Operating Characteristics of $M^X/G/1$ Queue with $N$-Policy

Ho Woo Lee

Department of Industrial Engineering
Sungkyunkwan University
Su Won, KOREA 440-746

Soon Seok Lee
Switching Method Section
ETRI
Tae Jeon, KOREA 305-606

ABSTRACT

We consider a $M^X/G/1$ queueing system with $N$-policy. The server is turned off as soon as the system empties. When the queue length reaches or exceeds a predetermined value $N$ (threshold), the server is turned on and begins to serve the customers. We place our emphasis on understanding the operational characteristics of the queueing system. One of our findings is that the system size is the sum of two independent random variables: one is the PGF of the stationary system size of the $M^X/G/1$ queueing system without $N$-policy and the other one has the probability generating function $\sum_{j=0}^{N-1} \pi_j z^j / \sum_{j=0}^{N-1} \pi_j$ in which $\pi_j$ is the probability that the system state stays at $j$ before reaching or exceeding $N$ during an idle period. Using this interpretation of the system size distribution, we determine the optimal threshold $N$ under a linear cost structure.

Key Words: $M^X/G/1$ queue, $N$-policy, system size, optimal operating policy
1. Introduction

We consider $M^X/G/1$ queueing system with $N$-policy that operates as follows. Customers arrive according to a compound Poisson Process where the arrival size $X$ is a random variable. The server is turned off each time the system empties. When the queue length reaches or exceeds $N$ (threshold), the server is turned on and begins to serve the customers exhaustively. Customers are served one at a time by a single server. We will call this queueing system an $M^X/G/1/N$-policy queue.

The first study of the $M^X/G/1/N$-policy queue was by Lee and Srinivasan [2]. They presented a procedure to find the optimal stationary operating policy under a linear cost structure. To do this, they derived the mean waiting time which was expressed by the ratio of the first and second factorial moments of the system size at a service initiation epoch. Lee[3] developed an efficient algorithm to compute the steady-state probabilities. For control policies including $N$-policy and their applications, readers are urged to see Lee and Srinivasan and the references therein.

Neither Lee and Srinivasan nor Lee attempted to provide an understanding of the operational characteristics of the queueing system. Lee and Srinivasan did not explain the terms involved in the mean waiting time and the optimal control policy, apparently due to their complexity. Lee did not explain what the terms involved in the system size distribution meant. Generally speaking, in queueing theory, meaningful interpretations of the terms involved in some performance measures make it much easier to analyze other similar systems. This is what motivates our study.

In this study, we focus on explaining the operational characteristics of the $M^X/G/1/N$-policy queue. We derive and give appropriate interpretations for the system size distribution, the waiting time distribution and other performance measures. Finally we adopt the same linear cost structure as the one in Lee and Srinivasan and derive the optimal stationary operating policy which provides a more meaningful interpretation.
2. System size distribution

In this section, we first set up the system equations for its stationary distribution by employing the remaining service time as the supplementary variable. Then we solve the equations and derive the probability generating function (PGF) of the system size distribution.

We will use the following notations and probabilities:

- $N$: threshold
- $\lambda$: group arrival rate
- $X$: arrival size random variable
- $g_k$: $\Pr(X = k)$
- $X(z)$: the probability generating function of $X$
- $S$: service time random variable
- $s(x)$: the probability density function of $S$
- $S'(\Theta)$: the Laplace–Stieltjes transform (LST) of $S$
- $S'_t$: remaining service time of the customer in service at time $t$
- $N(t)$: system size at time $t$
- $Y(t)$: \begin{align*}
1 & \text{ if server is busy} \\
0 & \text{ if server is idle}
\end{align*}
- $P_n(x, t) dt = \Pr\{N(t) = n, x \leq S'_t(t) \leq x + dt, Y(t) = 1\}, \quad n = 1, 2, \ldots$
- $P_n(x) = \lim_{t \to 0} P_n(x, t)$
- $R_n(t) = \Pr\{N(t) = n, Y(t) = 0\}, \quad n = 0, 1, \ldots, N-1$
- $R_n = \lim_{t \to \infty} R_n(t)$

In steady-state, we have the following system equations,

\begin{align}
0 & = -\lambda R_0 + P_1(0), \\
0 & = -\lambda R_n + \lambda \sum_{k=1}^{n} R_{n-k} \cdot g_k, \quad (n = 1, 2, \ldots, N-1), \\
- \frac{d}{dx} P_1(x) & = -\lambda P_1(x) + P_2(0)s(x), \\
- \frac{d}{dx} P_m(x) & = -\lambda P_m(x) + P_{m+1}(0)s(x) + \lambda \sum_{k=1}^{m-1} P_{m-k}(x)g_k \quad (m = 2, 3, \ldots, N-1), \\
- \frac{d}{dx} P_n(x) & = -\lambda P_n(x) + P_{n+1}(0)s(x) + \lambda \sum_{k=1}^{n-1} P_{n-k}(x)g_k \\
& + \lambda s(x) \sum_{k=0}^{n-1} R_k \cdot g_{n-k}, \quad (n \geq N).
\end{align}
Taking the LST of both sides of the equations (2-3), (2-4), and (2-5), we have

$$\Theta P_m^\prime(\Theta) - P_m(0) = \lambda P_m^\prime(\Theta) - P_2(0)S^*(\Theta), \quad (m=2, 3, \ldots, N-1)$$

(2-6)

$$\Theta P_n^\prime(\Theta) - P_n(0) = \lambda P_n^\prime(\Theta) - P_{n+1}(0)S^*(\Theta) - \lambda \sum_{k=1}^{n-1} P_{n-k}(0)g_k, \quad (n \geq N).$$

