

An Optimal P_λ^M -Service Policy for an $M/G/1$ Queueing System*

Jongho Bae[†] Jongwoo Kim[†] Eui Yong Lee[§]

Abstract

We consider an $M/G/1$ queueing system under P_λ^M -service policy. As soon as the workload exceeds threshold $\lambda > 0$, the service rate is increased from 1 to $M \geq 1$ and is kept until the system becomes empty. After assigning several costs, we show that there exists a unique M minimizing the long-run average cost per unit time.

Keywords: P_λ^M -service policy; $M/G/1$ queue; Optimal service rate

1 Introduction

Bae et al. [1] introduced a P_λ^M -service policy for an $M/G/1$ queueing system. Server starts to work with service rate 1 when a customer arrives. The arrival process of the customers is a Poisson process of rate $\nu > 0$ and the service times of customers are independent and identically distributed with distribution function G . The server increases his/her service rate to $M \geq 1$ instantaneously, if the workload exceeds threshold $\lambda > 0$, and keeps the service rate until the system becomes empty. Otherwise, the server finishes the busy period with service rate 1. The server restarts to work with service rate 1, if another customer arrives. Bae et al. [1] studied the workload process and obtained the stationary distribution of the workload process.

In this paper, we extended the earlier analysis by assigning costs related to the service rate, idle period, and workload, and then seeking to minimize the

*This work was supported by Korea Research Foundation Grant(KRF-2002-070-C00021).

[†]Department of Mathematics, Jeonju University, Jeonju 560-759, Korea

[‡]Department of Mathematics, Pohang University of Science and Technology, Pohang 790-784, Korea

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

long-run average cost by varying M . It is shown that there exists a unique M which minimizes the long-run average cost per unit time.

2 Long-run average cost per unit time

Let $\{X(t), t \geq 0\}$ be the workload process of $M/G/1$ queueing system with P_λ^M -service policy. Note that the time epochs when the server starts to work form embedded regeneration points of workload process. In this section, we calculate the long-run average cost per unit time, after assigning the following four costs to the system:

- $h(M)$: the operating cost per unit time while the service rate being M , $M \geq 1$.
- $g(M)$: the cost for increasing the service rate from 1 to M , $M \geq 1$.
- c_I : the penalty per unit time while the server being empty.
- c_H : the cost per unit time for the system holding a unit workload.

It is assumed that h and g are nonnegative, nondecreasing, and twice differentiable convex function including linear functions. We also assume that $h(1) = g(1) = 0$ and that neither h nor g is a constant function.

Let T be the length of a regeneration cycle in $\{X(t), t \geq 0\}$. Then, by the renewal reward theorem (Ross [2], p.133), the long-run average cost per unit time is given by

$$\frac{E[\text{cost during } T]}{E[T]}.$$

Note that T can be partitioned into three periods T_1 , T_2 , and T_3 , where T_1 is the period of service rate 1, T_2 of service rate M , and T_3 the idle period. Hence,

$$E[T] = \alpha E[T_1] + \beta E[T_2] + \frac{1}{\nu},$$

where α is the probability that there exists a period of service rate 1 and β of service rate M in a cycle of $\{X(t), t \geq 0\}$. By Bae et al. [1], $\alpha = G(\lambda)$, $\beta = \frac{H'(\lambda)}{\nu H(\lambda)}$, $E[T_1] = \frac{1}{\nu G(\lambda)} \left(H(\lambda) - 1 - \frac{H'(\lambda)}{H(\lambda)} \int_0^\lambda H(x) dx \right)$, and $E[T_2] = \frac{\lambda + E[L]}{M - \rho}$,

where $H(x) = \sum_{n=0}^{\infty} \rho^n G_e^{*n}(x)$, $\rho = \nu m$, the traffic intensity, $m = \int_0^{\infty} x dG(x)$, $G_e(x) = (1/m) \int_0^x (1 - G(u)) du$, the equilibrium distribution function of G , $*n$ is the n -fold recursive Stieltjes convolution with G_e^{*0} being Heaviside function, and L is the first exceeding amount of starting level over λ in a cycle of the period of service rate M . Therefore,

$$E[T] = A + \beta \frac{E[S_a]}{M - \rho},$$

where $A = \alpha E[T_1] + \frac{1}{\nu} = \frac{1}{\nu} \left(H(\lambda) - \frac{H'(\lambda)}{H(\lambda)} \int_0^{\lambda} H(x) dx \right)$, which is constant with respect to M , and $S_a = \lambda + L$. We assume $M > \rho$ and $M \geq 1$.

