Mathematical Problems in Engineering
Volume 6, pp. 495–503
Reprints available directly from the publisher
Photocopying permitted by license only

On an $M^1, M^2/G^r/1$ Queueing System

LOTFI TADJ

Department of Statistics and Operations Research, College of Science, King Saud University, P. O. Box 2455, Riyadh 11451, Saudi Arabia

(Received 11 May 1998; In final form 3 November 1999)

The author studies a service delayed queueing system with priority discipline. The joint queue size distribution is derived in the steady state.

Keywords: Queue; Embedded Markov chain; Priority discipline; Joint distribution; Steady-state distribution

AMS Subject Classifications: Primary: 60K10, 60K25; Secondary: 90B22, 90B25

1. INTRODUCTION

Queueing systems where the server does not start service until the queue size has reached a specified level have been extensively studied by different authors, sometimes under different appelations, such as service delayed models, models with accumulation level, quorum models, models under q-policy, *etc.* Results and applications of these models and of their various variants are reported by Dshalalow [3, 4].

Just as much studied, if not more, are priority queueing systems. Many real world queueing systems have their customers divided into classes. It is only reasonable to distinguish between express and regular mail, rush and ordinary jobs, and, generally speaking, important and less important customers. The use of priority discipline improves the measures of performance of the higher priority classes at the expense of the lower priority classes.

The present paper incorporates a priority discipline into a service delayed queueing system. Customers are divided into two classes: a class 1 (high-priority) customers and a class 2 (low-priority) customers. The interest is on the joint queue size distribution, and we derive it in the steady state.

The organization of this paper is as follows. In Section 2, we report some results related to the service delayed queueing system $M/G^r/1$. In Section 3, we introduce the service delayed queueing system with priority discipline $M^1, M^2/G^r/1$ and derive the joint queue size distribution in the steady state. A summary and directions for further research are given at the end of the paper.

2. THE M/G^r/1 MODEL

Consider a queueing system where customers arrive to the facility and wait for service on a first-come, first served basis. Arrivals occur according to a Poisson process with rate λ. Assume there is one server. Assume also that the server becomes idle after a departure if the queue size is smaller than a given level r and that he resumes work when the queue size has grown up to the level r. The size of the batch of customers taken for each service is fixed and equal to r. Let $Q(t)$ denote the number of customers in the system at an arbitrary time t and let Q_n be the number of customers in the system at the completion epoch T_n of the nth service, that is $Q_n = Q(T_n^+)$. Also, let A_{n+1} be the number of arrivals during the service period of the $(n+1)$th batch. Then, (Q_n) satisfies the recursion

$$Q_{n+1} = (Q_n - r)^+ + A_{n+1}. \tag{2.1}$$

(Q_n) is a Markov chain. It is ergodic if and only if $\lambda b < r$, where b is the mean service duration. Denote by $P_n(x) = E[x^{Q_n}]$ the probability generating function of Q_n, by $p_i = \lim_{n\to\infty} P\{Q_n = i\}$ the steady-state distribution. Also, let $A(x) = E[x^{A_n}] = B^*(\lambda - \lambda x)$, where $B^*(s)$ is the Laplace–Stieltjes transform of the service time distribution B, let $U = \{1, 2, \ldots, r-1\}$ and let $I_U(\cdot)$ denote the indicator function of set U.

THEOREM *The probability generating function in the steady state* $P(x) = \lim_{n\to\infty} P_n(x)$ *satisfies the following relation:*

$$P(x) = \frac{A(x)[x^r H(1) - H(x)]}{x^r - A(x)}. \qquad (2.2)$$

where

$$H(x) = \sum_{i=0}^{r-1} p_i x^i = \lim_{n\to\infty} E[x^{Q_n} I_U(Q_n)]. \qquad (2.3)$$

Proof

$$\begin{aligned} P_{n+1}(x) &= E[x^{Q_{n+1}}] \\ &= E[x^{(Q_n - r)^+ + A_{n+1}}] \\ &= A(x)(x^{-r} E[x^{Q_n} I_{U^c}(Q_n)] + E[I_U(Q_n)]) \\ &= A(x)(x^{-r}\{P_n(x) - E[x^{Q_n} I_U(Q_n)]\} + E[I_U(Q_n)]). \end{aligned}$$

Now, let $n \to \infty$ and solve for $P(x)$ to get (2.2).

