Analysis on $G/M/1$ queue with two-stage service policy *

SUNGGON KIM[1], JONGWOO KIM[1], EUI YONG LEE [2]

ABSTRACT

We consider a $G/M/1$ queue with two-stage service policy. The server starts to serve with rate of μ_1 customers per unit time until the number of customers in the system reaches λ. At this moment, the service rate is changed to that of μ_2 customers per unit time and this rate continues until the system is empty. We obtain the stationary distribution of the number of customers in the system.

Keywords: $G/M/1$ queue; two-stage service policy; stationary distribution; number of customers

1. INTRODUCTION

In this paper, a $G/M/1$ queue is considered. We adopt two-stage service policy. The customers arrive according to a renewal process with inter arrival times following distribution function A of mean a. The server is initially idle. On an arrival of a customer, the server starts to serve μ_1 customers per unit time. Note that the service times are exponentially distributed. If the number of customers reaches λ, then the server immediately changes his service rate to μ_2 customers per unit time and finishes the current busy period, otherwise he finishes the busy period with service rate μ_1. The same service policy is applied to the forthcoming customers. We assume that $\max(\mu_1, \mu_2) < 1/a$ for the stability of the queue.

Kim [3] analysed the $M/M/1$ queue with the two-stage service policy and obtained the stationary distribution of the number of customers in system. Bae *et al.* [2] showed that there is an optimal service rate. In case $\mu_1 = 0$, the two-stage service policy become to be the N-policy, which was introduced by Yadin and Naor [6]. Recently, Zhang and Tian [7] obtained the stationary number of customers in a $G/M/1$ queueing system with the N-policy.

It is obtained in this paper the stationary distribution of the number of customers in the system. We, first, decompose the process of number of customers into two processes according to the service rate and obtain the stationary distribution of each process by

*This work was supported by KOSEF through grant No. R01-2004-000-10284-0 and the Statistical Research Center for Complex Systems at Seoul National University.

[1]Statistical Research Center for Complex Systems, Seoul National University, Seoul 151-747, Republic of Korea

[2]Department of Statistics, Sookmyung Women's University, Seoul 140-742, Republic of Korea

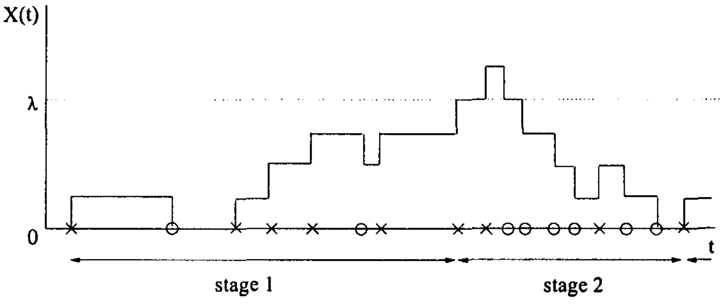

Figure 2.1: A sample path of queue length process $\{X(t),\, t \geq 0\}$

making use of an embedded Markov chain and the technique of level-crossing. We combine these two stationary distributions to obtain the stationary distribution of the original process.

2. Main Result

Let $\{X(t), t \geq 0\}$ be the process of the number of customers in the system. A sample path of $\{X(t), t \geq 0\}$ is illustrated in Figure 1. To obtain the stationary distribution of $\{X(t), t \geq 0\}$, we decompose $\{X(t), t \geq 0\}$ into two processes. First, the initial idle period is ignored. The system is called *in stage 1* as long as the service rate is kept μ_1 and is called *in stage 2* while the service rate is μ_2. The idle period after stage 2 is assumed to be still in stage 2. Meanwhile, the idle periods in the middle of stage 1 is assumed to be in stage 1. $\{X_1(t), t \geq 0\}$ is formed by separating the periods in stage 1 from $\{X(t), t \geq 0\}$ and connecting them together. The rests of $\{X(t), t \geq 0\}$ are connected to form $\{X_2(t), t \geq 0\}$.

