X_n = (j, k)] \\ &\quad \times P[S_{n+1} = S^1 | X_n = (j, k)] P[X_n = (j, k)]. \end{aligned}$$

For $n \to \infty$, we have

$$A_u^{\,1} = \sum_{(j,k)} \sum_v P[X_{n+1} = (u, v) | S_{n+1} = S^1, X_n = (j, k)] \times d^1(j, k) A_{jk},$$

and

$$P_u^{\,1} = \frac{1}{\lambda_u^{\,1}} \frac{A_u^{\,1}}{E[D]};$$

the conditional probability of X_{n+1} is given by (4), and $E[D]$ by (18).

For type 2 jobs, the similar results hold by replacing the superscripts.

Conclusions

When a computer system is considered as a single central server queue, the stationary stochastic behavior can be evaluated if the inter-arrival times are independent, exponential random variables. The results in this paper show the relationship between the departure average distribution and the time average distribution. For a single class of jobs, a simple relation is obtained. If there are two different classes of jobs, the distribution $\{P_{jk}\}$ is given as a function of $\{A_{jk}\}$ by (15)-(19). It seems not difficult to extend these results to the cases of more than two classes.

If both of the arrival processes are homogeneous (i.e., $\lambda_j^{\,1} = \lambda^1$ and $\lambda_k^{\,2} = \lambda^2$), then the event Z and the first passage time $T(r, m | j, k)$, $\forall\ r \geq u$, $m \geq v$, are independent. This fact is shown by equations (13) and (16). Unfortuantely, this is not true in the case of state-dependent arrival processes. It can be seen that if we let $u = 1$ and $v = 0$, $T(1, 1, | 0, 0)$ does depend upon Z. Thus, from a practical point of view, the major difficulty is in computing the transition probabilities $\{P[X_{n+1} = (u, v) | X_n = (j, k)]\}$ or equivalently $F_u^{\,i}(t | j)$.

If the arrival rate is a linear function of its own state variable for each class of jobs, one can use a binomial law to evaluate these probabilities. For arbitrary functions, however, much computational effort may be required. Consequently, an efficient method is most desirable.

Acknowledgment

I thank L. Woo for his valuable contribution to this paper.

References

1. P. J. Courtois and J. Georges, "On a Single-Server Finite Queueing Model with State-Dependent Arrival and Service Processes," *Operations Research* **19**, 424 (1971).
2. S. M. Ross, *Applied Probability Models with Optimization Applications*, Holden-Day, Inc., San Francisco, CA., 1970.
3. R. Syski, *Introduction to Congestion Theory in Telephone Systems*, Oliver and Boyd Ltd., Edinburgh, Scotland, 1960.
4. L. Takacs, *Introduction to the Theory of Queues*, Oxford University Press, London, England, 1962.
5. M. Reiser and H. Kobayashi, "The Effects of Service Time Distributions on System Performance in a Multi-programmed Computer System Model," *Research Report RC 4671*, IBM Thomas J. Watson Research Center, Yorktown Heights, New York, November 1973.
6. J. F. C. Kingman, "The Heavy Traffic Approximation in the Theory of Queues," Chapter 6 in *Proceedings of Symposium on Congestion Theory*, edited by W. L. Smith and W. E. Wilkson, University of North Carolina, 1965.

Received October 6, 1974

The author is located at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, 10598.