

Conditional sojourn time distributions in $M/G/1$ and $G/M/1$ queues under P_λ^M -service policy

Sunggon Kim^{1,a}

^aDepartment of Statistics, University of Seoul, Korea

Abstract

P_λ^M -service policy is a workload dependent hysteretic policy. The policy has two service states comprised of the ordinary stage and the fast stage. An ordinary service stage is initiated by the arrival of a customer in an idle state. When the workload of the server surpasses threshold λ , the ordinary service stage changes to the fast service state, and it continues until the system is empty. These service stages alternate in this manner. When the cost of changing service stages is high, the hysteretic policy is more efficient than the threshold policy, where a service stage changes immediately into the other service stage at either case of the workload's surpassing or crossing down a threshold. P_λ^M -service policy is a modification of P_λ^M -policy proposed to control finite dams, and also an extension of the well-known D -policy. The distributions of the stationary workload of P_λ^M -service policy and its variants are studied well. However, there is no known result on the sojourn time distribution. We prove that there is a relation between the sojourn time of a customer and the first up-crossing time of the workload process over the threshold λ after the arrival of the customer. Using the relation and the duality of $M/G/1$ and $G/M/1$ queues, we obtain conditional sojourn time distributions in $M/G/1$ and $G/M/1$ queues under the policy.

Keywords: P_λ^M -service policy, hysteretic service policy, sojourn time distribution, duality

1. Introduction

We consider queueing systems under P_λ^M -service policy. The policy is a workload-dependent hysteretic service policy with two service stages of different service rates 1 and $M (> 1)$, i.e., there are the ordinary service stage with rate 1 and the fast service stage with rate M . The server starts to serve with rate 1 if a customer arrives to the server in an idle state. When the workload of the server reaches over λ , the service rate changes to M , and the rate continues until the system is empty. We obtain the conditional sojourn time distribution of a randomly chosen customer given the amount of work present in the server just after the arrival of the customer.

Bae *et al.* (2002) proposed P_λ^M -service policy as a control policy for the $M/G/1$ queue, and obtained the distribution of the stationary workload. The policy is a modification of P_λ^M policy proposed by Faddy (1974) to control optimally a finite dam with the flow of water into the reservoir formed by a Wiener process. In a finite dam under P_λ^M policy, the release rate of water is 0 initially, and the rate changes to M instantaneously as soon as the level reaches λ . Once the dam becomes empty, the release rate remains zero until the level λ is reached again.

P_λ^M -service policy can be considered as an extension of D -policy proposed by Balachandran (1973) as a control policy for single server queueing systems. Under the policy, the server is turned

¹ Department of Statistics, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul 02504, Korea.
E-mail: sgkim@uos.ac.kr

off at the end of a busy period, waits for the workload of the server to exceed some fixed value D , and starts the service. D -policy is identical to P_λ^M policy if we consider the level of water in a dam as the workload of the server, and let $D = \lambda$ and $M = 1$.

Kim *et al.* (2006) considered the problem of finding the optimal value of M . They considered 4 types of costs; a switching cost for increasing service rate, an operating cost incurred when service rate is M , a holding cost per unit of workload, and a penalty cost due to being idle of the server. They showed that there exists the unique optimal value of M minimizing sum of the costs. Lee and Kim (2007) considered an $M/G/1$ queue under P_λ^M -service policy with an exponentially distributed set-up time. They obtained the stationary workload distribution. Applying the duality between the $M/G/1$ and the $G/M/1$ queues, Kim and Bae (2008) obtained the distribution of the stationary workload in the $G/M/1$ queue under P_λ^M -service policy. When the workload capacity is finite, Hur *et al.* (2006) developed an approximation method for some performance measures such as average workload, stationary workload distribution, and blocking probability.