2.1. Analysis of $\{X_1(t), t \geq 0\}$

Let $Y_n, n = 1, 2, \ldots$, be the number of customers in the system seen by the n-th arrival in $\{X_1(t), t \geq 0\}$. Then, $\{Y_n, n \geq 1\}$ is an aperiodic and irreducible Markov chain with state space $\{0, 1, 2, \ldots, \lambda - 1\}$ and the following transition matrix;

$$P_1 = \begin{bmatrix} r_0 & q_0 & 0 & \ldots & 0 \\ r_1 & q_1 & q_0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{\lambda-2} & q_{\lambda-2} & q_{\lambda-3} & \cdots & q_0 \\ r_0 & q_0 & 0 & \ldots & 0 \end{bmatrix},$$

where

$$q_k = \int_0^\infty \frac{e^{-\mu_1 t}(\mu_1 t)^k}{k!}\, dA(t), \quad k = 0, 1, 2, \ldots,$$

and $r_n = 1 - \sum_{k=0}^{n} q_k$. Observe that the last low in P_1 is the same as the first low. This is due to that $\{X_1(t), t \geq 0\}$ becomes 1 immediately after an arrival seeing $\lambda - 1$ customers.

Since $\{Y_n, n \geq 1\}$ is an irreducible Markov chain with finite state space, it is ergodic. The stationary distribution $\pi^{(1)} = (\pi_0^{(1)}, \pi_1^{(1)}, \ldots, \pi_{\lambda-1}^{(1)})$ is given by

$$\pi_i^{(1)} = \frac{D_{i+1}}{\sum_{j=1}^{\lambda} D_j}.$$

In the above equation, D_i is the determinant of the matrix obtained by striking out the i-th row and the i-th column of $I^{(\lambda)} - P_1$ and $I^{(\lambda)}$ is the λ-dimensional identity matrix. See Barlow and Proschan ([1], p129).

The process $\{X_1(t), t \geq 0\}$ is a regenerative process, where the regeneration points are the epochs of arrival seeing $\lambda - 1$ customers in the system. The following theorem gives the expectation of T_1, a cycle length between two regeneration points.

THEOREM 2.1. *Let $m_{ij}^{(1)}$, $i, j = 0, 1, 2, \ldots, \lambda-1$, be the expected time from an arrival seeing i customers to the next arrival seeing j customers in $\{X_1(t), t \geq 0\}$. Then,*

$$m_{ii}^{(1)} = \frac{a}{\pi_i^{(1)}}.$$

Thus, the expectation of T_1 is given by $E[T_1] = a/\pi_{\lambda-1}^{(1)}$

The expectation of T_1 is finite. Hence, $\{X_1(t), t \geq 0\}$ has a stationary distribution $\mathbf{p}^{(1)} = (p_0^{(1)}, p_1^{(1)}, \ldots, p_{\lambda-1}^{(1)})$.

Since the service times of customers are independent and exponentially distributed, by PASTA, the probability that a departure leaves $j-1$ customers in the system is $p_j^{(1)}$, for $j = 1, 2, \ldots$. This implies that the average number of departures leaving $j-1$ customers in a unit time is $\mu_1 p_j^{(1)}$. Moreover, notice that there is only 1 customer in the system immediately after the arrival seeing $\lambda - 1$ customers and that the rate of occurrence of this event is $\nu\pi_{\lambda-1}^{(1)}$, where $\nu = 1/a$. Thus, we can see that for $j = 2, 3, \ldots, \lambda - 1$,

$$\text{down-crossing rate to level } \left(j - \frac{1}{2}\right) = \mu_1 p_j^{(1)} + \nu\pi_{\lambda-1}^{(1)}, \tag{2.1}$$

and

$$\text{down-crossing rate to level } \frac{1}{2} = \mu_1 p_1^{(1)}. \tag{2.2}$$

We can also obtain the up-crossing rates from the definition of $\pi_j^{(1)}$'s. For $j = 1, 2, \ldots, \lambda - 1$,

$$\text{up-crossing rate to level } \left(j - \frac{1}{2}\right) = \nu\pi_{j-1}^{(1)}. \tag{2.3}$$

Since the process $\{X_1(t), t \geq 0\}$ has a stationary distribution, the up-crossing rate to a level $(j - 1/2)$ should be equal to the down-crossing rate to the same level, for $j = 1, 2, \ldots, \lambda - 1$. By equating these rates in eqs. (2.1), (2.2), and (2.3), we can derive that

$$p_1^{(1)} = \rho_1 \pi_0^{(1)},$$
$$p_j^{(1)} = \rho_1(\pi_{j-1}^{(1)} - \pi_{\lambda-1}^{(1)}), \quad j = 2, \ldots, \lambda - 1,$$

where $\rho_1 = \nu/\mu_1$. Using the identity that $\sum_{j=0}^{\lambda-1} p_j^{(1)} = 1$, we obtain that

