An $M/G/1$ Queue under the P_λ^M Policy with a Setup Time*

Jiyeon Lee† Jongwoo Kim‡

Abstract

We consider the P_λ^M service policy for an $M/G/1$ queue in which the service rate is increased from 1 to M at the exponential setup time after the level of workload exceeds λ. The stationary distribution of the workload is explicitly obtained through the level crossing argument.

Keywords: P_λ^M policy; $M/G/1$ queue; Stationary distribution; Setup time

1 Introduction

In this paper, we analyze the workload process of an $M/G/1$ queue under the P_λ^M policy with a random setup time. As in the ordinary $M/G/1$ queue, a server is initially idle and starts to work with service rate 1 if a customer arrives. The customers arrive according to a Poisson process of rate $\nu > 0$ and the service times of customers are independently and identically distributed with distribution function G and mean m. We assume that $\rho \equiv \nu m < 1$. If the workload of the system exceeds the level $\lambda(> 0)$, then after the random setup time is elapsed the server increases his service rate to $M(\geq 1)$ and continues to serve at rate M until he becomes idle. Before the setup time is finished, if the server is free, then he keeps his service rate to be 1. The setup time, denoted by S, is exponentially distributed with mean $1/\xi$ and independent of the workload. The server restarts to work with service rate 1, if another customer arrives.

We, in the next section, derive an explicit formula for the stationary distribution of the workload by using the level crossing argument of Brill and Posner [3] and Cohen [4].

*The first author was supported by Korea Research Foundation Grant(KRF-2004-002-C00056) and the second author was supported by KOSEF through Statistical Research Center for Complex Systems at Seoul National University.

†Department of Statistics, Yeungnam University, Kyungsan 712-749, Korea

‡Statistical Research Center for Complex Systems, Seoul National University, Seoul 151-742, Korea

2 Analysis of the Workload Process

Let $\mathbf{X} = \{X(t), t \geq 0\}$ be the workload process under the service policy described in the previous section. Then, $\mathbf{X}$ is regenerated at the time epochs when the server starts to work. We denote by C the length of a cycle, the interval between two successive regeneration points. Note that $\mathbf{X}$ is not, however, a Markov process. To analyze $\mathbf{X}$, we first decompose it into three processes. Let $\mathbf{X}_1$ be a process obtained from $\mathbf{X}$ by separating the periods of service rate 1, which start at the beginning of the busy period and end at the first exit from $(0, \lambda]$, and then connecting these together. Let $\mathbf{X}_2$ be formed by separating and connecting the residual busy periods of $\mathbf{X}$. Finally $\mathbf{X}_3$ is formed by connecting the idle periods of $\mathbf{X}$, that is, $\mathbf{X}_3 \equiv 0$. Clearly, both $\mathbf{X}_1$ and $\mathbf{X}_3$ are regenerative Markov processes whereas $\mathbf{X}_2$ is the only regenerative process. In these processes, we will call each separated segment a cycle of each process. Denote by C_i the length of a cycle in $\mathbf{X}_i$, for $i = 1, 2, 3$.

Let F_i be the stationary distribution function of $\mathbf{X}_i$ for $i = 1, 2, 3$ and F the stationary distribution function of $\mathbf{X}$. By applying the renewal reward theorem (Ross [5, p.133]), we can show that, for $x \geq 0$,

$$F(x) = p\frac{E[C_1]}{E[C]}F_1(x) + q\frac{E[C_2]}{E[C]}F_2(x) + \frac{1/\nu}{E[C]},$$

where p and q are the probabilities that there exist $\mathbf{X}_1$ and $\mathbf{X}_2$ in a cycle of $\mathbf{X}$, respectively. Note that $E[C] = pE[C_1] + qE[C_2] + 1/\nu$. We can easily see that $p = G(\lambda)$, since there exists $\mathbf{X}_1$ in a cycle of $\mathbf{X}$ if and only if the workload brought by the first customer after the idle period is less than or equal to λ. On the other hand, we can observe that q is the same as the probability that $\mathbf{X}$ crosses over the level λ during a cycle. Let $\mathbf{W} = \{W(t), t \geq 0\}$ be the workload process of the ordinary $M/G/1$ queue with ν the arrival rate of the customers and G the distribution function of the service times with mean m. Because $\mathbf{X}$ coincides with $\mathbf{W}$ until $\mathbf{X}$ upcrosses λ during a cycle, the probability q is given by $q = 1 - P\{D_\lambda = 0\} = H'_\rho(\lambda)/(\nu H_\rho(\lambda))$, where D_λ denotes the number of downcrossings of λ during a cycle of $\mathbf{W}$ of which distribution is obtained in (A.4), and the definition of H_ρ is given as follows:

$$H_\rho(x) = \sum_{n=0}^{\infty} \rho^n G_e^{*n}(x),$$

where $G_e^{*n}(x)$ is the n-fold recursive Stieltjes convolution of $G_e(x) = \int_0^x (1 - G(u))du/m$ with G_e^{*0} being Heaviside function.

