RESPONSE TIME DISTRIBUTIONS FOR A MULTI-CLASS QUEUE WITH FEEDBACK

Simon S. Lam and A. Udaya Shankar

Department of Computer Sciences
University of Texas at Austin
Austin, Texas 78712

Abstract

A single server queue with feedback and multiple customer classes is analyzed. Arrival processes are independent Poisson processes. Each round of service is exponentially distributed. After receiving a round of service, a customer may depart or rejoin the end of the queue for more service. The number of rounds of service required by a customer is a random variable with a general distribution. Our main contribution is characterization of response time distributions for the customer classes. Our results generalize in some respects previous analyses of processor-sharing models. They also represent initial efforts to understand response time behavior along paths with loops in local balanced queueing networks.

1. INTRODUCTION

Many service facilities can be modeled as a feedback queue such as shown in Figure 1. Of interest in this paper is a single-server queue with infinite waiting room and Q types of customers. The arrival process of type q customers is an independent Poisson process (q = 1, 2, ... Q). Each new arrival joins the end of the queue. The customer at the head of the queue receives from the server a round of service, which is an independent exponentially distributed random variable with mean $1/\mu$ seconds. After receiving a round of service, a customer may depart or rejoin the end of the queue for more service. The number of rounds of service required by a type q customer is a random variable with a general probability distribution $\{a_r^{(q)}, r = 1, 2, \ldots, R\}$ where $a_r^{(q)}$ is the probability of a type q customer requiring exactly r rounds of service.

The queue length distribution of the above model is readily available since the feedback queue described is an open queueing network satisfying local balance [1]. The contribution of this paper is to characterize response time distributions of the different types of customers; specifically, we solved for the conditional response time distributions of an arbitrary customer requiring r rounds of service for r = 1, 2, ..., R.

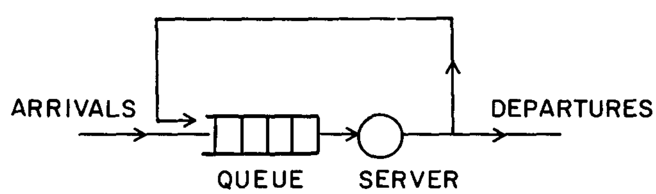

Fig. 1. A feedback queue model.

Relationship to prior work

Our feedback queue model is like a time-sharing model with exponentially distributed service "quantums." Time-sharing models were first studied by Kleinrock [2] who solved for the mean response time of a customer conditioning on his (total) service requirement. He considered two cases: (a) constant quantum size Δ, and (b) the limiting case of $\Delta \to 0$ called processor-sharing. Customers are assumed to arrive according to a Poisson process. In case (a), the number of service quantums required by a customer is geometrically distributed. In case (b), the service requirements are characterized by an exponential distribution. (This is called the processor-sharing M/M/1 queue.) Kleinrock's conditional mean response time result was later shown to hold for a processor-sharing M/G/1 queue (i.e. service requirements characterized by a general distribution) as well by Sakata, Noguchi and Oizumi [3]. Higher order response time statistics are much harder to get. The response time distribution for the processor-sharing M/M/1 queue was obtained by Coffman, Muntz, and Trotter [4]. The response time distribution for the constant quantum size case was obtained by Muntz [5] assuming exponentially distributed service requirements.

Our feedback queue model is different from the time-sharing models in several respects. A round of service in our model, corresponding to a service quantum in time-sharing models, is exponentially distributed. Our model can be used, however, to approximate processor-sharing by making $1/\mu$ very small relative to the mean service requirement.

Aside from the quantum size assumption, our model is more general than those of [4,5] in two respects: (i) multiple types of customers, and (ii) the number of rounds of service for each customer type has a general probability distribution. Specifically, distributions of service requirements that are admissible in our model are those with moment generating functions of the form

$$B_q^*(s) = \sum_{r=1}^{R} a_r^{(q)} \left(\frac{\mu}{s+\mu}\right)^r \tag{1}$$

Our model is also different from the feedback queue model of Takács [6]. In his model, each round of service can have a general distribution. However, he considered one type of customers only and the number of rounds of service required by a customer is geometrically distributed; in other words, after each round of service, a customer always departs with probability (1-p) and rejoins the end of the queue with probability p (memoryless behavior).

