Journal of the Korean Statistical Society (2005), **34**: *1, pp 39–47*

POISSON ARRIVAL QUEUE WITH ALTERNATING SERVICE RATES †

JONGWOO KIM[1], EUI YONG LEE[2] AND HO WOO LEE[3]

ABSTRACT

We adopt the $P_{\lambda,\tau}^M$ policy of dam to introduce a service policy with alternating service rates for a Poisson arrival queue, in which the service rate alternates depending on the number of customers in the system. The stationary distribution of the number of customers in the system is derived and, after operating costs being assigned to the system, the optimization of the policy is studied.

AMS 2000 subject classifications. Primary 60K25; Secondary 90B22.
Keywords. Poisson arrival queue, alternating service rates, long-run average cost.

1. INTRODUCTION

After Faddy (1974) introduced a P_λ^M policy for a finite dam, Yeh (1985) generalized the policy to introduce a $P_{\lambda,\tau}^M$ policy; the release rate of water is changed instantaneously from 0 to M (from M to 0) when the level of water in the reservoir upcrosses λ (downcrosses τ), where $M > 0$ and $\lambda > \tau \geq 0$. See Yeh (1985), Abdel-Hameed (2000), and Bae *et al.* (2003) for details.

We, in this paper, apply the $P_{\lambda,\tau}^M$ policy to introduce a service policy with alternating service rates for a Poisson arrival queue. The system is initially empty and customers arrive according to a Poisson process of rate $\nu > 0$. The server starts to work on an arrival of customer at a service rate $\mu_1 \geq 0$, that is, the customers are served for an exponential amount of time with mean $1/\mu_1$.

Received July 2004; accepted December 2004.
†This research was supported by KOSEF through Statistical Research Center for Complex Systems at Seoul National University.
[1]Statistical Research Center for Complex Systems, Seoul National University, Seoul 151-742, Korea (e-mail : jwkim@stats.snu.ac.kr)
[2]Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea
[3]Department of Systems Management Engineering, Sungkyunkwan University, Suwon 440-746, Korea

However, if the number of customers in the system reaches λ (integer-valued), the service rate is increased to $\mu_2 \geq \mu_1$, so that the customers, including the one being served at the moment, are now served for an exponential amount of time with mean $1/\mu_2$. The fast service rate is continued until the number of customers becomes τ (integer-valued, $0 \leq \tau < \lambda$). At this point the service rate is changed to the ordinary service rate μ_1. The service rate gets μ_2 again if the number of customers in the system reaches λ before the busy period ends, otherwise, the server finishes the present busy period with ordinary service rate μ_1. We assume that $\mu_2 > \nu$ for the stability of the system.

A similar model for finite message storage buffer was studied by Li (1989). The stationary distribution of the queue length was obtained through the generator matrices. In present paper, an explicit formula is obtained for the stationary distribution of the number of customers in the system by using the decomposition technique introduced by Lee and Ahn (1998). After assigning operating costs to the system, we calculate the long-run average cost per unit time and show that there exists a unique fast service rate which minimizes the long-run average cost per unit time.

2. Stationary Distribution of the Number of Customers

Let $\{N(t), t \geq 0\}$ be the process of the number of customers in the system. Note that the time 0 and the epochs where busy periods end form embedded regeneration points of $\{N(t), t \geq 0\}$. Denote by T the length of a cycle, the interval between two successive embedded regeneration points, and by $T_1(T_2)$ the total length of periods with service rate $\mu_1(\mu_2)$ in a cycle. That is,

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

where $1/\nu$ is the expected length of idle period. Note that $\{N(t), t \geq 0\}$ is non-Markovian, hence, we decompose $\{N(t), t \geq 0\}$ into three Markov processes $\{N_1(t), t \geq 0\}$, $\{N_2(t), t \geq 0\}$, and $\{N_3(t), t \geq 0\}$. Process $\{N_1(t), t \geq 0\}$ is formed by separating periods of service rate μ_1 from the original process and connecting them together. Process $\{N_2(t), t \geq 0\}$ is similarly formed by separating and connecting the periods of service rate μ_2. Process $\{N_3(t), t \geq 0\}$ is formed by connecting the rest of the original process, that is, $N_3(t) \equiv 0$ for all $t \geq 0$.