Lee and Kim (2006) studied a generalization of P_λ^M -service policy, where a fast service stage initiated by the up-crossing of level λ remains until the down-crossing of level τ instead of being empty of the server. The service rate depends on the workload and also on the service stage, i.e., there exist two service rate functions $r_1(x)$ for the ordinary service stage and $r_2(x)$ for the fast service stage. Under this service policy, they obtained the stationary workload distribution in an $M/G/1$ queue. The policy they considered is also a generalization of $P_{\lambda,\tau}^M$ policy proposed by Yeh (1985) as a control policy for a finite dam. In this policy, $r_1(x) = 0$ and $r_2(x) = M$.

There are some studies on $P_{\lambda,\tau}^M$ policy and its variants when input processes are more general than compound Poisson processes of $M/G/1$ queues. Abdel-Hameed (2000) studied $P_{\lambda,\tau}^M$ policy in an infinite dam when the input is a compound Poisson process with positive drift. Abdel-Hameed and Nakhi (2006) studied the optimal control of the finite dam under $P_{\lambda,\tau}^M$ policy when the input is a diffusion process and the release rate is state dependent. Abdel-Hameed (2011) analyzed an infinite dam under $P_{\lambda,\tau}^M$ policy when the input is an increasing Lévy process. Bekker (2009) considered a single server queue under a variant of $P_{\lambda,\tau}^M$ policy. He assumed Lévy input processes, and the Lévy exponents are different in each stage. He obtained the Laplace transform of the stationary workload.

Most studies on queues with workload dependent hysteretic service policy focus on obtaining a stationary workload distribution. In this paper, we obtain the conditional sojourn time distribution of a customer in $M/G/1$ and $G/M/1$ queues under P_λ^M -service policy when the workload of the server just after the arrival of the customer is given. We reveal that there is a relation between the sojourn time of a customer and the first up-crossing time of the workload process over the level λ after the arrival of the customer. From the relation and the duality of $M/G/1$ and $G/M/1$ queues, we obtain the conditional sojourn time distribution.

The paper is organized as follows. In Section 2, we represent the sojourn time distribution using the conditional sojourn time distribution and the steady state distribution of the workload. In Section 3, we show the relation between the conditional sojourn time of a customer and the first up-crossing time of the workload process over the level λ . In Section 4, we obtain the conditional sojourn time distributions of $M/G/1$ and $G/M/1$ queues under P_λ^M -service policy.

2. The conditional sojourn time

We consider a $G/G/1$ queuing system. Inter-arrival times of customers are independent and identically distributed (iid) with cumulative distribution function (cdf) $A(\cdot)$, and amounts of work brought by customers are also iid with cdf $B(\cdot)$. We assume that $A(0) = 0$ and $B(0) = 0$. Let α and β denote

expectations of the inter-arrival times and the amounts of work, respectively. We define the workload of the server at time t as the amount of work present in the queueing system at time t . We choose random a customer, and shift the arrival time point of the customer to 0. We call this tagged customer. Let Y be the amount of work present in the server before time 0, and let W_0 be the amount of work brought by the tagged customer. If we define $X = Y + W_0$, then T , the sojourn time of the tagged customer, is defined as

$$T = \min \left\{ t \mid \int_0^t r(u) du = X \right\},$$

where $r(t)$ is the service rate at time t . We define T_x as the conditional sojourn time of the tagged customer when $X = x$.

If $x > \lambda$, then the service rate is M until the system is empty. Thus,

$$T_x = \frac{x}{M}, \quad x > \lambda. \quad (2.1)$$

For the case of $0 < x \leq \lambda$, the service rate is 1 at time 0. Depending on the evolution of the workload process after time 0, the rate can be changed to M before the departure of the tagged customer. Thus, T_x is not a deterministic function of x , but a random variable dependent on x . We denote by $F(t|x)$ the cdf of T_x , $x > 0$.