$$p_0^{(1)} = 1 - \rho_1 + \rho_1(\lambda - 1)\,\pi_{\lambda-1}^{(1)}.$$

2.2. Analysis of $\{X_2(t), t \geq 0\}$

Let $Z_n, n = 1, 2, \ldots$, be the number of customers in the system seen by the n-th arriving customer in $\{X_2(t), t \geq 0\}$. Then, $\{Z_n, n \geq 1\}$ is an aperiodic and irreducible Markov chain and its state space is the set of all non-negative integers. We define q_k as follows:

$$q'_k = \int_0^\infty \frac{e^{-\mu_2 t}(\mu_2 t)^k}{k!}\, dA(t), \quad k = 0, 1, 2, \ldots,$$

and let $r'_n = 1 - \sum_{k=0}^{n} q'_k$. Then, the transition matrix of $\{Z_n, n \geq 1\}$ is given by

$$P_2 = \begin{bmatrix} r'_{\lambda-1} & q'_{\lambda-1} & q'_{\lambda-2} & q'_{\lambda-3} & \cdots \\ r'_1 & q'_1 & q'_0 & 0 & \cdots \\ r'_2 & q'_2 & q'_1 & q'_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

Note that the first low is the same as the λ-th low in P_2. This is due to that $\{X_2(t), t \geq 0\}$ becomes λ immediately after an arrival seeing the system empty. Using the Pakes's lemma [4], we can obtain the following theorem:

THEOREM 2.2. *The Markov chain* $\{Z_n, n \geq 1\}$ *is ergodic.*

By the above theorem, the stationary distribution $\boldsymbol{\pi}^{(2)} = \{\pi_0^{(2)}, \pi_1^{(2)}, \pi_2^{(2)} \ldots\}$ of the Markov chain $\{Z_n, n \geq 1\}$ exists. To obtain an explicit form of $\pi_i^{(2)}$'s, we need the following lemma and theorem.

LEMMA 2.1. *Let* $P_2^{(n)}$ *be the submatrix of* P_2 *having the same first* n *rows and* n *columns as* P_2 *and let* $I^{(n)}$ *is the* n*-dimensional identity matrix. Then,* $M^{(\lambda+1)} = I^{(\lambda+1)} - P_2^{(\lambda+1)}$ *is invertible.*

THEOREM 2.3. *Let* α *be the unique solution of* $x = \sum_{k=0}^{\infty} q_k x^k$ *and let* $\mathbf{b}$ *be a* $(\lambda+1)$*-dimensional vector such that*

$$\mathbf{b} = c\,\boldsymbol{\eta}\,[M^{(\lambda+1)}]^{-1},$$

where $\boldsymbol{\eta} = (1-\alpha)(r_{\lambda-1} - r_0, q_{\lambda-1} - q_0, q_{\lambda-2}, \ldots, q_0)$, $c = 1/(1+\boldsymbol{\eta}\,[M^{(\lambda+1)}]^{-1}\mathbf{1}')$, *and* $\mathbf{1}'$ *is the transpose of* $(\lambda+1)$*-dimensional vector* $(1,1,\ldots,1)$. *Then, the explicit form of* $\pi^{(2)}$ *is given by*

$$\pi_j^{(2)} = \begin{cases} b_j + c(1-\alpha)\alpha^j, & 0 \le j \le \lambda \\ c(1-\alpha)\alpha^j, & j \ge \lambda+1. \end{cases}$$

The process $\{X_2(t), t \ge 0\}$ is a regenerative process, where the regeneration points are the epochs of arrival seeing the system empty. Let $m_{ij}^{(2)}$, $i,j = 0,1,2,\ldots$, be the expected time from an arrival seeing i customers to the next arrival seeing j customers in $\{X_2(t),\ t \ge 0\}$. Then, by the similar argument to obtain Theorem 2.1, we can obtain that

$$m_{ii}^{(2)} = \frac{a}{\pi_i^{(2)}}.$$

If we let T_2 be the length of a cycle of $\{X_2(t),\ t \ge 0\}$, then the expectation of T_2 is given by $m_{00}^{(2)}$. Thus,

$$E[T_2] = \frac{a}{\pi_0^{(2)}}.$$

Since $\{X_2(t), t \ge 0\}$ is a regenerative process with finite expected cycle length, the process has a stationary distribution $\mathbf{p}^{(2)} = (p_0^{(2)}, p_1^{(2)}, p_2^{(2)} \ldots)$.