In the following subsections, we evaluate the stationary distributions F_1 and F_2. The expected values $E[C_1]$ and $E[C_2]$ are also obtained.

2.1 Stationary distribution of $\mathbf{X}_1$

Let f_1 be the probability density function of the stationary distribution of $\mathbf{X}_1$ and $D_x^{(1)}$ denote the number of downcrossings of x during a cycle of $\mathbf{X}_1$. By the level crossing argument, we have

$$f_1(x) \equiv \frac{d}{dx}F_1(x) = \frac{E[D_x^{(1)}]}{E[C_1]}, \quad 0 < x \le \lambda.$$

Observe that $\mathbf{X}_1$ in a cycle coincides with $\mathbf{W}$ until $\mathbf{W}$ either crosses over λ or reaches 0, provided that both processes start at the same level. Therefore, if we denote by $D_{0\lambda;yx}$ the number of downcrossings of x until $\mathbf{W}$, starting at y $(0 < y \le \lambda)$, either crosses over λ or reaches 0, then $E[D_x^{(1)}]$ can be calculated by conditioning on the starting level y of a cycle of $\mathbf{X}_1$ as follows:

$$E[D_x^{(1)}] = \int_0^\lambda E[D_{0\lambda;yx}]\frac{dG(y)}{G(\lambda)}, \quad 0 < x \le \lambda.$$

Substituting $E[D_{0\lambda;yx}]$ obtained in (A.2) and recalling $\int_0^\lambda f_1(x)dx = 1$ yield

$$E[C_1] = \frac{1}{\nu G(\lambda)}\left(H_\rho(\lambda) - 1 - \frac{H_\rho'(\lambda)}{H_\rho(\lambda)}\int_0^\lambda H_\rho(x)dx\right).$$

2.2 Stationary distribution of $\mathbf{X}_2$

First, we define by L, the exceeding amount of starting level over λ in a cycle of $\mathbf{X}_2$. By using the Markovian property of $\mathbf{X}_1$, Bae et al. [1] derived the probability density function $r(l)$ of L given by

$$r(l) \equiv -\frac{d}{dl}P\{L > l\} = -\frac{d}{dl}\left(\frac{1 - G(\lambda + l) + \int_0^\lambda P(l,t)dG(t)}{H_\rho'(\lambda)/\nu H_\rho(\lambda)}\right),$$

where $P(l,t) = c_l H_\rho(\lambda - t) - \rho\int_0^{\lambda - t} J_l(\lambda - t - u)dH_\rho(u)$, $c_l = \rho(H_\rho * J_l)(\lambda)/H_\rho(\lambda)$, and $J_l(x) = G_e(x + l) - G_e(l)$. We notice that the starting levels of each cycle in $\mathbf{X}_2$ are independent and have the same distribution as the random variable $\lambda + L$.

Let $C_2(y)$ be the length of cycle of $\mathbf{X}_2$ when it starts at level y $(y > \lambda)$. We denote by T_y the elapsed time to reach the level 0 when $\mathbf{W}$ starts at y. Then, we have that

$$C_2(y) \stackrel{\mathcal{D}}{=} \begin{cases} S + C_2'(y) & T_y > S, \\ T_y & T_y \le S, \end{cases}$$

where $C_2'(y)$ denotes the rest period of $C_2(y)$ excluding the setup time and $\stackrel{\mathcal{D}}{=}$ denotes the equality in distribution. Obviously during $C_2'(y)$ the customers are served at rate M. Notice that the starting level of the period $C_2'(y)$ in $\mathbf{X}_2$, denoted by W_y, can be expressed by

$$W_y \stackrel{\mathcal{D}}{=} W(S) \mid T_y > S, W(0) = y$$

and so its LST is given by

$$E[e^{-\theta W_y}] = \frac{E[e^{-\theta W(S)} \mid W(0) = y] - E[e^{-\theta W(S)} \cdot 1_{\{T_y \le S\}} \mid W(0) = y]}{1 - \tilde{T}_y(\xi)}, \qquad (1)$$

where $\tilde{T}_y(\xi) \equiv E[e^{-\xi T_y}]$, the LST of T_y, which is derived in (A.6).