The expected costs during a cycle are as follows:

$$E[\text{operating cost during } T] = \beta E[T_2] h(M) = \beta E[S_a] \frac{h(M)}{M - \rho},$$

$$E[\text{cost for increasing the service rate during } T] = \beta g(M),$$

$$E[\text{penalty during idle period in } T] = \frac{c_I}{\nu},$$

and

$$E[\text{holding cost during } T]$$

$$= c_H (\alpha E[\text{total work during } T_1] + \beta E[\text{total work during } T_2]).$$

We can obtain average total works during T_1 and T_2 as follows:

$$\begin{aligned} & E[\text{total work during } T_1] \\ &= \frac{1}{\nu G(\lambda)} \left(\lambda H(\lambda) - \int_0^{\lambda} H(x) dx - \frac{H'(\lambda)}{H(\lambda)} \int_0^{\lambda} x H(x) dx \right), \end{aligned}$$

$$E[\text{total work during } T_2] = \frac{E[S_a^2]}{2(M - \rho)} + \frac{\nu E[S_a] E[S^2]}{2(M - \rho)^2},$$

where S is the service time of the first customer after an idle period. Finally, we obtain $C(M)$, the long-run average cost per unit time, as follows:

$$C(M) = \frac{\beta E[S_a] h(M) + (M - \rho) \{ \beta g(M) + c_H \beta u(M) + c_I / \nu + c_H \alpha B \}}{A(M - \rho) + \beta E[S_a]},$$

for $M > \rho$, $M \geq 1$, where

$$B = E[\text{total work during } T_1],$$

and

$$u(M) = E[\text{total work during } T_2].$$

3 Optimal fast service rate

We now show the uniqueness of the fast service rate which minimizes $C(M)$.

By differentiating $C(M)$ by M , we have

$$C'(M) = \frac{\beta E[S_a] N(M)}{(A(M - \rho) + \beta E[S_a])^2},$$

where

$$\begin{aligned} N(M) &= \left\{ \frac{A(M - \rho)}{E[S_a]} + \beta \right\} \{E[S_a]h'(M) + (M - \rho)(g'(M) + c_H u'(M))\} \\ &\quad - Ah(M) + \beta(g(M) + c_H u(M)) + \frac{c_I}{\nu} + c_H \alpha B. \end{aligned}$$

We show $N(M)$ is a strictly increasing function on M .

$$\begin{aligned} N'(M) &= \left\{ \frac{A(M - \rho)}{E[S_a]} + \beta \right\} \{E[S_a]h''(M) + 2g'(M) + (M - \rho)g''(M) \\ &\quad + c_H(2u'(M) + (M - \rho)u''(M))\}. \end{aligned}$$

Recall that $h''(M)$, $g'(M)$, and $g''(M)$ are nonnegative, and observe that

$$\begin{aligned} 2u'(M) + (M - \rho)u''(M) &= \{(M - \rho)u(M)\}'' \\ &= \frac{\nu E[S_a]E[S^2]}{(M - \rho)^3} > 0. \end{aligned}$$

Note that the sign of $C'(M)$ is the same as that of $N(M)$.

We investigate the sign of $N(1)$ in case that $M \geq 1 > \rho$ and that of $\lim_{M \rightarrow \rho^+} N(M)$ in case that $M > \rho \geq 1$. When $M \geq 1 > \rho$,

$$\begin{aligned} N(1) &= \left\{ \frac{A(1 - \rho)}{E[S_a]} + \beta \right\} \{E[S_a]h'(1) + (1 - \rho)g'(1)\} + \frac{c_I}{\nu} + c_H \alpha B \\ &\quad - c_H \left\{ A \left(\frac{E[S_a^2]}{2E[S_a]} + \frac{\nu E[S^2]}{1 - \rho} \right) + \frac{\nu \beta E[S_a]E[S^2]}{2(1 - \rho)^2} \right\}, \end{aligned}$$

which may assume positive or negative value. When $M > \rho \geq 1$,

$$\lim_{M \rightarrow \rho^+} N(M) = \beta E[S_a]h'(\rho) - Ah(\rho) + \beta g(\rho) + \frac{c_I}{\nu} + c_H \alpha B$$

$$\begin{aligned}
& +c_H \frac{A}{E[S_a]} \lim_{M \rightarrow \rho^+} (M - \rho)^2 u'(M) \\
& +c_H \beta \lim_{M \rightarrow \rho^+} ((M - \rho)u'(M) + u(M)) \\
& = -\infty,
\end{aligned}$$

since

$$\lim_{M \rightarrow \rho^+} (M - \rho)^2 u'(M) = \lim_{M \rightarrow \rho^+} \left(-\frac{E[S_a^2]}{2} - \frac{\nu E[S_a]E[S^2]}{(M - \rho)} \right) = -\infty,$$

and

$$\lim_{M \rightarrow \rho^+} ((M - \rho)u'(M) + u(M)) = \lim_{M \rightarrow \rho^+} \left(-\frac{\nu E[S_a]E[S^2]}{2(M - \rho)^2} \right) = -\infty.$$