Since Y and W_0 are independent, the cdf of X is the Stieltjes convolution of $F_Y(\cdot)$ and $B(\cdot)$, i.e.,

$$F_X(x) = \int_0^x B(x-y) dF_Y(y), \quad x > 0.$$

Then, the stationary sojourn time distribution is given by

$$\Pr\{T \leq t\} = \int_0^{\infty} F(t|x) dF_X(x). \quad (2.2)$$

Due to equation (2.1), $T_x = x/M$ for $x > \lambda$. Then, we have

$$\Pr\{T \leq t\} = \int_0^{\lambda} F(t|x) dF_X(x) + \int_{\lambda}^{\infty} I\left(\frac{x}{M} \leq t\right) dF_X(x), \quad (2.3)$$

where $I(\cdot)$ is the indicator function. The second term in the right hand side of equation (2.3) is evaluated as:

$$\int_{\lambda}^{\infty} I\left(\frac{x}{M} \leq t\right) dF_X(x) = \begin{cases} F_X(Mt) - F_X(\lambda), & t \geq \frac{\lambda}{M}, \\ 0, & t < \frac{\lambda}{M}. \end{cases}$$

Note that $T_x \in (x/M, x]$ for $x \leq \lambda$, which will be shown in Proposition 1. It implies that $F(t|x) = 0$ for $t \leq x/M$, and that $F(t|x) = 1$ for $t \geq x$. The first term in the right hand side of equation (2.3) is evaluated as:

$$\int_0^{\lambda} F(t|x) dF_X(x) = \begin{cases} F_X(\lambda), & t \geq \lambda, \\ F_X(t) + \int_t^{M t \wedge \lambda} F(t|x) dF_X(x), & t < \lambda, \end{cases}$$

where $Mt \wedge \lambda$ means $\min\{Mt, \lambda\}$.

The sojourn time distribution is obtained by conditioning on the value of X as follows. From the above, we can see that for $t < \lambda/M$,

$$\Pr\{T \leq t\} = F_X(t) + \int_t^{Mt} F(t|x) dF_X(x),$$

and for $\lambda/M < t \leq \lambda$,

$$\Pr\{T \leq t\} = F_X(t) + \int_t^\lambda F(t|x) dF_X(x) + F_X(Mt) - F_X(\lambda),$$

and for $t \geq \lambda$,

$$\Pr\{T \leq t\} = F_X(Mt).$$

3. The first upcrossing time over the level λ

Let W_i , $i = 1, 2, \dots$, be the amount of work brought by the i^{th} customer after time 0, and U_i , $i = 2, 3, \dots$, the inter-arrival time of the $(i-1)^{\text{th}}$ and i^{th} customers. U_1 denotes the arrival time of the first customer after time 0. Then, $A(\cdot)$ is the cdf of W_i , $i = 1, 2, \dots$, and $B(\cdot)$ is the cdf of U_i , $i = 1, 2, \dots$. We consider the following process, for $0 < x \leq \lambda$,

$$X(t) = x + \sum_{i=1}^{N(t)} W_i - t, \quad t \geq 0, \quad (3.1)$$

where $N(t)$ is the number of customers arrived during $(0, t]$, and $N(0) = 0$. Then, $\{X(t), t \geq 0\}$ is the workload process of the queueing system until it crosses up the level λ or becomes to be negative. Note that the process $\{X(t), t \geq 0\}$ is right-continuous.

The first up-crossing time over the level λ is defined as

$$\tau_x = \inf\{t > 0 : X(t) > \lambda\}, \quad 0 < x \leq \lambda. \quad (3.2)$$

For $x > \lambda$, we let $\tau_x = 0$. The assumption of $B(0) = 0$ says that there is no arrival except the tagged customer at time 0, and that for $0 < x \leq \lambda$, $\tau_x > 0$ almost surely. We denote by $G(t|x)$ the cdf of τ_x . Then, $G(0|x) = 0$ for $0 < x \leq \lambda$. From the following Proposition 1, we can see that it is sufficient to obtain the explicit form of $G(t|x)$, $0 < x \leq \lambda$, in order to obtain the conditional sojourn time distribution $F(t|x)$.