The original motivation of this work stems from our efforts to characterize the response time in a network of queues. For a network of FCFS queues that satisfies local balance, J. Wong [7] found the response time distribution of customers traversing loop-free paths. Our results in this paper represent efforts to understand the response time behavior along paths with loops in the simplest form of queueing networks satisfying local balance.

Assumptions and definitions

Consider the following example of 2 types of customers. Type 1 customers arrive according to a Poisson process with rate α_1 customers per second. The number of rounds of service required by a type 1 customer has the probability distribution

$$a_r^{(1)} = \begin{cases} 1/100 & r = 1, 2, \ldots, 100 \\ 0 & \text{otherwise} \end{cases}$$

Type 2 customers arrive according to a Poisson process with rate α_2 customers per second. The number of rounds of service required by a type 2 customer has the probability distribution

$$a_r^{(2)} = \begin{cases} 1/10 & r = 1, 2, \ldots, 10 \\ 0 & \text{otherwise} \end{cases}$$

Using the properties of Poisson processes, the above model is equivalent to the following model with 100 types of customers. Type r customers ($r = 1, 2, \ldots, 100$) require exactly r rounds of service and arrive according to a Poisson process with rate

$$\gamma_r = \begin{cases} 0.01\alpha_1 + 0.1\alpha_2 & r = 1, 2, \ldots, 10 \\ 0.01\alpha_1 & r = 11, 12, \ldots, 100 \\ 0 & \text{otherwise} \end{cases}$$

We shall, without any loss of generality, consider the following model. There are R types of customers. The arrival process of the r^{th} type is Poisson at rate γ_r customers per second. A type r customer requires exactly r rounds of service. It should be obvious that if we can derive response time distributions for this model, response time distributions for any model with Q customer types and service time requirements characterized by Eq. (1) can be easily obtained.

Let t_r be the response time of attaining exactly r rounds of service; $r = 1, 2, \ldots R$ and obviously $t_0 = 0$. We shall solve for its moment generating function

$$T_r^*(s) = E[e^{-st_r}]$$

where $E[\cdot]$ denotes the expectation of the function of random variable(s) inside the brackets.

We shall only consider steady-state results. For a single-server queue, stationarity is assured if the traffic intensity $\rho < 1$ where $\rho = \sum_{r=1}^{R} \gamma_r(r/\mu)$; see Cohen [8].

Customers in the queue are differentiated into R different classes; class k consists of all those customers in the queue with exactly k more rounds of service to go, where $k = 1, 2, \ldots, R$.

Let us follow the progress of a "tagged" customer and introduce some more notation. Upon his initial arrival, the tagged customer finds n_k class k customers in the queue ($k = 1, 2, \ldots, R$). The system state thus found at an arrival instant is denoted by $\underline{n} = (n_1, n_2, \ldots, n_R)$ and is described by the moment generating function

$$P^*(\underline{z}) = E[z_1^{n_1} z_2^{n_2} \ldots. z_R^{n_R}]$$

where $\underline{z}$ is the shorthand notation for $(z_1, z_2, \ldots, z_R)$.

At the end of the tagged customer's r^{th} round of service (given that he requires at least r rounds), let the system state at that instant be denoted by $\underline{m}^{(r)} = (m_1^{(r)}, m_2^{(r)}, \ldots, m_R^{(r)})$ where $m_k^{(r)}$ is the number of customers who have exactly k more rounds of service to go. Define $M^{(r)} = \sum_{k=1}^{R} m_k^{(r)}$.

In order to characterize $T_r^*(s)$, we shall need to first characterize the joint distribution of t_r and $\underline{m}^{(r)}$, which is described by

$$U_r^*(s, z) = E[e^{-st_r} z_1^{m_1^{(r)}} z_2^{m_2^{(r)}} \ldots . z_R^{m_R^{(r)}}]$$

Summary of results

We derived a recursive equation relating $U_{r+1}^*(s, \underline{z})$ to $U_r^*(s, \underline{z})$ [Lemma 2]. An explicit solution of $U_r^*(s, \underline{z})$ was found, from which $T_r^*(s)$ was obtained [Theorem 1]. We then proved that the stationary distribution of $\underline{m}^{(r)}$, r = 1, 2, ... R, is the same as that of $\underline{n}$ [Theorem 2]. With this result, we solved for the mean value of t_r [Theorem 3]; this last result is similar to the mean response time result of processor-sharing models. We also obtained an efficient recursive algorithm to calculate the second order statistics of t_r [Theorem 4]. Some numerical results are shown in Section 3.