The result above is given for example in Dshalalow and Russell [5]. They also show how to get the unknown probabilities $p_0, \ldots, p_{r-1}$ in (2.2) using a variant of Rouche's theorem.

We will be needing in Section 3 the distribution of the first return time for the subset $\{0, 1, \ldots, r-1\}$ of states to itself. It was shown by Tadj and Rikli [7] that the transition probability matrix of (Q_n)

$$Q = \begin{pmatrix} f_0 & f_1 & f_2 & f_3 & \cdots \\ f_0 & f_1 & f_2 & f_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f_0 & f_1 & f_2 & f_3 & \cdots \\ 0 & f_0 & f_1 & f_2 & \cdots \\ 0 & 0 & f_0 & f_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \qquad (2.4)$$

where

$$f_k = \int_0^\infty \frac{(\lambda u)^k}{k!} e^{-\lambda u} dB(u), \quad k \geq 0, \qquad (2.5)$$

can be partitioned into blocks of dimensions $r \times r$ as follows

$$Q = \begin{pmatrix} B_0 & B_1 & B_2 & B_3 & \cdots \\ A_0 & A_1 & A_2 & A_3 & \cdots \\ 0 & A_0 & A_1 & A_2 & \cdots \\ 0 & 0 & A_0 & A_1 & \cdots \\ 0 & 0 & 0 & A_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \tag{2.6}$$

where, for $n \geq 0$, the matrices B_n have r identical rows with elements

$$(B_n)_{ij} = f_{nr+j}, \quad i,j = 0, \ldots, r-1, \tag{2.7}$$

for $n \geq 1$, the matrices A_n have elements

$$(A_n)_{ij} = f_{nr+j-i}, \quad i,j = 0, \ldots, r-1, \tag{2.8}$$

and A_0 is upper triangular with elements

$$(A_0)_{ij} = f_{j-i}, \quad i \leq j, \; i,j = 0, \ldots, r-1. \tag{2.9}$$

Using the terminology of Neuts [6], Q is said to be a stochastic matrix of M/G/1 type. Informally speaking, level i, $i \geq 0$, consists of the r states in the ith row of matrix Q, as represented in (2.6). For $k \geq 1, 0 \leq i,j < r-1$, let $K_{ij}(k)$ be the conditional probability that the Markov chain, starting in state i of level 0, returns to level 0 by hitting state j, after exactly k transitions. Because the matrices B_n, $n \geq 0$, have identical rows, the probabilities $K_{ij}(k)$ are independent of the initial state i and will be denoted $K_j(k)$. For $k \geq 1$, let $K(k) = [K_0(k), K_1(k), \ldots, K_{r-1}(k)]$ and let $K^*(z) = \sum_{k=1}^{\infty} K(k) z^k$. Then, using Neuts' method [6], the vector $K^*(z)$ is given by (see for example Choi and Lee [1]):

$$K^*(z) = z[f_0, \ldots, f_{r-1}] + \sum_{n=1}^{\infty} z[f_{rn}, \ldots, f_{rn+r-1}] G^n(z), \tag{2.10}$$

where the matrix G is the solution to the matrix equation

$$G(z) = z \sum_{n=0}^{\infty} A_n G^n(z). \tag{2.11}$$

3. THE $M^1, M^2/G^r/1$ MODEL

We want to generalize and "refine" the previous model by assuming that customers are divided into two classes: high-priority and low-priority customers. Customers of each class follow an independent Poisson arrival process with rates λ_1 and λ_2, respectively. They are accommodated in an infinite capacity waiting room. The service time distribution of each size r batch of customers is B, the same for both priority classes. At time T_n (the end of a service), the server takes r class 1 customers for service, if available. Otherwise, if there are i ($i<r$) class 1 customers, then the server completes the batch by taking $(r-i)$ class 2 customers. Finally, if less than r class 1 and class 2 customers are waiting, then the server idles until the level r is reached.