Note that, by the memoryless property of the exponential distribution of S, the conditional distribution of $W(S)$, given that $T_y \leq S$ and $W(0) = y$, is equal to the unconditional distribution of $W(S)$, where $\mathbf{W}$ starts at 0. Hence,

$$E[e^{-\theta W(S)} \cdot 1_{\{T_y \leq S\}} \mid W(0) = y] = \tilde{T}_y(\xi) E[e^{-\theta W(S)} \mid W(0) = 0]. \tag{2}$$

Adopting the results in Boxma et al. [2] yields that for $z \geq 0$,

$$E[e^{-\theta W(S)} \mid W(0) = z] = \frac{\xi}{\xi - \varphi(\theta)} \left(e^{-\theta z} - \frac{\theta e^{-\theta_0(\xi) z}}{\theta_0(\xi)} \right), \quad \theta \geq 0, \tag{3}$$

where $\varphi(\theta) = \theta - \nu + \nu \tilde{G}(\theta)$ with $\tilde{G}(\theta) = \int_0^\infty e^{-\theta x} dG(x)$, the LST of G, and $\theta_0(\xi)$ is the solution to the equation $\varphi(\theta) = \xi$. Substituting (2) into (1) and using (3) result in the following LST of W_y:

$$E[e^{-\theta W_y}] = \frac{\xi \left\{ \theta_0(\xi) e^{-\theta y} - \theta e^{-\theta_0(\xi) y} - \tilde{T}_y(\xi)[\theta_0(\xi) - \theta] \right\}}{\theta_0(\xi)(\xi - \varphi(\theta))(1 - \tilde{T}_y(\xi))},$$

from which we can calculate both the expected value of W_y and the distribution function $K_y(w)$ of W_y.

Notice that if we change the scale of time by considering $1/M$ as a unit time, then the arrival rate of $\mathbf{X}_2$ during the period $C_2'(y)$ becomes ν/M, the service rate 1, and the traffic intensity $\rho' \equiv \nu m/M$. Then, it follows from (A.7) that

$$E[C_2'(y) \mid T_y > S] = \frac{1}{M} \frac{E[W_y]}{1 - \rho'} = \frac{E[W_y]}{M - \nu m}.$$

Thus, the expected value of $C_2(y)$ is given by

$$E[C_2(y)] = \left(1 - \tilde{T}_y(\xi)\right) \left(\frac{1}{\xi} + \frac{E[W_y]}{M - \nu m} \right).$$

Conditioning on the starting level y, we have

$$E[C_2] = \int_\lambda^\infty E[C_2(y)] r(y - \lambda) dy.$$

Now, we need to investigate $E[\int_0^{C_2} 1_{\{X_2(t) \leq x\}} dt]$ in order to get the stationary density function of $\mathbf{X}_2$ given by

$$f_2(x) \equiv \frac{d}{dx} F_2(x) = \frac{\frac{d}{dx} E[\int_0^{C_2} 1_{\{X_2(t) \leq x\}} dt]}{E[C_2]}. \tag{4}$$

Let $D_{yx}(t)$ denote the number of downcrossings of x during the time interval $(0, t]$ by $\mathbf{W}$ which starts at y and let D_{yx} be the number of downcrossings of x until $\mathbf{W}$, starting at y, returns to 0. Then,

$$\frac{d}{dx} \int_0^{C_2(y)} 1_{\{X_2(t) \leq x\}} dt \overset{\text{a.s.}}{=} \begin{cases} D_{yx} & T_y \leq S, \\ D_{yx}(S) + D_x(C_2'(y), \rho')/M & T_y > S, \end{cases}$$

where $D_x(C_2'(y), \rho')$ denotes the number of downcrossings of x by $\mathbf{W}$, which starts at y, during the residual busy period $C_2'(y)$, of which the traffic intensity is ρ' and $\overset{\text{a.s.}}{=}$ denotes the almost sure equality. By taking expectations and applying the dominated convergence theorem, we obtain

$$\begin{aligned}
&\frac{d}{dx}E\left[\int_0^{C_2(y)} 1_{\{X_2(t)\le x\}}dt\right] \\
&= E[D_{yx}] + \left(1-\tilde{T}_y(\xi)\right)\left\{\frac{1}{M}H_{\rho'}(x) - H_\rho(x)\right. \\
&\qquad\qquad \left. - \int_0^x \left(\frac{1}{M}H_{\rho'}(x-w) - H_\rho(x-w)\right)dK_y(w)\right\}. \qquad (5)
\end{aligned}$$

Hence, conditioning on the starting level y and employing the dominated convergence theorem, we get an expression for the numerator of (4) as

$$\int_\lambda^\infty \frac{d}{dx}E\left[\int_0^{C_2(y)} 1_{\{X_2(t)\le x\}}dt\right] r(y-\lambda)dy,$$

where $\frac{d}{dx}E[\int_0^{C_2(y)} 1_{\{X_2(t)\le x\}}dt]$ has been derived in (5).

