International Journal of Reliability and Applications
Vol. 3, No. 4, pp. 165-171, 2002

Optimization of the M/M/1 Queue with Impatient Customers

Kyung Eun Lim and **Eui Yong Lee** *
Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea

Abstract. An optimization of the M/M/1 queue with impatient customers is studied. The impatient customer does not enter the system if his or her virtual waiting time exceeds the threshold $K > 0$. After assigning three costs to the system, a cost proportional to the virtual waiting time, a penalty to each impatient customer, and also a penalty to each unit of the idle period of the server, we show that there exists a threshold K which minimizes the long-run average cost per unit time.

Key Words : *optimization, M/M/1 queue, impatient customer, virtual waiting time*

1. INTRODUCTION

We, in this paper, consider the M/M/1 queue with impatient customers. The customers arrive according to a Poisson process with rate $\lambda > 0$ and the service times of customers are independent and exponentially distributed random variables with mean μ. The impatient customer is the one who leaves the system without being served if he or she has waited for more than a period $K > 0$. In the analysis of virtual waiting time, however, the impatient customer can be considered as the customer who enters the system only when his or her waiting time does not exceed K. A sample path of the virtual waiting time of the M/M/1 queue with impatient customers is shown in Figure 1.

The queue with impatient customers has been studied by many authors. Daley(1965) and De Kok and Tijms(1985) obtained the limiting distributions of the virtual waiting time for M/G/1 and G/G/1 queues. However, their results were restricted to some special cases. Recently, Lee and Lim(2000) and Bae, Kim and

*Corresponding author.
E-mail address: eylee@sookmyung.ac.kr

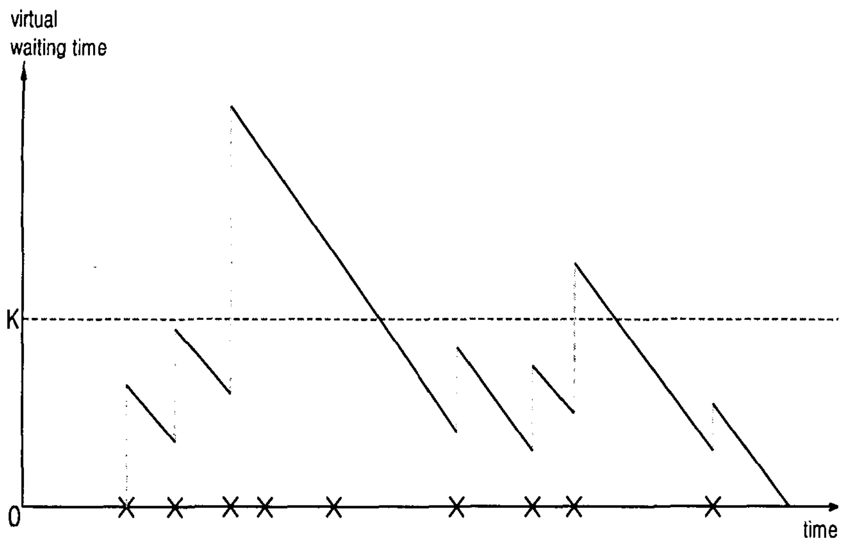

Figure 1. A sample path of the virtual waiting time

Lee(2001) have obtained the explicit and complete formulas for the busy period of the server and the limiting distribution of the virtual waiting time.

In this paper, we assign three costs to the system, a cost proportional to the expected virtual waiting time, a penalty to each impatient customer, and another penalty to each unit of the idle period of the server. We, then, find the value of K which minimizes the long-run average cost. To do this, in Section 2, we obtain the explicit formulas for the expected busy period, the number of impatient customers during a busy period, and the expected virtual waiting time. In Section 3, after computing the long-run average cost per unit time, we show that there exists an optimal value of K.

Our result can be directly applicable when the queue is a transmitting station in a network and the customers are now the sets of data. The transmitting station collects the sets of data from other stations and sends them continuously to the main station. In this case, the service time of the customer is now the length of the data and the virtual waiting time is the pile of data that not been sent. If the current amount of data not being sent exceeds the threshold K, the incoming sets of data are no longer admitted until the amount of data presented in the station becomes less than K again. In this transmitting station, it is important for the manager to find the optimal value of K.

2. INTERESTING CHARACTERISTICS

In this section, we calculate some characteristics of the system, the expected busy period, the number of impatient customers during a busy period, and the expected virtual waiting time, which are needed later to find the long-run average cost per unit and of themselves interesting.