The analysis used in this section is not new. It generalizes that of the previous section, making use of indicator functions. It has been used, for example, by Choi and Lee [1] in studying a priority queueing system, by Dshalalow [2] in studying the theory of fluctuations, *etc.*

Let $Q^i(t)$ be the number of class i ($i=1,2$) customers in the system at an arbitrary time t and denote by $Q_n^i = Q^i(T_n^+)$ the number of class i ($i=1,2$) customers in the system at a service completion epoch T_n. Then (Q_n^1, Q_n^2) is a two-dimensional Markov chain satisfying the following recursive equations:

$$Q_{n+1}^1 = (Q_n^1 - r)^+ + A_{n+1}, \tag{3.1}$$

$$Q_{n+1}^2 = (Q_n^2 - (r - Q_n^1)^+)^+ + B_{n+1}, \tag{3.2}$$

where A_{n+1} (respectively B_{n+1}) is the number of high priority (respectively low-priority) arrivals during the service period of the $(n+1)$th batch. The necessary and sufficient condition for the system to reach steady-state is obviously $(\lambda_1+\lambda_2)b<r$. Let $P_n(x,y) = E[x^{Q_n^1}y^{Q_n^2}]$ denote joint probability generating function of (Q_n^1, Q_n^2), and let $A(x)=E[x^{A_n}]=B^*(\lambda_1-\lambda_1 x)$ and $B(x)=E[x^{B_n}]=B^*(\lambda_2-\lambda_2 x)$.

THEOREM *The joint probability generating function in the steady state* $P(x,y)=lim_{n\to\infty}P_n(x,y)$ *satisfies the following relation*:

$$P(x,y) = \frac{x^r A(x)B(y)}{x^r - A(x)B(y)}\left(E[I_U(Q^1+Q^2)] - \frac{1}{y^r}E[y^{Q^1+Q^2}I_U(Q^1+Q^2)] - \frac{1}{x^r}R(x,y) + \frac{1}{y^r}R(y,y)\right), \tag{3.3}$$

where

$$R(x,y) = E[x^{Q^1} y^{Q^2} I_U(Q^1)]. \tag{3.4}$$

Proof

$$\begin{aligned}
P_{n+1}(x,y) &= E[x^{Q^1_{n+1}} y^{Q^2_{n+1}}] \\
&= E[x^{(Q^1_n - r)^+ + A_{n+1}} y^{(Q^2_n - (r - Q^1_n)^+)^+ + B_{n+1}}] \\
&= E[x^{A_{n+1}}]E[y^{B_{n+1}}]E[x^{(Q^1_n - r)^+} y^{(Q^2_n - (r - Q^1_n)^+)^+}] \\
&= A(x)B(y)(E[x^{Q^1_n - r} y^{Q^2_n} I_{U^c}(Q^1_n)] + E[y^{(Q^1_n + Q^2_n - r)^+} I_U(Q^1_n)]) \\
&= A(x)B(y)\bigg(\frac{1}{x^r} E[x^{Q^1_n} y^{Q^2_n} I_{U^c}(Q^1_n)] \\
&\qquad + E[y^{Q^1_n + Q^2_n - r} I_U(Q^1_n) I_{U^c}(Q^1_n + Q^2_n)] \\
&\qquad + E[I_U(Q^1_n + Q^2_n)] \bigg) \\
&= A(x)B(y)\bigg(\frac{1}{x^r} \{P_n(x,y) - E[x^{Q^1_n} y^{Q^2_n} I_{U^c}(Q^1_n)]\} \\
&\qquad + E[I_U(Q^1_n + Q^2_n)] \\
&\qquad + \frac{1}{y^r} \{E[y^{Q^1_n + Q^2_n} I_U(Q^1_n)] \\
&\qquad - E[y^{Q^1_n + Q^2_n} I_U(Q^1_n + Q^2_n)]\} \bigg) \\
&= A(x)B(y)\bigg(\frac{1}{x^r} P_n(x,y) + E[I_U(Q^1_n + Q^2_n)] \\
&\qquad - \frac{1}{y^r} E[y^{Q^1_n + Q^2_n} I_U(Q^1_n + Q^2_n)] \\
&\qquad + \frac{1}{y^r} E[y^{Q^1_n + Q^2_n} I_U(Q^1_n)] \\
&\qquad - \frac{1}{x^r} E[x^{Q^1_n} y^{Q^2_n} I_U(Q^1_n)] \bigg).
\end{aligned}$$

Now, let $n \to \infty$ and solve for $P(x,y)$ to get (3.3).