2. THE ANALYSIS

Consider the system state $\underline{n} = (n_1, n_2, \ldots, n_R)$ at arrival instants. Recall that n_k is the number of class k customers with exactly k more rounds of service to go. The aggregate arrival rate of customers to the k^{th} class is

$$\lambda_k = \sum_{i=k}^{R} \gamma_i \tag{2}$$

since any new arrival who requires at least k rounds of service must enter and leave the k^{th} class exactly once.

Lemma 1. The moment generating function of $\underline{n}$ is

$$P^*(\underline{z}) = \frac{1-\rho}{1 - \sum_{k=1}^{R} \rho_k z_k} \tag{3}$$

where $\rho_k = \lambda_k/\mu$ and $\rho = \sum_{k=1}^{R} \rho_k$.

Proof. Given Poisson arrival processes, the system state probabilities at an arrival instant are the same as system state probabilities at a random time instant [9]. With each round of service being exponentially distributed with the same mean $(1/\mu)$, we have an open queueing network that satisfies local balance [1]. Eq. (3) has been obtained by Reiser and Kobayashi [10]. (Q. E. D.)

Since each round of service is exponentially distributed, it has the moment generating function

$$B^*(s) = \frac{\mu}{s + \mu} \tag{4}$$

A recursive solution of $U_r^*(s, \underline{z})$ is next given.

Lemma 2.

$$U_0^*(s, \underline{z}) = P^*(\underline{z}) \tag{5}$$

$$U_{r+1}^*(s, \underline{z}) = y_1(s,\underline{z})\, U_r^*(s, \underline{y}(s,\underline{z})) \qquad r \geq 0 \tag{6}$$

where

$$\underline{y}(s,\underline{z}) = (y_1(s,\underline{z}), y_2(s,\underline{z}), \ldots, y_R(s,\underline{z})),$$

$$y_1(s,\underline{z}) = B^*(s + \sum_{i=1}^{R} \gamma_i(1 - z_i)),$$

and

$$y_k(s,\underline{z}) = z_{k-1}\, y_1(s,\underline{z}) \qquad \text{for } 2 \le k \le R$$

Proof. For $r = 0$, $t_0 = 0$ and $\underline{m}^{(0)} = \underline{n}$. This and the definition of $U_r^*(s, \underline{z})$ yield (5) at once.

To show (6), consider the time period between t_r and t_{r+1} during which the server served $M^{(r)} + 1$ customers, where $M^{(r)} = \sum_{k=1}^{R} m_k^{(r)}$ and the extra one is for the tagged customer's $(r + 1)^{st}$ round. During the same time period, each class k customer became a class $(k - 1)$ customer where $k = 2, 3, \ldots R$. Furthermore, let $A_k(t)$ be the number of external new arrivals to class k during time $t(= t_{r+1} - t_r)$ according to a Poisson process of rate γ_k customers per second. We note that class R is an exception in that its $m_R^{(r+1)}$ customers are all new arrivals. Thus, conditioning on t_r and $\underline{m}^{(r)}$, we have

$$U_{r+1}^*(s, \underline{z}/t_r, \underline{m}^{(r)}) = E[e^{-s(t+t_r)}\, z_1^{m_2^{(r)} + A_1(t)}\, z_2^{m_3^{(r)} + A_2(t)} \ldots z_R^{A_R(t)} / t_r, \underline{m}^{(r)}]$$

$$= e^{-st_r} (\prod_{k=2}^{R} z_{k-1}^{m_k^{(r)}})\, E[e^{-st}\, z_1^{A_1(t)}\, z_2^{A_2(t)} \ldots z_R^{A_R(t)} / M^{(r)}].$$

The last quantity on the right hand side is $(y_1(s,\underline{z}))^{M^{(r)}+1}$ because t is the sum of $M^{(r)} + 1$ independent identically distributed random variables with the moment generating function $B^*(s)$. The above equation can be rewritten as

$$U_{r+1}^*(s, \underline{z}/t_r, \underline{m}^{(r)}) = y_1(s,\underline{z}) \;\{e^{-st_r}\, y_1(s,\underline{z})^{m_1^{(r)}} \prod_{k=2}^{R} [z_{k-1}\, y_1(s,\underline{z})]^{m_k^{(r)}}\}$$

Unconditioning on t_r and $\underline{m}^{(r)}$, (6) follows. (Q. E. D.)