2.1 Busy Period

Lee and Lim (2000) have obtained the explicit formula of the expected busy period starting from level $x(0 < x \leq K)$ by using the optional stopping theorem of matingale as follows :

$$E(B_x) = \frac{Z(x,s,\mu,\lambda,K)}{(\lambda\mu - 1)(e^{-sK} - s\mu - 1)s\mu e^{-sK}},$$

where

$$\begin{aligned} Z(x,s,\mu,\lambda,K) = {} & \lambda\mu^2[(s\mu+1)s\mu e^{-sx} + s\mu(e^{-sK} - s\mu - 2) - 1]] \\ & -(e^{-sK} - s\mu - 1)[xs\mu e^{-sK} - K(e^{-sx} - 1)(s\mu + 1)] \end{aligned}$$

with $s = \frac{\lambda\mu - 1}{\mu}$. Here, x is the service time of the first customer after an idle period.

Let B be the busy period which are not conditioned on x, then, by the double expectation formula,

$$\begin{aligned} E(B) &= E[E(B|X)] \\ &= \int_0^K E(B_x)\frac{1}{\mu}e^{-\frac{x}{\mu}}dx + \int_K^\infty [\mu + E(B_K)]\frac{1}{\mu}e^{-\frac{x}{\mu}}dx \\ &= \frac{\mu[1 - \lambda\mu e^{sK}]}{1 - \lambda\mu}. \end{aligned} \tag{2.1}$$

Observe that the amount of service time exceeding K, when $x > K$, is still an exponential random variable with mean μ due to the memoryless property of the exponential random variable.

2.2 Number of Impatient Customers

Let N be the number of impatient customers during a busy period. Lee and Lim(2000) obtained the expectation of the same number when the busy period started at $x(0 < x \leq K)$, which is given by

$$E(N_x) = \frac{P_K^x P_0^K}{(1 - P_K^K)^2},$$

where $P_K^x = \dfrac{(e^{-sx} - 1)(s\mu + 1)}{e^{-sK} - s\mu - 1}$ with $P_0^x = 1 - P_K^x$.

Conditioning on the service time of the first customer gives

$$\begin{aligned}
E(N) &= E[E(N|X)] \\
&= \int_0^K E(N_x)\frac{1}{\mu}e^{-\frac{x}{\mu}}dx + \int_K^\infty [(x-K)\lambda + E(N_K)]\frac{1}{\mu}e^{-\frac{x}{\mu}}dx \\
&= \frac{\lambda(e^{-\lambda K} - \lambda\mu e^{-\frac{K}{\mu}}) + s\lambda\mu}{se^{-sK}} + \frac{s\lambda\mu e^{-\frac{K}{\mu}} - \lambda e^{-\lambda K} + \lambda e^{-\frac{K}{\mu}}}{se^{-sK}} \\
&= \lambda\mu e^{sK}. \qquad (2.2)
\end{aligned}$$

Notice that $E(N) = \lambda\mu$ when $K = 0$, which is the case when at most one customer is allowed in the system.

2.3 Virtual Waiting Time

Let V be the virtual waiting time. The limiting distribution of the virtual waiting time process was found by Bae, Kim and Lee(2001), which was given by

$$F(x) = \frac{1 - \lambda\mu e^{sx}}{1 - (\lambda\mu)^2 e^{sK}}, \qquad 0 < x \le K$$

$$F(K+y) = 1 - \frac{\lambda\mu(1-\lambda\mu)e^{sK}}{1-(\lambda\mu)^2 e^{sK}}e^{-\frac{y}{\mu}}, \qquad y > 0.$$

From this, we can calculate the long-run average virtual waiting time, $E(V)$ say, as follows :

$$\begin{aligned}
E(V) &= \int_0^\infty \bar{F}(x)dx \\
&= \int_0^K \bar{F}(x)dx + \int_0^\infty \bar{F}(K+y)dy \\
&= \frac{\lambda\mu(e^{sK}-1) - sK(\lambda\mu)^2 e^{sK}}{s(1-(\lambda\mu)^2 e^{sK})} + \frac{\lambda\mu^2(1-\lambda\mu)e^{sK}}{1-(\lambda\mu)^2 e^{sK}} \\
&= \frac{(\lambda\mu)^2 e^{sK}[2-\lambda\mu] - \lambda\mu - sK(\lambda\mu)^2 e^{sK}}{s[1-(\lambda\mu)^2 e^{sK}]}. \qquad (2.3)
\end{aligned}$$

Remarks (i) When $K = \infty$, that is, when there are no impatient customers,

$$E(V) = \frac{\lambda\mu^2}{1-\lambda\mu}$$

which is a well known result in the ordinary M/M/1 queue.