Let $P(n)$ be the stationary distribution of $\{N(t), t \geq 0\}$ and $P_i(n)$ be the stationary distribution of $\{N_i(t), t \geq 0\}$, for $i = 1, 2$. Since $E[T_i]/E[T]$ is the long-run proportion of time that $\{N_i(t), t \geq 0\}$ takes in the original process $\{N(t), t \geq 0\}$, for $i = 1, 2$,

$$P(n) = \frac{E[T_1]}{E[T]}P_1(n) + \frac{E[T_2]}{E[T]}P_2(n) + \frac{1/\nu}{E[T]}I_{\{n=0\}},$$

for $n = 0, 1, 2, \ldots$, where I_A is the indicator of event A.

Define T_1^1 as the length of the first passage time from the starting point of busy period to the epoch where the number of customers reaches either 0 or λ, while p_1 as the probability of reaching λ, T_2^λ as that from the epoch where the service rate is changed to μ_2 to the epoch where the number of customers reaches τ, and T_1^τ as that from the epoch where the service rate is changed to μ_1 from μ_2 to epoch where the number of customers reaches either 0 or λ, while p_2 as the probability of reaching λ. Then, due to the strong Markovian property, we have

$$E[T_1] = E[T_1^1] + \frac{p_1}{1-p_2}E[T_1^\tau]$$

and

$$E[T_2] = \frac{p_1}{1-p_2}E[T_2^\lambda].$$

It is well-known in the theory of Markov process that

$$p_1 = \begin{cases} \dfrac{1-(\mu_1/\nu)}{1-(\mu_1/\nu)^\lambda}, & \text{if } \mu_1 \neq \nu, \\ \dfrac{1}{\lambda}, & \text{if } \mu_1 = \nu, \end{cases}$$

and

$$p_2 = \begin{cases} \dfrac{1-(\mu_1/\nu)^\tau}{1-(\mu_1/\nu)^\lambda}, & \text{if } \mu_1 \neq \nu, \\ \dfrac{\tau}{\lambda}, & \text{if } \mu_1 = \nu. \end{cases}$$

To derive $E[T_1^1]$, we need the following two lemmas:

LEMMA 2.1. *Consider a process $M(t) = N(t) - (\nu - \mu_1)t$, after a busy period begins in $\{N(t), t \geq 0\}$. Then, $M(t)$ is a martingale with $E[M(t)] = 1$ until $N(t)$ reaches either state 0 or λ.*

PROOF. Denote by $\{\mathcal{F}_t\}$ the filtration of $\sigma\{N(s), 0 \le s \le t\}$. For $s \le t$,

$$\begin{aligned}
E[M(t) \mid \mathcal{F}_s] &= E[N(t) - (\nu - \mu_1)t \mid \mathcal{F}_s] \\
&= E[N(t) - N(s) + N(s) - (\nu - \mu_1)t \mid \mathcal{F}_s] \\
&= E[N(t) - N(s) \mid \mathcal{F}_s] + E[N(s) - (\nu - \mu_1)t \mid \mathcal{F}_s] \\
&= (\nu - \mu_1)(t - s) + N(s) - (\nu - \mu_1)t \\
&= N(s) - (\nu - \mu_1)s \\
&= M(s).
\end{aligned}$$

□

LEMMA 2.2. $E[\sup_{t \ge 0} |M(T_1^1 \wedge t)|] < \infty$, *where* $T_1^1 \wedge t = \min(T_1^1, t)$.

PROOF.

$$\begin{aligned}
E\left[\sup_{t \ge 0} |M(T_1^1 \wedge t)|\right] &= E\left[\sup_{0 \le t \le T_1^1} |M(t)|\right] \\
&= E\left[\sup_{0 \le t \le T_1^1} |N(t) - (\nu - \mu_1)t|\right] \\
&\le E\left[\sup_{0 \le t \le T_1^1} (N(t) + |\nu - \mu_1|t)\right] \\
&\le E\left[\sup_{0 \le t \le T_1^1} N(t)\right] + |\nu - \mu_1| E\left[T_1^1\right] \\
&< \lambda + |\nu - \mu_1| E[T_1^1] \\
&< \infty,
\end{aligned}$$

since $E[T_1^1] < \infty$. □

Since T_1^1 is a Markov time and $\Pr\{T_1^1 < \infty\} = 1$, applying the optional sampling theorem(Karlin and Taylor (1975, p.259)) to $M(t)$ gives $E[M(T_1^1)] = 1$, which yields

$$E[T_1^1] = \begin{cases} \dfrac{\lambda/\nu}{1 - (\mu_1/\nu)^\lambda} - \dfrac{1/\nu}{1 - (\mu_1/\nu)}, & \text{if } \mu_1 \ne \nu, \\ \dfrac{\lambda - 1}{2\nu}, & \text{if } \mu_1 = \nu. \end{cases}$$