Note that in the quantity between parentheses on the right-hand-side of (3.3), the first two terms can be derived from (2.2) and that only $R(x,y)$ needs to be determined for $P(x,y)$ to be fully determined.

In order to obtain $R(x,y)$ we follow the analysis of Choi and Lee [1], and consider the joint distribution of the numbers of high and low priority customers at a service completion in which $Q^1 < r$. So let t_n denote a service completion epoch in which $Q^1 < r$ and for $i = 1, 2$, let $\bar{Q}_n^i = Q^i(t_n)$. Also, let x_n be the time period between t_n and t_{n+1}. Define the joint probability generating function $\bar{R}_n(x,y) = E[x^{\bar{Q}_n^1} y^{\bar{Q}_n^2}]$ and the joint probability generating function in the steady state $\bar{R}(x,y) = \lim_{n\to\infty} R_n(x,y)$. Then R and $\bar{R}$ are related by

$$R(x,y) = P\{Q^1 < r\}\bar{R}(x,y), \tag{3.5}$$

so that $P(x,y)$ is fully determined if $\bar{R}(x,y)$ is. Note that for $l \geq 1$ and $j = 0, \ldots, r-1$, $P\{x_n = l, \bar{Q}_{n+1}^1 = j\} = K_j(l)$ is determined by (2.10). Let C_n denote the number of class 2 arrivals during the time period x_n. Then the embedded processes $\bar{Q}_n^1$ and $\bar{Q}_n^2$ satisfy

$$\bar{Q}_{n+1}^2 = (\bar{Q}_n^1 + \bar{Q}_n^2 - r)^+ + C_n, \tag{3.6}$$

THEOREM *The joint probability generating function in the steady state $\bar{R}(x,y) = lim_{n\to\infty} R_n(x,y)$ satisfies the following relation*:

$$\bar{R}(x,y) = (\bar{R}(y,y) - E[y^{\bar{Q}^1+\bar{Q}^2} I_U(\bar{Q}^1 + \bar{Q}^2)] + y^r E[I_U(\bar{Q}^1 + \bar{Q}^2)]) \times \frac{1}{y^r}\sum_{j=0}^{r-1} x^j K_j^*(B(y)). \tag{3.7}$$

Proof

$$\begin{aligned}
\bar{R}_{n+1}(x,y) &= E[x^{\bar{Q}_{n+1}^1} y^{\bar{Q}_{n+1}^2}] = E[x^{\bar{Q}_{n+1}^1} y^{(\bar{Q}_n^1+\bar{Q}_n^2-r)^+ + C_n}] \\
&= \sum_{l=1}^{\infty} E[x^{\bar{Q}_{n+1}^1} y^{(\bar{Q}_n^1+\bar{Q}_n^2-r)^+ + C_n} I_{\{l\}}(x_n)] \\
&= E\left[y^{(\bar{Q}_n^1+\bar{Q}_n^2-r)^+}\right] \sum_{l=1}^{\infty} [B(y)]^l E[x^{\bar{Q}_{n+1}^1} I_{\{l\}}(x_n)] \\
&= E\left[y^{(\bar{Q}_n^1+\bar{Q}_n^2-r)^+}\right] \sum_{j=0}^{r-1}\sum_{l=1}^{\infty} x^j [B(y)]^l K_j(l) \\
&= \left(E[y^{\bar{Q}_n^1+\bar{Q}_n^2-r} I_{U^c}(\bar{Q}_n^1 + \bar{Q}_n^2)] + E[I_U(\bar{Q}_n^1 + \bar{Q}_n^2)]\right) \sum_{j=0}^{r-1} x^j K_j^*(B(y))
\end{aligned}$$

$$= \left(E[y^{\bar{Q}_n^1+\bar{Q}_n^2} I_{U^c}(\bar{Q}_n^1+\bar{Q}_n^2)] + y^r E[I_U(\bar{Q}_n^1+\bar{Q}_n^2)]\right) \frac{1}{y^r} \sum_{j=0}^{r-1} x^j K_j^*(B(y))$$
$$= \left(\bar{R}_n(y,y) - E[y^{\bar{Q}_n^1+\bar{Q}_n^2} I_U(\bar{Q}_n^1+\bar{Q}_n^2)] + y^r E[I_U(\bar{Q}_n^1+\bar{Q}_n^2)]\right)$$
$$\times \frac{1}{y^r} \sum_{j=0}^{r-1} x^j K_j^*(B(y)).$$

Now, let $n \to \infty$ to get (3.7).