Proposition 1. For $0 < x \leq \lambda$, the support of T_x is $(x/M, x]$. The distribution of T_x is given by

$$\Pr\{T_x = x\} = 1 - G(x|x), \quad (3.3)$$

and

$$F(t|x) = G\left(\frac{M}{M-1} \left(t - \frac{x}{M}\right) \middle| x\right), \quad \frac{x}{M} < t < x. \quad (3.4)$$

Proof: Suppose that $0 < x \leq \lambda$. If $\tau_x \geq x$, then the service rate during $[0, x)$ is 1. The amount of work served during the time interval is x , which means that $T_x = x$. Conversely, the event $\{T_x = x\}$ implies that $\tau_x \geq x$ in the same reason. Thus, we can see that

$$\Pr\{T_x = x\} = \Pr\{\tau_x \geq x\}, \quad (3.5)$$

which gives equation (3.3).

If $0 < \tau_x < x$, then the amount of work served during $[0, \tau_x]$ is τ_x . After time τ_x , $(x - \tau_x)$ amount of work needs to be served before the completion of the service for the tagged customer. Therefore, we have

$$\begin{aligned} T_x &= \tau_x + \frac{x - \tau_x}{M} \\ &= \frac{x}{M} + \frac{M-1}{M} \tau_x. \end{aligned} \quad (3.6)$$

From equation (3.6), we can see that T_x is an increasing function of τ_x and has values in $(x/M, x)$ for $0 < \tau_x < x$. equation (3.4) follows from equation (3.6). Since $\tau_x > 0$ almost surely, we can see that the support of T_x is $(x/M, x]$. \square

From equations (3.5) and (3.6), it follows that for $0 < x \leq \lambda$,

$$T_x = \frac{x}{M} + \frac{M-1}{M} \tau_x \wedge x,$$

which gives

$$E[T_x] = \frac{x}{M} + \frac{M-1}{M} (E[\tau_x | \tau_x < x] \Pr\{\tau_x < x\} + x \Pr\{\tau_x \geq x\}).$$

To obtain another representation of τ_x in equation (3.2) for $0 < x \leq \lambda$, we consider the following random walk $\{S_n, n = 0, 1, 2, \dots\}$ associated with the $G/G/1$ queue:

$$\begin{aligned} S_0 &= x, \\ S_n &= x + \sum_{i=1}^n W_i - U_i, \quad i = 1, 2, \dots \end{aligned}$$

Since we have assumed that the tagged customer arrives at time 0 and the inter-arrival times are U_1, U_2, \dots , the arrival time of the n^{th} customer after time 0 is $t_n = \sum_{k=1}^n U_k$ for $n \geq 1$. From equation (3.1), it follows that

$$S_n = X(t_n), \quad n = 0, 1, 2, \dots,$$

where $t_0 = 0$. Therefore, for an n , $S_n > \lambda$ if and only if $X(t_n) > \lambda$. We define

$$N = \min\{n > 0 : S_n > \lambda\}. \quad (3.7)$$

Then, the process $\{X(t), t \geq 0\}$ crosses up the level λ first at time t_N , i.e.,

$$\tau_x = \sum_{i=1}^N U_i. \quad (3.8)$$

The first up-crossing time over the level λ is related to the initial busy period of a dual queue of the $G/G/1$ queue. In the dual queue, there is $\lambda - x$ amount of work before $t = 0$, and the first customer arrives at time 0. The n^{th} customer arrives at $\tilde{t}_n = \sum_{i=1}^{n-1} W_i$, $n = 2, 3, \dots$, i.e., inter-arrival times of

customers are W_1, W_2, \dots . The amount of work brought by the n^{th} customer is U_n , $n = 1, 2, \dots$. The dual queue is a $G/G/1$ queue with initial workload $\lambda - x$, where the inter-arrival times are iid with cdf $B(\cdot)$ and amounts of work brought by customers are iid with cdf $A(\cdot)$. The initial busy period means the period from time 0 to the first epoch of being empty of the server. Applying the argument used in Cohen (1969) and Prabhu (1997), we show that τ_x can be represented using the length of the initial busy period of the dual $G/G/1$ queue as follows.

Proposition 2. *Let $\tilde{\tau}_y$ be the length of the initial busy period of the dual $G/G/1$ queue when there is y amount of work in the dual queue before $t = 0$. Then, for $0 < x \leq \lambda$,*

$$\tau_x =_d \tilde{\tau}_{\lambda-x} - (\lambda - x), \quad (3.9)$$

where $=_d$ means the same in distribution.