A similar argument to the above yields

$$E[T_1^\tau] = \begin{cases} \dfrac{\lambda\left\{1-(\mu_1/\nu)^\tau\right\}/\nu}{\left\{1-(\mu_1/\nu)\right\}\left\{1-(\mu_1/\nu)^\lambda\right\}} - \dfrac{\tau/\nu}{1-(\mu_1/\nu)}, & \text{if } \mu_1 \neq \nu, \\ \dfrac{\tau(\lambda-\tau)}{2\nu}, & \text{if } \mu_1 = \nu. \end{cases}$$

$E[T_2^\lambda]$ can be easily derived from the queueing theory as follows:

$$E[T_2^\lambda] = \frac{\lambda-\tau}{\mu_2-\nu}.$$

To evaluate $P(n)$, it remains to obtain $P_1(n)$ and $P_2(n)$, which can be done through establishing balance equations. The balance equations for $P_1(n)$ are

$$\begin{aligned}
&(\nu+\mu_1)P_1(1) = \mu_1 P_1(1) + \mu_1 P_1(2), \\
&(\nu+\mu_1)P_1(n) = \nu P_1(n-1) + \mu_1 P_1(n+1), \quad \text{for } 2 \leq n \leq \lambda-2,\ n \neq \tau, \\
&(\nu+\mu_1)P_1(\tau) = \nu P_1(\tau-1) + \mu_1 P_1(\tau+1) + \nu P_1(\lambda-1), \\
&(\nu+\mu_1)P_1(\lambda-1) = \nu P_1(\lambda-2),
\end{aligned}$$

which, with $\sum_{n=1}^{\lambda-1} P_1(n) = 1$, have a unique solution given by, for $n = 1, 2, \ldots, \tau$,

$$P_1(n) = \begin{cases} \dfrac{(\mu_1/\nu)^{\tau-n}\left\{1-(\mu_1/\nu)\right\}\left\{1-(\mu_1/\nu)^{\lambda-\tau}\right\}}{(\lambda-\tau)\left\{1-(\mu_1/\nu)\right\} - (\mu_1/\nu)^\tau + (\mu_1/\nu)^\lambda}, & \text{if } \mu_1 \neq \nu, \\ \dfrac{2(\lambda-\tau)}{\lambda(\lambda-1)-\tau(\tau-1)}, & \text{if } \mu_1 = \nu, \end{cases}$$

for $n = \tau+1, \tau+2, \ldots, \lambda-1$,

$$P_1(n) = \begin{cases} \dfrac{\left\{1-(\mu_1/\nu)\right\}\left\{1-(\mu_1/\nu)^{\lambda-n}\right\}}{(\lambda-\tau)\left\{1-(\mu_1/\nu)\right\} - (\mu_1/\nu)^\tau + (\mu_1/\nu)^\lambda}, & \text{if } \mu_1 \neq \nu, \\ \dfrac{2(\lambda-n)}{\lambda(\lambda-1)-\tau(\tau-1)}, & \text{if } \mu_1 = \nu, \end{cases}$$

and the balance equations for $P_2(n)$ are

$$\begin{aligned}
&(\nu+\mu_2)P_2(\tau+1) = \mu_2 P_2(\tau+2), \\
&(\nu+\mu_2)P_2(n) = \nu P_2(n-1) + \mu_2 P_2(n+1), \quad \text{for } n \geq \tau+2,\ n \neq \lambda, \\
&(\nu+\mu_2)P_2(\lambda) = \mu_2 P_2(\tau+1) + \nu P_2(\lambda-1) + \mu_2 P_2(\lambda+1),
\end{aligned}$$

which, with $\sum_{n=\tau+1}^{\infty} P_2(n) = 1$, have a unique solution given by

$$P_2(n) = \begin{cases} \dfrac{1-(\nu/\mu_2)^{n-\tau}}{\lambda-\tau}, & \text{for } n = \tau+1,\dots,\lambda-1, \\ \dfrac{(\nu/\mu_2)^{n-\lambda}\left\{1-(\nu/\mu_2)^{\lambda-\tau}\right\}}{\lambda-\tau}, & \text{for } n = \lambda, \lambda+1, \dots. \end{cases}$$

3. OPTIMAL FAST SERVICE RATE

In this section, after assigning costs to the system, we calculate the long-run average cost per unit time and show that there exists a unique fast service rate minimizing the long-run average cost per unit time. We consider the following four costs:

- $c_1(\mu)$: the operating cost per unit time while the server is working with service rate μ, where $c_1(0) = 0$.
- $c_2(\mu_2)$: the cost needed to increasing the service rate to μ_2, where $c_2(\mu_1) = 0$.
- $c_3(n)$: the penalty per unit time for the heavy traffic when $N(t) = \lambda + n$, for $n = 1, 2, \dots$.
- c_4: the penalty per unit time while the server is idle.