A Appendix

A.1 The number of downcrossings in $M/G/1$ queue

We define by $P_{0\lambda;yx}$ the probability that $\mathbf{W}$ downcrosses x before either crossing over λ or reaching 0 when the process starts at y, $0 < x, y \le \lambda$. Bae et al. [1] showed that

$$P_{0\lambda;yx} = \begin{cases} \frac{H_\rho(\lambda-y)}{H_\rho(\lambda-x)} & 0 < x < y \le \lambda, \\ \frac{H_\rho(\lambda-y)}{H_\rho(\lambda-x)} - \frac{H_\rho(\lambda)H_\rho(x-y)}{H_\rho(\lambda-x)H_\rho(x)} & 0 < y \le x \le \lambda. \end{cases} \qquad (A.1)$$

Since

$$P\{D_{0\lambda;yx} = n\} = \begin{cases} 1 - P_{0\lambda;yx} & n = 0, \\ P_{0\lambda;yx}(P_{0\lambda;xx})^{n-1}(1 - P_{0\lambda;xx}) & n \ge 1, \end{cases}$$

it follows from (A.1) that

$$E[D_{0\lambda;yx}] = \begin{cases} \frac{H_\rho(x)H_\rho(\lambda-y)}{H_\rho(\lambda)} & 0 < x < y \le \lambda, \\ \frac{H_\rho(x)H_\rho(\lambda-y)}{H_\rho(\lambda)} - H_\rho(x-y) & 0 < y \le x \le \lambda. \end{cases} \qquad (A.2)$$

If we let $P_{0;yx} \equiv \lim_{\lambda\to\infty} P_{0\lambda;yx}$, then it represents the probability of returning to level x without reaching 0 when starting at y. When $y > x$, D_{yx} follows the geometric distribution with parameter $1 - P_{0;xx}$ and hence, $E[D_{yx}] = H_\rho(x)$, because $H_\rho(0) = 1$ and $\lim_{x\to\infty} H_\rho(x) = 1/(1-\rho)$. When $y \le x$,

$$P\{D_{yx} = n\} = \begin{cases} \frac{H_\rho(x-y)}{H_\rho(x)} & n = 0, \\ \left(1 - \frac{H_\rho(x-y)}{H_\rho(x)}\right)\left(1 - \frac{1}{H_\rho(x)}\right)^{n-1}\left(\frac{1}{H_\rho(x)}\right) & n \ge 1, \end{cases} \qquad (A.3)$$

so that

$$E[D_{yx}] = H_\rho(x) - H_\rho(x-y).$$

Conditioning on the workload brought by the first customer and noting, for $y > x$, $P\{D_{yx} = 0\} = 0$, from (A.3) we have

$$P\{D_x = 0\} = 1 - \frac{H'_\rho(x)}{\nu H_\rho(x)}. \tag{A.4}$$

A.2 The first hitting time of $M/G/1$ queue

Now, we need the distribution of T_y. From the Markov property of $\mathbf{W}$, we can see that

$$T_y \overset{\mathcal{D}}{=} \begin{cases} y & N(y) = 0, \\ y + \sum_{i=1}^{N(y)} B_i & N(y) \geq 1, \end{cases} \tag{A.5}$$

where $N(y)$ is the number of customers who arrive during the time y, which is the Poisson random variable with parameter νy, and B_i denotes the busy period of the $M/G/1$ queue. It is well known that the LST of B_i, denoted by $\tilde{B}(\theta)$, is the solution to the equation $\tilde{B}(\theta) = \tilde{G}(\theta + \nu - \nu\tilde{B}(\theta))$ (Wolff [6, p.390]).

By Wald's equation (Ross [5, p.105]), it follows from (A.5) that the LST of T_y is given by

$$\tilde{T}_y(\theta) = \exp\left\{-(\theta + \nu - \nu\tilde{B}(\theta))y\right\}, \tag{A.6}$$

and the expected value of T_y is given by

$$E[T_y] = \frac{y}{1-\rho}. \tag{A.7}$$

References

[1] Bae, J., Kim, S. and Lee, E.Y. (2002) "A P_λ^M-policy for an $M/G/1$ queueing system", *Applied Mathematical Modelling*, **26**, 929-939.

[2] Boxma, O.J., Perry, D. and Stadje, W. (2001) "Clearing Models for $M/G/1$ Queues", *Queueing Systems*, **38**, 287-306.

[3] Brill, P.H. and Posner, M.J.M. (1977) "Level Crossings in Point Processes Applied to Queues: Single-Server Case", *Operations Research*, **25**, 662-674.

[4] Cohen, J.W. (1977) "On up- and downcrossings", *Journal of Applied Probability*, **14**, 405-410.

[5] Ross, S.M. (1996) *Stochastic Processes*, 2nd ed., John Wiley & Sons, Inc., New York.

[6] Wolff, R.W. (1989) *Stochastic Modeling and the Theory of Queues*, Prentice Hall, New Jersey.