Explicit solutions for $U_r^*(s, \underline{z})$ and $T_r^*(s)$ can now be shown.

Theorem 1. (i) $$U_r^*(s, \underline{z}) = \frac{1 - \rho}{P_r(s) - \sum_{k=1}^{R} Q_{k,r}(s) z_k} \qquad r \ge 0 \tag{7}$$

where $P_r(s)$ and $Q_{k,r}(s)$ are polynomials in s given by

$$\begin{bmatrix} P_r(s) \\ Q_{1,r}(s) \\ Q_{2,r}(s) \\ \vdots \\ Q_{R-1,r}(s) \\ Q_{R,r}(s) \end{bmatrix} = \begin{bmatrix} (1 + \frac{s}{\mu} + \rho_1) & -1 & 0 & 0 & \ldots & 0 \\ \gamma_1/\mu & 0 & 1 & 0 & & 0 \\ \gamma_2/\mu & 0 & 0 & 1 & & 0 \\ \vdots & \vdots & & & \ddots & \vdots \\ & & & & 1 & 0 \\ \gamma_{R-1}/\mu & 0 & 0 & \ldots & 0 & 1 \\ \gamma_R/\mu & 0 & 0 & \ldots & 0 & 0 \end{bmatrix}^r \begin{bmatrix} 1 \\ \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{R-1} \\ \rho_R \end{bmatrix} \tag{8}$$

(ii) $$T_r^*(s) = \frac{1-\rho}{P_r(s) - \sum_{k=1}^{R} Q_{k,r}(s)} \tag{9}$$

Proof. (i) Because of (3) and (5), (7) holds for $r = 0$ with $P_0(s) = 1$ and $Q_{k,0}(s) = \rho_k$ for $1 \leq k \leq R$. Assuming that (7) holds for r, we use (6) and (4) to express $U_{r+1}^*(s, \underline{z})$ as follows.

$$U_{r+1}^*(s, \underline{z}) = \frac{1}{1 + \frac{s}{\mu} + \sum_{i=1}^{R} \frac{\gamma_i}{\mu}(1-z_i)} \cdot \frac{1-\rho}{P_r(s) - \dfrac{Q_{1,r}(s) - \sum_{k=1}^{R-1} Q_{k+1,r}(s)\, z_k}{1 + (s/\mu) + \sum_{i=1}^{R}(\gamma_i/\mu)(1-z_i)}}$$

$$= \frac{1-\rho}{\{(1 + \frac{s}{\mu} + \sum_{i=1}^{R} \frac{\gamma_i}{\mu})P_r(s) - Q_{1,r}(s)\} - \sum_{k=1}^{R-1} [\frac{\gamma_k}{\mu} P_r(s) + Q_{k+1,r}(s)] z_k - \frac{\gamma_R}{\mu} z_R P_r(s)}$$

Thus, the form of (7) is maintained, and it is evident from the above that

$$\begin{bmatrix} P_{r+1}(s) \\ Q_{1,r+1}(s) \\ \cdot \\ \cdot \\ \cdot \\ Q_{R,r+1}(s) \end{bmatrix} = \begin{bmatrix} (1 + \frac{s}{\mu} + \rho_1) & -1 & 0 & \cdots & 0 \\ \gamma_1/\mu & 0 & 1 & & \cdot \\ \cdot & & & \ddots & 0 \\ \cdot & & & 0 & 1 \\ \gamma_R/\mu & \cdots & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} P_r(s) \\ Q_{1,r}(s) \\ \cdot \\ \cdot \\ \cdot \\ Q_{R,r}(s) \end{bmatrix} \tag{10}$$

The recursion in (10) started at $r = 0$ clearly yields (8).

(ii) (9) follows from (7) and $T_r^*(s) = U_r^*(s, \underline{1})$. (Q. E. D.)