Proof: Suppose that $0 < x \leq \lambda$. We consider the following process:

$$\tilde{X}(t) = \lambda - x + \sum_{i=1}^{\tilde{N}(t)} U_i - t, \quad t \geq 0,$$

where $\tilde{N}(t)$ is the number of customers arrived during $[0, t]$ in the dual queue. From the definition of the dual queue, $\tilde{N}(0) = 1$. We can see that $\tilde{X}(t)$ is the workload of the dual queue at time t during the initial busy period. The process $\{\tilde{X}(t), t \geq 0\}$ is a right continuous process with left limit. Since $\tilde{t}_n = \sum_{i=1}^{n-1} W_i$, $n = 2, 3, \dots$, we have

$$\tilde{X}(\tilde{t}_n-) = \lambda - x + \sum_{i=1}^{n-1} U_i - W_i, \quad n = 2, 3, \dots$$

If $\tilde{X}(\tilde{t}_i-) > 0$ for $i \leq n$, then the initial busy period of the dual queue continues after time \tilde{t}_n . Let \tilde{N} be the number of the customers served during the initial busy period. Then, $\tilde{X}(\tilde{t}_n-) > 0$ for $n \leq \tilde{N}$, and $\tilde{X}(\tilde{t}_n-) < 0$ for $n = \tilde{N} + 1$, which implies

$$\tilde{N} = \min\{n > 0 : \lambda - x + \sum_{i=1}^n U_i - W_i < 0\}. \quad (3.10)$$

By comparing equation (3.10) with equation (3.7), we can see that N and \tilde{N} have the same distribution. Since the length of the initial busy period of the dual $G/G/1$ queue is equal to the amount of work served during the initial busy period, we have

$$\tilde{\tau}_{\lambda-x} = \lambda - x + \sum_{i=1}^{\tilde{N}} U_i. \quad (3.11)$$

Applying above equation to equation (3.8) completes the proof. \square

4. Conditional sojourn time distributions in $M/G/1$ and $G/M/1$ queues

Suppose that the customers arrive according to a Poisson process with rate ν , i.e., $\nu = 1/\alpha$ and $A(x) = 1 - e^{-\nu x}$, $x \geq 0$. Then, the explicit form of $G(t|x)$, $0 < x \leq \lambda$, is given in the following proposition.

Proposition 3. *Suppose that $A(x) = 1 - e^{-\nu x}$, $x \geq 0$, and define*

$$H(u, t) = \sum_{k=0}^{\infty} \frac{(\nu t)^k}{k!} e^{-\nu t} B^{k*}(u),$$

where $B^{n*}(\cdot)$ is the n -fold Stieltjes-convolution of $B(\cdot)$. Then,

$$G(t|\lambda) = 1 - \frac{1}{t} \int_0^t H(u, t) du,$$

and for $0 < x < \lambda$,

$$G(t|x) = 1 - H(\lambda - x + t, t) + \int_0^t \frac{1}{t-u} \left(\int_0^{t-u} H(y, t-u) dy \right) d_u H(\lambda - x + u, u). \quad (4.1)$$

Proof: We can rewrite τ_x in equation (3.2) as: for $0 < x \leq \lambda$,

$$\begin{aligned} \tau_x &= \inf\{t > 0 : \lambda - X(t) < 0\} \\ &= \inf \left\{ t > 0 : \lambda - x + t - \sum_{i=1}^{N(t)} W_i < 0 \right\}. \end{aligned}$$

We denote $\lambda - x + t - \sum_{i=1}^{N(t)} W_i$ by $Y(t)$, $t \geq 0$. Then, $\{Y(t), t \geq 0\}$ is a Cramér-Lundberg surplus process with initial surplus $\lambda - x$ and premium rate 1. Then, τ_x is the time to the ruin of the process $\{Y(t), t \geq 0\}$, and $G(t|x)$ is the cdf of the ruin time. An explicit form of it is given in Asmussen (2000, pp.104–106), which completes the proof. \square