Note that in the quantity between parentheses on the right-hand-side of (3.7), the first two terms are unknown.

COROLLARY

$$\bar{R}(y,y) = \frac{\left(y^r E[I_U(\bar{Q}^1+\bar{Q}^2)] - E[y^{\bar{Q}^1+\bar{Q}^2} I_U(\bar{Q}^1+\bar{Q}^2)]\right) \sum_{j=0}^{r-1} y^j K_j^*(B(y))}{y^r - \sum_{j=0}^{r-1} y^j K_j^*(B(y))} \tag{3.8}$$

Proof Letting $x=y$ in (3.6) and solving for $\bar{R}(y,y)$ we get (3.8).

The only unknown term in (3.8), $E[y^{\bar{Q}^1+\bar{Q}^2} I_U(\bar{Q}^1+\bar{Q}^2)]$, can be determined by applying Rouche's theorem to $y^r - \sum_{j=0}^{r-1} y^j K_j^*(B(y))$, (see Choi and Lee [1]). Indeed, $y^r - \sum_{j=0}^{r-1} y^j K_j^*(B(y))$ has r roots in the region $|y| \leq 1$. These roots must also be roots of the numerator in (3.8) since $\bar{R}(y,y)$ is analytic in $|y| \leq 1$. Having determined $\bar{R}(y,y)$ by (3.8), one can backtrack to get $\bar{R}(x,y)$ by (3.7) then $R(x,y)$ by (3.5) and finally $P(x,y)$ by (3.3).

4. CONCLUSION

The probability generating function of the joint queue size is derived in a service delayed queueing system with two priority classes. Among the possible research directions:

- study the continuous time version of the process considered here.
- check the conservation law that the intensity of the system is equal to the server capacity, as in most queueing systems.
- assume a different service distribution for each class.
- consider more than two priority classes.

Extensions of this model are as numerous as the variants of the service delayed queueing system. One may consider, for example:

- a bulk arrival process
- a modulated input process
- a state dependence of the service process
- an hysteretic control policy
- an N-policy
- a vacationing server
- a random server capacity
- customer impatience, *etc.*

References

[1] Choi, B. D. and Lee, Y. (1998). $\mathrm{Geo}^{X_1}/\mathrm{Geo}^{X_2}/\mathrm{D}/c$ HOL priority queueing system with random order selection within each priority class, *Probability in the Engineering and Informational Sciences*, **12**, 125–139.

[2] Dshalalow, J. H. (1992). A first passage time problem and its applications to the analysis of a class of stochastic models, *Journal of Applied Mathematics and Stochastic Analysis*, **5**(1), 83–98.

[3] Dshalalow, J. H. (1995). Excess level processes in queueing, In: *Advances in Queueing: Theory, Methods and Open Problems* (Ed. Dshalalow, J. H.), pp. 243–262, CRC Press, Boca Raton.

[4] Dshalalow, J. H. (1997). Queueing systems with state dependent parameters, In: *Frontiers in Queueing: Models and Applications in Science and Engineering* (Ed. Dshalalow, J. H.), pp. 61–116, CRC Press, Boca Raton.

[5] Dshalalow, J. H. and Russell, G. (1992). On a single-server queue with fixed accumulation level, state dependent service, and semi-Markov modulated input flow, *International Journal of Mathematics and Mathematical Sciences*, **15**(3), 593–600.

[6] Neuts, M. F. (1989). *Structured Stochastic Matrices of M/G/1-Type and Their Applications*, Marcel Dekker, Inc., New York.

[7] Tadj, L. and Rikli, N.-E., Matrix analytic solution to a quorum queueing system, *Mathematical and Computer Modelling* (to appear).