(ii) In the case that $K = 0$,

$$E(V) = \frac{\lambda\mu^2}{1+\lambda\mu}.$$

4. OPTIMIZATION

We, in this section, find the optimal value of K after calculating the long-run average cost per unit time. Let C_1 be the holding cost per unit time of a unit amount of service time. Note that in the long-run, the average holding cost per unit time will be $C_1E(V)$. Let C_2 be the penalty to each impatient customer whom we lost without serving. Let C_3 be the penalty for each unit of the idle period of the server. Define $C(K)$ as the long-run average cost per unit time. Then, by the renewal reward theorem [Ross(1996), p.133], $C(K)$ is given by

$$\begin{aligned} C(K) &= \frac{E[\text{total cost during a cycle}]}{E[\text{length of a cycle}]} \\ &= E(V)C_1 + \frac{E(N)C_2 + \frac{1}{\lambda}C_3}{E(B)+\frac{1}{\lambda}} \\ &= \frac{[(\lambda\mu)^2 e^{sK}[2-\lambda\mu] - \lambda\mu - sK(\lambda\mu)^2 e^{sK}]C_1}{s[1-(\lambda\mu)^2 e^{sK}]} \\ &\quad + \frac{[\lambda^2\mu e^{sK} C_2 + C_3](1-\lambda\mu)}{1-(\lambda\mu)^2 e^{sK}}. \end{aligned} \tag{2.4}$$

Differentiating the above equation with respect to K gives

$$C'(K) = \frac{A(K)}{s^2[1-(\lambda\mu)^2 e^{sK}]^2}$$

where

$$\begin{aligned} A(K) &= [s^2(\lambda\mu)^2 e^{sK} - 2s^2(\lambda\mu)^3 e^{sK} + s^2(\lambda\mu)^4 e^{2sK} - s^3K(\lambda\mu)^2 e^{sK}]C_1 \\ &\quad - s^4(\lambda\mu)^2 e^{sK} C_2 - s^4\lambda^2\mu^3 e^{sK} C_3. \end{aligned} \tag{2.5}$$

After some tedious algebras, we can show that

i) When $\mu^2 C_1 \geq C_2 + \mu C_3$,
$A(K) \geq 0$, for all $K \geq 0$ (see appendix), and hence, $C(K)$ is minimized at $K = 0$.

ii) When $\mu^2 C_1 < C_2 + \mu C_3$,
$A(0) < 0, \lim_{K\to\infty} A(K) = 0+$, and hence there exists a value $K^* > 0$ which minimizes $C(K)$. K^* is the solution of equation $A(K) = 0$ which can be found by a routine numerical calculation.

Remark Notice that C_1 is the cost related to serving customers and C_2 and C_3 may be considered as the penalties of not serving customers. Hence, the condition $\mu^2 C_1 \geq C_2 + \mu C_3$ implicates that the serving cost of customers is relatively larger than the penalty of losing customers, in which case it is optimal to serve as small number of customers as possible. Otherwise, that is, when $\mu^2 C_1 < C_2 + \mu C_3$, there exists an optimal value $K^*(0 < K^* < \infty)$ minimizing the long-run average cost.

The following two figures show the graphs of $A(K)$ in both cases. In Figure 3, we can see that K^* is about 7.4.

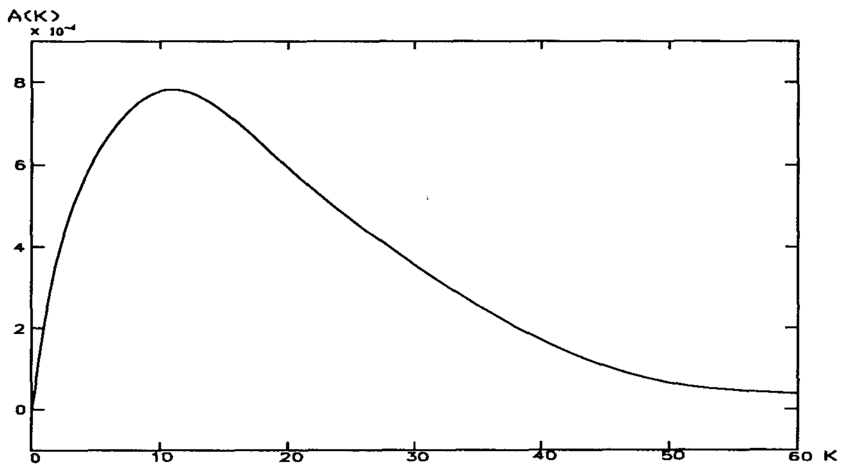

Figure 2. $\mu^2 C_1 \geq C_2 + \mu C_3, \quad (\lambda = 0.1, \mu = 5, C_1 = 1, C_2 = 10, C_3 = 3)$