We assume that c_1, c_2, and c_3 are all nonnegative strictly increasing functions, and that c_1 and c_2 are twice differentiable convex functions.

First of all, from the results in section 2, the expected length of a cycle can be expressed as a function of μ_2,

$$E[T] = E[T_1] + \frac{p_1}{1-p_2}\frac{\lambda-\tau}{\mu_2-\nu} + \frac{1}{\nu}.$$

The expected costs during a cycle corresponding to the above four costs are obtained as follows:

- $E[\text{operating cost}] = E[T_1]c_1(\mu_1) + \frac{p_1}{1-p_2}\frac{\lambda-\tau}{\mu_2-\nu}c_1(\mu_2)$.
- $E[\text{switching cost}] = \frac{p_1}{1-p_2}c_2(\mu_2)$.
- $E[\text{penalty for heavy traffic}] = \frac{p_1}{1-p_2}\frac{1-(\nu/\mu_2)^{\lambda-\tau}}{\mu_2-\nu}\sum_{n=1}^{\infty} c_3(n)\left(\frac{\nu}{\mu_2}\right)^n$.

- E[penalty for idle period] $= \frac{c_4}{\nu}$.

The foregoing enable us to establish the following long-run average cost per unit time:

$$C(\mu_2) = \frac{p_1}{1-p_2} \frac{(\mu_2 - \nu)(A + c_2(\mu_2)) + (\lambda - \tau)c_1(\mu_2) + a(\mu_2)}{(E[T_1] + 1/\nu)(\mu_2 - \nu) + p_1(\lambda - \tau)/(1-p_2)}, \tag{3.1}$$

where $A = (1-p_2)(E[T_1]c_1(\mu_1) + c_4/\nu)/p_1$ and

$$a(\mu_2) = \left\{1 - \left(\frac{\nu}{\mu_2}\right)^{\lambda-\tau}\right\} \sum_{n=1}^{\infty} c_3(n) \left(\frac{\nu}{\mu_2}\right)^n.$$

An algebra shows that $a(\mu_2)$ can be written as

$$a(\mu_2) = \sum_{n=1}^{\lambda-\tau} c_3(n) \left(\frac{\nu}{\mu_2}\right)^n + \sum_{n=\lambda-\tau+1}^{\infty} (c_3(n) - c_3(n-\lambda+\tau)) \left(\frac{\nu}{\mu_2}\right)^n. \tag{3.2}$$

From equation (3.1), we have

$$C'(\mu_2) = \frac{p_1}{1-p_2} \frac{b_1(\mu_2) - b_2(\mu_2)}{\{(E[T_1] + 1/\nu)(\mu_2 - \nu) + p_1(\lambda-\tau)/(1-p_2)\}^2},$$

where

$$\begin{aligned} b_1(\mu_2) = &\left\{\left(E[T_1] + \frac{1}{\nu}\right)(\mu_2 - \nu) + \frac{p_1}{1-p_2}(\lambda - \tau)\right\}(\mu_2 - \nu)c_2'(\mu_2) \\ &+ \left\{\left(E[T_1] + \frac{1}{\nu}\right)(\mu_2 - \nu) + \frac{p_1}{1-p_2}(\lambda - \tau)\right\}(\lambda - \tau)c_1'(\mu_2) \\ &- \left(E[T_1] + \frac{1}{\nu}\right)(\lambda - \tau)c_1(\mu_2) + \frac{p_1}{1-p_2}(\lambda - \tau)(A + c_2(\mu_2)) \end{aligned}$$

and

$$\begin{aligned} b_2(\mu_2) = &\left(E[T_1] + \frac{1}{\nu}\right) a(\mu_2) \\ &- \left\{\left(E[T_1] + \frac{1}{\nu}\right)(\mu_2 - \nu) + \frac{p_1}{1-p_2}(\lambda - \tau)\right\} a'(\mu_2). \end{aligned}$$