For $r = 1$, 2 and 3, we show $U_r^*(s, \underline{z})$ below.

$$U_1^*(s, \underline{z}) = \frac{1-\rho}{1 + \frac{s}{\mu} - \sum_{k=1}^{R} \rho_k z_k}$$

$$U_2^*(s, \underline{z}) = \frac{1-\rho}{(1 + \frac{s}{\mu})^2 + \frac{s}{\mu}\rho_1 - \sum_{k=1}^{R} (\rho_k + \frac{s}{\mu^2}\gamma_k) z_k}$$

$$U_3^*(s, \underline{z}) = (1-\rho) / \Big\{ (1 + \tfrac{s}{\mu})^3 + 2(\tfrac{s}{\mu})^2 \rho_1 + \tfrac{s}{\mu}(\rho_2 + 2\rho_1 + \rho_1^2) - \sum_{i=1}^{R-1} \{\tfrac{\gamma_i}{\mu}[(1+\tfrac{s}{\mu})^2 + \tfrac{s}{\mu}\rho_1] + [\rho_{i+1} + \tfrac{\gamma_{i+1}}{\mu}\tfrac{s}{\mu}]\} z_i - \tfrac{\gamma_R}{\mu}[(1 + \tfrac{s}{\mu})^2 + \tfrac{s}{\mu}\rho_1] z_R \Big\}$$

From the above, we obtain $T_r^*(s)$ for $r = 1$, 2 and 3 by letting $\underline{z} = \underline{1}$ in $U_r^*(s, \underline{z})$.

$$T_1^*(s) = \frac{1-\rho}{(1 + \frac{s}{\mu}) - \rho}$$

$$T_2^*(s) = \frac{1-\rho}{(1 + \frac{s}{\mu})^2 - \rho}$$

$$T_3^*(s) = \frac{1-\rho}{(1+\frac{s}{\mu})^3 + \rho_1(\frac{s}{\mu})^2 - \rho}$$

We note that the solutions for $U_r^*(s, \underline{z})$ and $T_r^*(s)$ become quite complex if one tries to solve for $P_r(s)$ and $Q_{k,r}(s)$ explicitly using the matrix equation (8) when $r \geq 4$. In what follows, we turn our attention to finding the moments of t_r. To do so, we need the following result concerning the distribution of $\underline{m}^{(r)}$.

Theorem 2. For any $r \geq 0$, $\underline{m}^{(r)}$ and $\underline{n}$ have the same stationary distribution. That is

$$U_r^*(0, \underline{z}) = E[z_1^{m_1^{(r)}} z_2^{m_2^{(r)}} \ldots z_R^{m_R^{(r)}}] = P^*(\underline{z}) \tag{11}$$

Proof. By (5), (11) holds true for $r = 0$. Assume that (11) holds true for some r so that $U_r^*(0, \underline{z}) = P^*(\underline{z})$. By (3), (6) and the induction hypothesis,

$$U_{r+1}^*(0, \underline{z}) = y_1(0,\underline{z}) \cdot \frac{1-\rho}{1 - \sum_{k=1}^{R} \rho_k\, y_k(0,\underline{z})} = \frac{1-\rho}{\frac{1}{y_1(0,\underline{z})} - (\rho_1 + \sum_{k=1}^{R-1} \rho_{k+1}\, z_k)}$$

$$= \frac{1-\rho}{1 + \sum_{i=1}^{R} \frac{\gamma_i}{\mu}(1-z_i) - \rho_1 - \sum_{k=1}^{R-1} \rho_{k+1}\, z_k} = \frac{1-\rho}{1 - \sum_{k=1}^{R} \rho_k\, z_k}$$

which is $P^*(\underline{z})$. The last equality is obtained using the following relationships:

$$\rho_1 = \frac{\lambda_1}{\mu} = \sum_{i=1}^{R} \frac{\gamma_i}{\mu} \quad \text{and} \quad \rho_k = \frac{\lambda_k}{\mu} + \rho_{k+1} \quad \text{for} \quad 1 \leq k \leq R-1. \qquad \text{(Q. E. D.)}$$

The moments of t_r can be obtained from the moment generating function of t_r and $\underline{m}^{(r)}$ as follows.