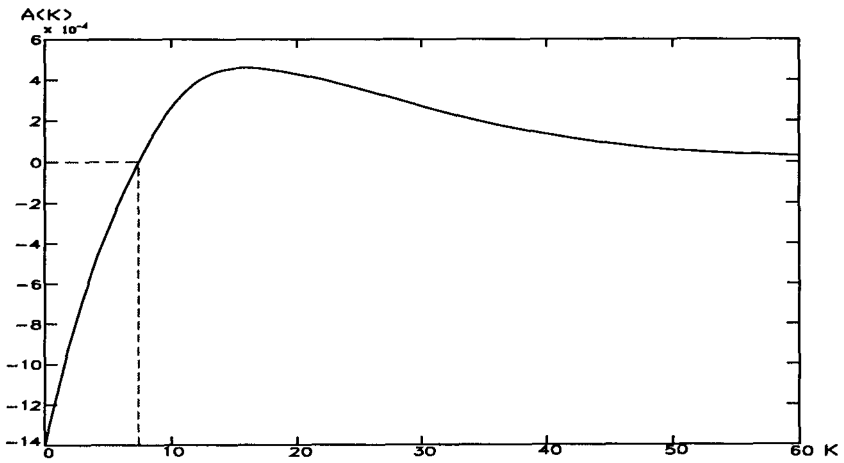

Figure 3. $\mu^2 C_1 < C_2 + \mu C_3, \quad (\lambda = 0.1, \mu = 5, C_1 = 1, C_2 = 50, C_3 = 5)$

APPENDIX

If $\mu^2 C_1 \geq C_2 + \mu C_3$, then, $A(K) \geq 0$, for all $K \geq 0$.

proof) Observing that

$$-s^4(\lambda\mu)^2 e^{sK}(C_2 + \mu C_3) \geq -s^4(\lambda\mu)^2 e^{sK}\mu^2 C_1,$$

we can see that

$$\begin{aligned}
A(K) &\geq [s^2(\lambda\mu)^2 e^{sK} - 2s^2(\lambda\mu)^3 e^{sK} + s^2(\lambda\mu)^4 e^{2sK} - s^3 K(\lambda\mu)^2 e^{sK}]C_1 \\
&\quad -s^4(\lambda\mu)^2 e^{sK}\mu^2 C_1 \\
&= s^2(\lambda\mu)^4 e^{2sK} C_1 - s^3 K(\lambda\mu)^2 e^{sK} C_1 - s^2(\lambda\mu)^4 e^{sK} C_1, \text{ since } s^2 = \left(\frac{\lambda\mu - 1}{\mu}\right)^2 \\
&\geq [s^3(\lambda\mu)^4 K e^{sK} - s^3(\lambda\mu)^2 K e^{sK}]C_1 \\
&= s^3[(\lambda\mu)^2 - 1](\lambda\mu)^2 K e^{sK} C_1 \\
&\geq 0
\end{aligned}$$

where the third inequality comes from the fact that $e^{sK} \geq 1 + sK$ and the fifth inequality follows from the fact that $s \geq 0 \Leftrightarrow \lambda\mu \geq 1$.

ACKNOWLEDGEMENT

This research was supported by the Sookmyung Women's University research grant, 2001.

REFERENCES

Bae, J. H., Kim, S. G. and Lee, E. Y. (2001). The virtual waiting time of the M/G/1 queue with impatient customers. *Queueing Systems*, **38**, 485-494.

Daley, D. J. (1965). General customer impatience in the queue GI/G/1. *Journal of Applied Probability*, **2**, 186-205.

De Kok, A. G. and Tijms, H. C. (1985). A queueing system with impatient customers. *Journal of Applied Probability*, **22**, 688-696.

Karlin, S. and Taylor, H. M. (1975). *A first course in stochastic processes.* Academic Press, New York.

Lee, E. Y. and Lim, K. E. (2000). The analysis of the M/M/1 queue with impatient customers. *The Korean Communications in Statistics*, **7**, 489-497.

Ross, S. M. (1996). *Stochastic Processes.* 2nd edition John Wiley & Sons, New York.