Notice that $b_1(\mu_2)$ is strictly increasing, since c_1 and c_2 are strictly increasing and twice differentiable convex functions. By differentiating equation (3.2) twice with respect to μ_2, we observe that $a''(\mu_2) \geq 0$ and thus $b_2(\mu_2)$ is decreasing. It is obvious $\lim_{\mu_2 \to \infty} b_2(\mu_2) = 0$. In order to see $\lim_{\mu_2 \to \infty} b_1(\mu_2) > 0$, we need the following lemma:

LEMMA 3.1.
$$\liminf_{\mu_2\to\infty} \frac{(\mu_2-\nu)c_1'(\mu_2)}{c_1(\mu_2)} \geq 1.$$

PROOF.
$$\begin{aligned}
\liminf_{\mu_2\to\infty} \frac{(\mu_2-\nu)c_1'(\mu_2)}{c_1(\mu_2)} &= \liminf_{\mu_2\to\infty} \frac{(\mu_2-\nu)c_1'(\mu_2)}{c_1(\mu_2)-c_1(\nu)} \frac{c_1(\mu_2)-c_1(\nu)}{c_1(\mu_2)} \\
&\geq \left(\liminf_{\mu_2\to\infty} \frac{(\mu_2-\nu)c_1'(\mu_2)}{c_1(\mu_2)-c_1(\nu)}\right)\left(\liminf_{\mu_2\to\infty} \frac{c_1(\mu_2)-c_1(\nu)}{c_1(\mu_2)}\right) \\
&= \liminf_{\mu_2\to\infty} \frac{(\mu_2-\nu)c_1'(\mu_2)}{(c_1(\mu_2)-c_1(\nu))},
\end{aligned}$$

since $\liminf_{\mu_2\to\infty}(c_1(\mu_2)-c_1(\nu))/c_1(\mu_2) = 1$. By the mean value theorem, there exists $\tilde{\mu_2}$ ($\nu < \tilde{\mu_2} < \mu_2$), which depends on μ_2, such that

$$c_1(\mu_2) - c_1(\nu) = (\mu_2-\nu)c_1'(\tilde{\mu_2}).$$

Hence,

$$\begin{aligned}
\liminf_{\mu_2\to\infty} \frac{(\mu_2-\nu)c_1'(\mu_2)}{c_1(\mu_2)} &\geq \liminf_{\mu_2\to\infty} \frac{c_1'(\mu_2)}{c_1'(\tilde{\mu_2})} \\
&\geq 1.
\end{aligned}$$

□

By lemma 3.1, we can see that $b_1(\mu_2)$ and $b_2(\mu_2)$ cross each other at most once. Therefore, we have the following theorem:

THEOREM 3.1. *There exists a unique fast service rate μ_2^* which minimize the long-run average cost per unit time $C(\mu_2)$.*

REMARK 3.1. When $\mu_2 \geq \mu_1 > \nu$,

(i) if $b_1(\mu_1) \geq b_2(\mu_1)$, $C(\mu_2)$ is an increasing function for $\mu_2 \geq \mu_1$ and hence minimized at $\mu_2 = \mu_1$.

(ii) if $b_1(\mu_1) < b_2(\mu_1)$, there exists $\mu_2^* > \mu_1$ such that $C'(\mu_2) < 0$ for $\mu_2 < \mu_2^*$ and $C'(\mu_2) > 0$ for $\mu_2 > \mu_2^*$. Therefore, $C(\mu_2)$ is minimized at $\mu_2 = \mu_2^*$, which is the unique solution of equation $b_1(\mu_2) = b_2(\mu_2)$.

ACKNOWLEDGEMENTS

We would like to thank the referees for their helpful comments.

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.

BAE, J., KIM, S. AND LEE, E. Y. (2003). "Average cost under the $P_{\lambda,\tau}^{M}$ policy in a finite dam with compound Poisson inputs", *Journal of Applied Probability*, **40**, 519–526.

FADDY, M. J. (1974). "Optimal control of finite dams: Discrete (2-stage) output procedure", *Journal of Applied Probability*, **11**, 111–121.

KARLIN, S. AND TAYLOR, H. M. (1975). *A First Course in Stochastic Processes*, 2nd edition, Academic Press, New York.

LEE, E. Y. AND AHN, S. K. (1998). "P_{λ}^{M}-policy for a dam with input formed by a compound Poisson process", *Journal of Applied Probability*, **35**, 482–488.

LI, S.-Q. (1989). "Overload control in a finite message storage buffer", *IEEE Transactions on Communications*, **37**, 1330–1338.

YEH, L. (1985). "Optimal control of a finite dam: Average-cost case", *Journal of Applied Probability*, **22**, 480–484.