By the same reason as was mentioned in the previous subsection, the average number of departures leaving $j-1$ customers in a unit time is $\mu_2 p_j^{(2)}$. That is, for $j = 1,2,3,\ldots$

$$\text{down-crossing rate to level } \left(j - \frac{1}{2}\right) = \mu_2 p_j^{(2)}. \tag{2.4}$$

By definition, the probability that an arriving customer sees $j-1$ customers in the system is $\pi_{j-1}^{(2)}$. Moreover, there are λ customers in the system just after the arrival seeing the system empty. Hence, the up-crossing rates are given as follows:

$$\text{up-crossing rate to level } \left(j - \frac{1}{2}\right) = \begin{cases} \nu\pi_0^{(2)}, & j = 1, \\ \nu(\pi_{j-1}^{(2)} + \pi_0^{(2)}), & 2 \le j \le \lambda, \\ \nu\pi_{j-1}^{(2)}, & j > \lambda+1. \end{cases} \tag{2.5}$$

By equating the up- and down-crossing rates in eqs. (2.4) and (2.5), we have that

$$p_j^{(2)} = \begin{cases} \rho_2\pi_0^{(2)}, & j = 1, \\ \rho_2(\pi_{j-1}^{(2)} + \pi_0^{(2)}), & 2 \le j \le \lambda, \\ \rho_2\pi_{j-1}^{(2)}, & j \ge \lambda+1. \end{cases} \tag{2.6}$$

From the identity that $\sum_{j=0}^{\infty} p_j^{(2)} = 1$, we can derive that

$$p_0^{(2)} = 1 - \rho_2 - \rho_2(\lambda - 1)\pi_0^{(2)}.$$

2.3. Stationary distribution of $\{X(t), t \geq 0\}$

We are now ready to obtain the stationary distribution of $\{X(t), t \geq 0\}$, $\mathbf{p} = (p_0, p_1, p_2, \ldots)$, where $p_i = \lim_{t\to\infty} \Pr\{X(t) = i\}$. Note that the epochs where the number of customers reaches λ after the system being empty form a sequence of regeneration points in $\{X(t), t \geq 0\}$. The length of a cycle between two successive regeneration points is $T = T_1 + T_2$. We now assume that a reward at rate of one is given per unit time while $\{X(t), t \geq 0\}$, $\{X_1(t), t \geq 0\}$, and $\{X_2(t), t \geq 0\}$ stay at state i, $i = 0, 1, 2, \ldots$. By applying the renewal reward theorem(Ross [5, p.133]) repeatedly, we have that

$$\begin{aligned} p_i &= \frac{E[\text{reward during a cycle}T]}{E[T]} \\ &= \frac{E[\text{reward during a cycle}T_1] + E[\text{reward during a cycle}T_2]}{E[T]} \\ &= \frac{E[T_1]}{E[T]} \frac{E[\text{reward during a cycle}T_1]}{E[T_1]} + \frac{E[T_2]}{E[T]} \frac{E[\text{reward during a cycle}T_2]}{E[T_2]} \\ &= \frac{E[T_1]}{E[T_1] + E[T_2]} p_i^{(1)} + \frac{E[T_2]}{E[T_1] + E[T_2]} p_i^{(2)}, \end{aligned}$$

where $p_i^{(1)} = 0$ for i larger than $\lambda - 1$.

References

(1) R. G. Barlow and F. Proschan, *Mathematical Theory of Reliability*, Wiley, New York, 1965.

(2) J. Bae, J. Kim, and E. Y. Lee, An optimal service rate in a Poisson arrival queue with two-stage service policy, Mathematical Methods of Operations Research **58**(2003) 477-482.

(3) J. Kim, Possion Arrival Queue with Two-Stage Service Policy, Ph.D. thesis, POSTECH, 2004.

(4) A. G. Pakes, Some conditions for ergodicity and recurrence of Markov chains, Operations Research **17**(1969) 1058–1061.

(5) S.M. Ross, Stochastic Processes, 2nd ed., Wiley, New York, 1996.

(6) M. Yandin and P. Naor, Queueing systems with a removable service station, Operationas Research Quarterly **14**(1963) 393-405.

(7) Z. G. Zhang and N. Tian, The N threshold policy for the $GI/M/1$ queue, Operations Research Letters 32(2004) 77-84.