$$E[t_r^n] = (-1)^n \frac{\partial^n}{\partial s^n} U_r^*(s, \underline{z})\Big|_{s=0,\ \underline{z}=\underline{1}}$$

$$= (-1)^n \frac{\partial^n}{\partial s^n} U_r^*(s,z,z,\ldots,z)\Big|_{s=0,\ z=1} \tag{12}$$

Theorem 3. The conditional mean response time is

$$E[t_r] = \frac{r/\mu}{1-\rho} \tag{13}$$

The above theorem is proved by first expressing $E[t_{r+1}]$ in terms of $E[t_r]$ using (6), (11) and (12). (13) is then obtained by induction starting with $E[t_0] = 0$. (See [11] for details of Proof.)

Theorem 4. The second order statistics of the conditional response time can be found recursively using

$$Var(t_{r+1}) = Var(t_r) + \frac{1-2\rho r}{\mu^2(1-\rho)^2} + \frac{2}{\mu} E[t_r M^{(r)}] \tag{14}$$

$$E[t_{r+1} M^{(r+1)}] = \sum_{i=1}^{R} E[t_{r+1} m_i^{(r+1)}] \tag{15}$$

and

$$E[t_{r+1} m_i^{(r+1)}] = \frac{2\rho_i}{\mu\,(1-\rho)^2} + \frac{r\gamma_i}{\mu^2(1-\rho)} + \frac{\gamma_i}{\mu}\, E[t_r M^{(r)}] + E[t_r\, m_{i+1}^{(r)}] \qquad 1 \le i \le R \tag{16}$$

where $Var(t_r)$ is the variance of t_r and $E[t_r\, m_{R+1}^{(r)}]$ is zero, with the initial condition

$$Var(t_0) = 0$$

$$E[t_0\, m_i^{(0)}] = 0 \quad \text{for} \quad 1 \le i \le R .$$

The above theorem is proved by taking derivatives of (6), using the moment generating properties of transforms; (11) and (13) are used to simplify the resulting expressions. (See [11] for details of proof.)

3. DISCUSSIONS AND NUMERICAL EXAMPLES

The conditional mean response time result in Theorem 3 is analogous to results from analyses of a processor-sharing queue [2,3]. The mean response time

$$E[t_r] = \frac{r/\mu}{1-\rho}$$

of a type r job varies linearly as its (expected) service requirement r/μ.

The contribution of this paper is the derivation of higher order statistics for the response times of different types of jobs; also the service requirements (in number of rounds of service) of each type of jobs can have a general probability distribution.

By assuming that each round of service is exponentially distributed, the multi-class feedback queue considered is an open queueing network satisfying local balance. Each type of jobs corresponds to customers following a fixed path. The key idea in our solution approach is to develop a recursive relationship between the response time of a path and the response time of the same path extended by one more transition.

To illustrate our results, we apply the recursive algorithm in Theorem 4 to calculate the standard deviation of t_r for the following two examples.

Example 1. The service requirements of customers have a small coefficient of variation. The probability of a customer requiring r rounds of service is

$$a_r = \begin{cases} 1/3 & r = 19,\ 20,\ 21 \\ 0 & \text{otherwise} \end{cases}$$

Example 2. The service requirements of customers have a large coefficient of variation. The probability of a customer requiring r rounds of service is

$$a_r = \begin{cases} 80/99 & r = 1 \\ 19/99 & r = 100 \\ 0 & \text{otherwise} \end{cases}$$

The standard deviation of t_r is plotted versus ρ in Figures 2 and 3 for Examples 1 and 2 respectively, for different values of r.

For comparison, we have plotted two additional curves in Figures 2 and 3. One is the standard deviation of the response time versus ρ of an arbitrary job for our feedback queue (a round-robin system). The other is the standard deviation of the response time of an arbitrary job in a FCFS system (no feedback; a customer requiring r rounds of service gets all of them at the same time).

In both examples, the FCFS system gives rise to a smaller standard deviation for the response time of an arbitrary customer than the round-robin system.

In Figure 2, note that all customers require r=19,20 or 21 rounds of service. (The r=1 and r=10 curves correspond to no customers.) Therefore, FCFS gives rise to a smaller standard deviation for the response time of all customers than round-robin.