In equation (4.1), $d_u H(\lambda - x + u, u)$ means $H(\lambda - x + du, u)$. Note that $H(u, t) = \Pr\{\sum_{i=1}^{N(t)} W_i \leq u\}$, i.e., for $t > 0$, $H(u, t)$ is the cdf of the total amount of work arrived during $(0, t]$. Clearly, it has a point mass $H(0, t) = e^{-\nu t}$ at 0. When $B(\cdot)$ has a density $b(\cdot)$, the total amount also has the density given by $h(u, t) = \partial H(u, t)/\partial u$, i.e.,

$$h(u, t) = \sum_{k=1}^{\infty} \frac{(\nu t)^k}{k!} e^{-\nu t} b^{(k)}(u),$$

where $b^{(k)}(x)$ is the k -fold convolution of $b(x)$, i.e.,

$$\begin{aligned} b^{(1)}(u) &= b(u), \\ b^{(k)}(u) &= \int_0^u b^{(k-1)}(u-y)b(y) du, \quad k \geq 2. \end{aligned}$$

In this case, equation (4.1) is rewritten as:

$$G(t|x) = 1 - H(\lambda - x + t, t) + \int_0^t \frac{1}{t-u} \left(\int_0^{t-u} H(y, t-u) dy \right) h(\lambda - x + u, u) du.$$

By change of variable $w = t - u$, we have

$$G(t|x) = 1 - H(\lambda - x + t, t) + \int_0^t h(\lambda - x + w, w)e^{-\nu(t-w)} dw + \int_0^t \frac{1}{w} h(t - w + \lambda - x, t - w) \left(\int_0^w yh(w - y, w) dy \right) dw. \tag{4.2}$$

The above form of $G(t|x)$ is also obtained by Stadje and Zacks (2003) in a different manner from Asmussen (2000). Let $g(t|x)$ be the probability density function of the conditional sojourn time T_x . To obtain $g(t|x)$ from equation (4.2) is not easy. Stadje and Zacks (2003) also obtained the explicit form of $g(t|x)$ as:

$$g(t|x) = \nu e^{-\nu t} \bar{B}(\lambda - x + t) + \nu \int_0^{\lambda - x + t} h(u, t) \bar{B}(\lambda - x + t - u) du - \nu \int_0^t h(\lambda - x + u, u) e^{-\nu(t-u)} \bar{B}(t - u) du - \nu \int_0^t \frac{1}{u} \int_0^u (u - y) h(y, u) \bar{B}(u - y) dy du. \tag{4.3}$$

Applying Proposition 3 to Proposition 1, we obtain $F(t|x)$ for $0 < x \leq \lambda$. Since the tagged customer is chosen randomly, the system is in the stationary state at the arrival of the tagged customer due to the property of PASTA. It implies that Y follows the stationary workload distribution. An explicit form of $F_Y(\cdot)$ is obtained by Bae *et al.* (2002).

Now, we consider the case that the amount of the work brought by customers are exponentially distributed, i.e., $B(x) = 1 - e^{-x/\beta}$, $x \geq 0$. In Section 2, we have assumed that the tagged customer arrives at time 0 and the workload of the server just after the arrival of the tagged customer is x . Then, the queueing system under consideration is the $G/M/1$ queue with $X(0) = x$. The dual queue of the $G/M/1$ queue is an $M/G/1$ queue with initial workload $\lambda - x$, which is defined in Section 3. The Laplace transform and the cdf of $\tilde{\tau}_y$, the length of the initial busy period of the dual $M/G/1$ with $y = \lambda - x$, is obtained in equations (4.94) and (4.95) of Cohen (1969, p.261), respectively, as:

$$E[e^{-s\tilde{\tau}_y}] = \exp \left\{ -y \left(s + \frac{1 - \psi(s)}{\beta} \right) \right\}, \tag{4.4}$$

where $\psi(s)$ is the Laplace transform of the length of the initial busy period of the dual $M/G/1$ with $y = 0$, i.e., the length of the ordinary busy period, and