In order to see the sign of $\lim_{M \rightarrow \infty} C'(M)$, notice that

$$C'(M) = \frac{\beta E[S_a]N(M)/(M - \rho)^2}{\{A + \beta E[S_a]/(M - \rho)\}^2}.$$

Now,

$$\begin{aligned}
\frac{N(M)}{(M - \rho)^2} & = A \frac{(M - \rho)h'(M) - h(M)}{(M - \rho)^2} + c_H \beta \left\{ \frac{u'(M)}{M - \rho} + \frac{u(M)}{(M - \rho)^2} \right\} \\
& + \frac{c_H A u'(M)}{E[S_a]} + \frac{A g'(M)}{E[S_a]} \\
& + \frac{\beta h'(M)/E[S_a] + \beta g(M) + c_I/\nu + c_H \alpha B}{(M - \rho)^2} + \frac{\beta g'(M)}{M - \rho}.
\end{aligned}$$

The last two terms of the right side of the above equation are nonnegative for $M > \rho$, and $\lim_{M \rightarrow \infty} A g'(M)/E[S_a] > 0$, since g is not constant. We have only to show the limits of the first three terms as M goes to infinity are nonnegative for the proof of $\lim_{M \rightarrow \infty} C'(M) > 0$. $(M - \rho)h'(M) - h(M)$ is nondecreasing function in $M > \rho$, since its derivative is $(M - \rho)h''(M)$. Hence, it is bounded or goes to infinity as M goes to infinity. If it is bounded,

$$\lim_{M \rightarrow \infty} \frac{(M - \rho)h'(M) - h(M)}{(M - \rho)^2} = 0,$$

and if it goes to infinity as M , the above limit goes to infinity, by the L'hospital's law,

$$\lim_{M \rightarrow \infty} \frac{(M - \rho)h'(M) - h(M)}{(M - \rho)^2} = \lim_{M \rightarrow \infty} \frac{h''(M)}{2} \geq 0.$$

Notice that

$$\lim_{M \rightarrow \infty} \left(\frac{u'(M)}{M - \rho} + \frac{u(M)}{(M - \rho)^2} \right) = \lim_{M \rightarrow \infty} \left(-\frac{\nu E[S_a]E[S^2]}{2(M - \rho)^4} \right) = 0,$$

and

$$\lim_{M \rightarrow \infty} u'(M) = \lim_{M \rightarrow \infty} \left(-\frac{E[S_a^2]}{2(M - \rho)^2} - \frac{\nu E[S_a]E[S^2]}{(M - \rho)^3} \right) = 0.$$

Since $\lim_{M \rightarrow \infty} C'(M) > 0$, $N(M) > 0$ for M large enough even though $\lim_{M \rightarrow \rho^+} N(M) = -\infty$ or $N(1) < 0$. Therefore, we have the following conclusions:

- When $M \geq 1 > \rho$,
 - If $N(1) \geq 0$, $C'(M) > 0$ for all $M > 1$ and hence $C(M)$ is minimized at $M = 1$.
 - If $N(1) < 0$, there exists M^* ($M^* > 1$) such that $C'(M) < 0$ for $M < M^*$ and $C'(M) > 0$ for $M > M^*$. Therefore, there exists unique M which minimizes $C(M)$.
- When $M > \rho \geq 1$, there exists unique M^* ($M^* > \rho$) such that $C(M)$ is minimized at $M = M^*$.

The M^* is the solution of the equation $N(M) = 0$.

References

- [1] Bae, J., Kim, S. and Lee, E.Y. (2002) A P_λ^M -Policy for an $M/G/1$ Queuing System, *Applied Mathematical Modelling*, **26**, 929–939.
- [2] Ross, S.M. (1996), *Stochastic Processes*, 2nd edition, John Wiley & Sons.
- [3] Wolff, R.W. (1988), *Stochastic Modeling and the Theory of Queues*, Prentice Hall.