In Figure 3, the standard deviation of t_r for small values of r is smaller than the FCFS standard deviation at the same ρ. The exact crossover point depends upon ρ.

Acknowledgments

This work was supported by National Science Foundation under Grant No. ENG78-01803. The authors thank Patsy McMains for her typing assistance.

REFERENCES

[1] Baskett, F., K. Chandy, R. Muntz and F. Palacios, "Open, Closed and Mixed Networks of Queues with Different Classes of Customers," J. ACM, Vol. 22, 1975.

[2] Kleinrock, L., "Time-Shared Systems: A Theoretical Treatment," J. ACM, Vol. 14, April 1967.

[3] Sakata, M., S. Noguchi and J. Oizumi, "An Analysis of the M/G/1 Queue Under Round-Robin Discipline," Operations Research, Vol. 19, 1971.

[4] Coffman, E., R. Muntz and H. Trotter, "Waiting Time Distributions for Processor-Sharing Systems," J. ACM, Vol. 17, January 1970.

[5] Muntz, R., "Waiting Time Distribution for Round-Robin Queueing Systems," Proc. Symp. on Computer-Communications Networks and Teletraffic, Polytechnic Institute of Brooklyn, April 1972.

[6] Takacs, L., "A Single-Server Queue with Feedback," The Bell System Technical Journal, March 1963.

[7] Wong, J. W., "Distribution of End-to-End Delay in Message-Switched Networks," Computer Networks, Vol. 2, 1978.

[8] Cohen, J. W., The Single Server Queue, North-Holland Publishing Co., Amsterdam 1969.

[9] Kleinrock, L., Queueing Systems, Vol. 1: Theory, Wiley-Interscience, 1975, Page 118.

[10] Reiser, M. and H. Kobayashi, "Queueing Networks with Multiple Closed Chains: Theory and Computational Algorithms," IBM Journal of Research and Development, Vol. 19, May 1975.

[11] Lam, S. S. and A. U. Shankar, "Response Time Distributions for a Multi-Class Queue with Feedback," Department of Computer Sciences, University of Texas at Austin, Technical Report TR-114, Nov. 1979.

Fig. 2. Standard deviation of response time versus ρ for service requirements with a small coefficient of variation.

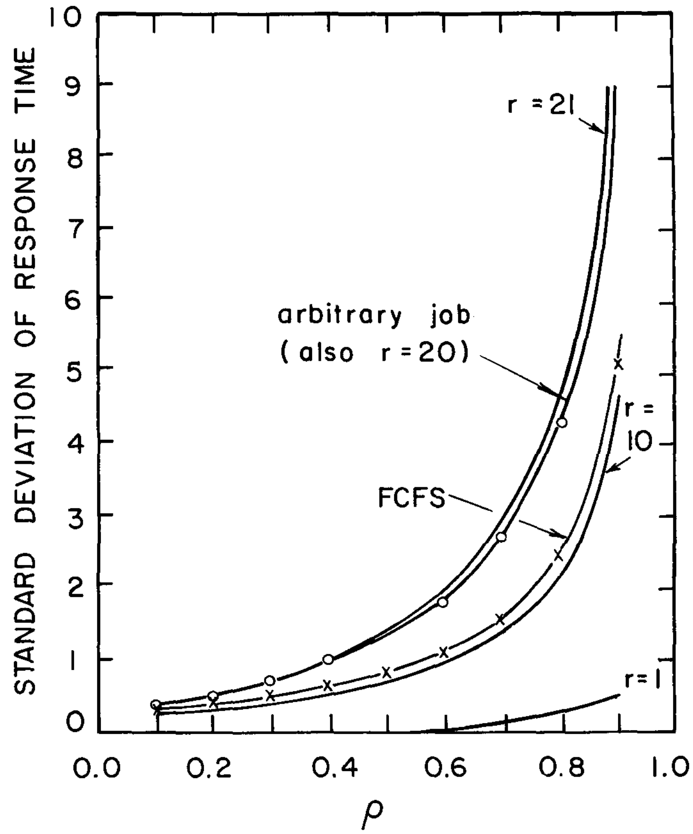

Fig. 3. Standard deviation of response time versus ρ for service requirements with a large coefficient of variation.