RESPONSE TIME DISTRIBUTIONS FOR A MULTI-CLASS QUEUE WITH FEEDBACK*

S. S. Lam and A. U. Shankar

TR - 114 November 1979

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

*This work was supported by National Science Foundation under Grant No. ENG78-01803.

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 with a probability that is dependent upon his class membership and number of rounds of service achieved. By properly defining customer classes, a wide range of non-exponential service time requirements are admissible in this model. 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 R classes of customers. The arrival process of the r^{th} class is an independent Poisson process ($r = 1, 2, \ldots R$). 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, depending upon his class membership and number of rounds of service achieved.

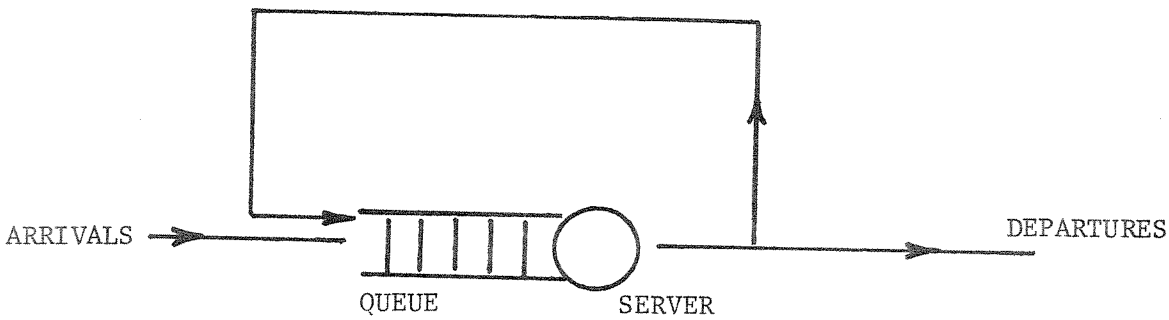

Figure 1. A feedback queue model.

where $a_i^{(r)}$ can be interpreted as the probability of a class r customer requiring exactly i rounds of service. $\left\{a_i^{(r)},\ 1 \leq i \leq R\right\}$ can be an arbitrary set of probabilities that sum to one and may be different for different customer classes.

Our model is also different from the feedback queue model of Takács [5]. In his model, each round of service can have a general distribution. However, he considered a single class 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 [6] 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

We shall, without any loss of generality, consider the following model. There are R classes of customers. The arrival process of the rth class is Poisson at rate γ_r customers per second. A class 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 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

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

algorithm to calculate the second order statistics of t_r [Theorem 4].

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 customers with exactly k more rounds of service to go. Let us redefine the meaning of customer classes to correspond to $n_1, n_2, \ldots, n_R$. Hence 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 and that each round of service is 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 [8]. (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

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

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

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

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.

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

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

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

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

<u>Proof.</u> (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.

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

$$+ [\rho_{i+1} + \frac{\gamma_{i+1}}{\mu}\frac{s}{\mu}]\Big\} z_i - \frac{\gamma_R}{\mu}[(1+\tfrac{s}{\mu})^2 + \tfrac{s}{\mu}\rho_1]\, z_R\Big\}$$

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 $\underline{m}^{(r)}$.

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

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

<u>Proof</u>. 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 (6) and the induction hypothesis,

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

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

<u>Theorem 3</u>. The conditional mean response time is

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

<u>Proof</u>. Using (6) and (12), we have

$$E[t_{r+1}] = -\frac{\partial}{\partial s}\{B^*(s+\lambda_1(1-z))\ U_r^*(s,\ \underline{y}(z,z,\ \ldots,\ z))\}\Big|_{s\,=\,0,\ z\,=\,1}$$

$$= \frac{1}{\mu} - E[\frac{\partial}{\partial s}\{e^{-st_r}(B^*(s))^{M^{(r)}}\}]_{s\,=\,0}$$

$$= \frac{1}{\mu} - \{-E[t_r] - E[M^{(r)}].\frac{1}{\mu}\}$$

By (11),

$$E[M^{(r)}] = \frac{\partial}{\partial z} P^*(z,z,\ \ldots,\ z)\Big|_{z\,=\,1} = \frac{\rho}{1-\rho}$$

Substituting this into the above expression for $E[t_{r+1}]$, we have

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

which yields (13) by induction starting with $E[t_0] = 0$. (Q.E.D.)

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

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

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

and

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

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

$$\mathrm{Var}(t_0) = 0$$

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

$$E[t_{r+1}\, m_i^{(r+1)}] = -\frac{\partial^2}{\partial s\, \partial z_i} \{y_1(s,\underline{z})\; U_r^*(s,\underline{y}(s,\underline{z}))\}\Big|_{s=0,\ \underline{z}=\underline{1}}$$

$$= \Bigg[-\frac{\partial}{\partial z_i} \Bigg\{[\frac{\partial}{\partial s}\, y_1(s,\underline{z})].U_r^*(s,\underline{y}(s,\underline{z})) + y_1(s,\underline{z}).E\Big[z_1^{m_2^{(r)}} \cdots\cdots z_{R-1}^{m_R^{(r)}}$$

$$.\ \frac{\partial}{\partial s}\ \{e^{-st_r}(y_1(s,\underline{z}))^{M^{(r)}}\}\Big]\Bigg\}_{s=0,\ z_j=1 \text{ for } j\neq i}\Bigg]_{z_i=1}$$

$$= \frac{\partial}{\partial z_i}\Bigg\{\frac{1}{\mu}\,(y_1(z_i))^2 E[z_i^{m_{i+1}^{(r)}}(y_1(z_i))^{M^{(r)}}] + y_1(z_i)E[z_i^{m_{i+1}^{(r)}}\{t_r(y_1(z_i))^{M^{(r)}}$$

$$+ M^{(r)}(y_1(z_i))^{M^{(r)}-1}.\frac{1}{\mu}.\,(y_1(z_i))^2\}]\Bigg\}_{z_i=1}$$

$$\left(\text{where } y_1(z_i) = y_1(0,\underline{z})\big|_{z_j=1 \text{ for } j\neq i} = \frac{1}{1+\frac{\gamma_i}{\mu}\,(1-z_i)}\right)$$

$$= \frac{2\gamma_i}{\mu^2} + \frac{1}{\mu}\,\{E[m_{i+1}^{(r)}] + E[M^{(r)}]\,\frac{\gamma_i}{\mu}\} + \frac{\gamma_i}{\mu}\,\{E[t_r] + \frac{1}{\mu}\,E[M^{(r)}]\}$$

$$+ \{\frac{\gamma_i}{\mu}\quad E[t_r\; M^{(r)}] + \frac{\gamma_i}{\mu^2}\,E[M^{(r)}(M^{(r)}+1)]\} + \{E[t_r\; m_{i+1}^{(r)}] + \frac{1}{\mu}\,E[M^{(r)} m_{i+1}^{(r)}]\}$$

$$= \Big\{\frac{2\gamma_i}{\mu^2} + \frac{1}{\mu}\,E[m_{i+1}^{(r)}] + \frac{3\gamma_i}{\mu^2}\,E[M^{(r)}] + \frac{\gamma_i}{\mu}\,E[t_r]$$

$$+ \frac{\gamma_i}{\mu^2}\quad E[(M^{(r)})^2]\Big\} + \frac{\gamma_i}{\mu}\,E[t_r\; M^{(r)}] + E[t_r\; m_{i+1}^{(r)}]$$

where the terms bracketted by {} can be evaluated using (11) and (13) to yield (16).

(Q.E.D.)

Acknowledgment

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

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

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

[8] 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.