$$\Pr\{\tilde{\tau}_y \leq t\} = \begin{cases} \sum_{k=0}^{\infty} \int_y^t e^{-\frac{s}{\beta}} \frac{s^k}{\beta^k k!} \left(\frac{y}{s}\right) d_s B^{k*}(s - y), & t \geq y, \\ 0, & t < y, \end{cases} \tag{4.5}$$

where $B^{0*}(\cdot)$ is the Heaviside step function. Proposition 2 says that

$$E[e^{-s\tau_x}] = e^{s(\lambda-x)} E[e^{-s\tau_{\lambda-x}}],$$

and that $G(t|x) = \Pr\{\tilde{\tau}_{\lambda-x} \leq t + \lambda - x\}$. Then, from equations (4.4) and (4.5), we have

$$E[e^{-s\tau_x}] = \exp \left\{ -(\lambda - x) \frac{1 - \psi(s)}{\beta} \right\}, \tag{4.6}$$

and

$$G(t|x) = \sum_{k=0}^{\infty} \int_0^t e^{-\frac{s+\lambda-x}{\beta}} \frac{(s+\lambda-x)^k}{\beta^k k!} \left(\frac{\lambda-x}{s+\lambda-x} \right) dB^{k*}(s), \quad t > 0. \quad (4.7)$$

Applying equation (4.7) to Proposition 1, we obtain the conditional sojourn time distribution of a customer in the $G/M/1$ queue with P_λ^M -service policy.

References

- Abdel-Hameed M (2000). Optimal control of a dam using $P_{\lambda,\tau}^M$ policies and penalty cost when the input process is a compound Poisson process with positive drift, *Journal of Applied Probability*, **37**, 408–416.
- Abdel-Hameed M (2011). Control of dams using $P_{\lambda,\tau}^M$ policies when the input process is a nonnegative Lévy process, *International Journal of Stochastic Analysis*, **2011**, Article ID 916952.
- Abdel-Hameed MS and Nakhi YA (2006). Optimal control of a finite dam with diffusion input and state dependent release rates, *Computers & Mathematics with Applications*, **51**, 317–324.
- Asmussen S (2000). *Ruin Probabilities*, World Scientific, Singapore.
- Bae J, Kim S, and Lee EY (2002). A P_λ^M -policy for an $M/G/1$ queueing system, *Applied Mathematical Modelling*, **26**, 929–939.
- Balachandran, KR (1973). Control policies for a single server system, *Management Science*, **19**, 1013–1018.
- Bekker R (2009). Queues with Lévy input and hysteretic control, *Queueing Systems*, **63**, 281–299.
- Cohen JW (1969). *The Single Server Queue*, North-Holland Amsterdam, Amsterdam.
- Faddy MJ (1974). Optimal control of finite dams: discrete (2-stage) output procedure, *Journal of Applied Probability*, **11**, 111–121.
- Hur S, Lee H, and Kim S (2006). A simple approximation method for workload analyses in some queueing systems with control policies, *Computers & Industrial Engineering*, **51**, 183–195.
- Kim J, Bae J, and Lee EY (2006). An optimal P_λ^M -service policy for an $M/G/1$ queueing system, *Applied Mathematical Modelling*, **30**, 38–48.
- Kim S and Bae J (2008). A $G/M/1$ queueing system with P_λ^M -service policy, *Operations Research Letters*, **36**, 201–204.
- Lee J and Kim J (2006). A workload-dependent $M/G/1$ queue under a two-stage service policy, *Operations Research Letters*, **34**, 531–538.
- Lee J and Kim J (2007). Workload analysis of an $M/G/1$ queue under the P_λ^M policy with a set-up time, *Applied Mathematical Modelling*, **31**, 236–244.
- Prabhu NU (1997). *Stochastic Storage Processes: Queues, Insurance Risk, Dams, and Data Communication* (2nd ed), Springer, New York.
- Stadje W and Zacks S (2003). Upper first-exit times of compound Poisson processes revisited, *Probability in the Engineering and Informational Sciences*, **17**, 459–465.
- Yeh L (1985). Optimal control of a finite dam: average-cost case, *Journal of Applied Probability*, **22**, 480–484.