(2-7)

Consider the probability generating functions

$$P^*(z, \Theta) = \sum_{n=1}^\infty P_n^*(\Theta)z^n,$$

$$P(z, 0) = \sum_{n=1}^\infty P_n(0)z^n.$$

After some algebraic manipulations with equations, (2-6), (2-7) and (2-8), it follows that,

$$[\Theta - \lambda + \lambda X(z)]P^*(z, \Theta) = P(z, 0) - S^*(\Theta) \left\{ \frac{P(z, 0)}{z} - P_1(0) + \lambda \sum_{n=N}^{\infty} \left\{ \sum_{k=0}^{N} R_k \cdot g_{n-k} \right\} z^n \right\}.$$

(2-9)

From equations (2-1) and (2-2), we have

$$\lambda \sum_{n=1}^{\infty} \left( \sum_{k=0}^{N-1} R_k \cdot g_{n-k} \right) z^n - P_1(0) = \lambda \left\{ \sum_{n=1}^{\infty} \left( \sum_{k=0}^{N-1} R_k \cdot g_{n-k} \right) z^n - R_0 \right\}$$

$$= \lambda \left\{ \sum_{k=0}^{N-1} R_k z^k \sum_{n=1}^{\infty} g_{n-k} z^n - R_0 \right\}$$

$$= \lambda \sum_{k=0}^{N-1} R_k z^k \left\{ X(z) - \sum_{n=1}^{N-1} g_{n-k} z^n \right\} - \lambda R_0$$

$$= \lambda \left\{ X(z) R(z) - \sum_{n=1}^{N-1} \sum_{k=1}^{n} g_{k} z^k R_{N-k-n} z^{N-k-n} - R_0 \right\}$$

$$= \lambda \left\{ X(z) R(z) - \sum_{n=1}^{N-1} R_{N-n} z^{N-n} - R_0 \right\}$$

$$= \lambda R(z) [X(z) - 1],$$

where $$R(z) = \sum_{n=0}^{N-1} R_n z^n.$$

Thus, equation (2-9) becomes
\[ [\Theta - \lambda + \lambda X(z)] P'(z, \Theta) = P(z, 0) - S'(\Theta) \left( \frac{P(z, 0)}{z} + \lambda R(z)[X(z) - 1] \right). \]  

(2-10)

Letting \( \Theta = \lambda - \lambda X(z) \), we have

\[ P(z, 0) = \frac{\lambda z S^*[\lambda - \lambda X(z)] [X(z) - 1] R(z)}{z - S^*[\lambda - \lambda X(z)]}. \]

Then from equation (2-10), we have

\[ P^*(z, \theta) = \frac{\lambda z S^* [\lambda - \lambda X(z)] - S'(\theta) [X(z) - 1] R(z)}{z - S^*[\lambda - \lambda X(z)]}. \]  

(2-11)

Thus the PGF of the system size distribution in steady state is

\[ P(z) = P^*(z, 0) + R(z) = \frac{(z - 1) S^*[\lambda - \lambda X(z)]}{z - S^*[\lambda - \lambda X(z)]} \cdot R(z). \]  

(2-12)

3. Analysis of the system size distribution

In this section we analyze the system size PGF \( P(z) \) and provide appropriate interpretations.

**THEOREM 3.1.** \( R(z) = R_0 \sum_{n=0}^{N-1} \pi_n z^n \), where \( \pi_0 = 1 \), and \( \pi_n = \sum_{i=1}^{\xi_n} g_i \cdot \pi_{n-i} \). Then the system size becomes

\[ P(z) = (1 - \rho) (z - 1) S^*[\lambda - \lambda X(z)] \frac{\sum_{n=0}^{N-1} \pi_n z^n}{z - S^*[\lambda - \lambda X(z)]} \]

\[ = P(z, M^\infty | G/1 \text{ without } N- \text{ policy}) \cdot \frac{\sum_{n=0}^{N-1} \pi_n z^n}{\sum_{n=0}^{N-1} \pi_n}. \]  

(3-1)

**PROOF.** First we show that \( R_\infty = R_0 \sum_{i=1}^{\xi_i} g_i \cdot \pi_{n-i} = R_0 \pi_n \) by mathematical induction. For
\( n = 1 \), it is obvious from equation (2.2). Now assume that it holds for some \( n \). Then

\[
R_{n+1} = \sum_{i=1}^{n+1} R_{n+1-i} \cdot g_i = \sum_{i=1}^{n+1} (R_0 \cdot \pi_{n+1-i}) g_i = R_0 \pi_{n+1}.
\]

Multiplying both sides by \( z^n \) and summing over \( n \) from 0 to \( N-1 \) yields

\[
R(z) = R_0 \sum_{n=0}^{N-1} \pi_n z^n.
\]

From \( P(1) = 1 \), we have \( R_0 = \frac{1 - \rho}{\sum_{n=0}^{N-1} \pi_n} \). Finally we have

\[
P(z) = \frac{(1 - \rho)(z-1)S^*[\lambda - \lambda X(z)]}{z - S^*[\lambda - \lambda X(z)]} \cdot \frac{\sum_{n=0}^{N-1} \pi_n z^n}{\sum_{n=0}^{N-1} \pi_n}.
\]

**REMARK 3.1.** From **THEOREM 1**, we see that the stationary system size of \( M^X|G|1/N- \) policy queue is the sum of two independent random variables one of which is the system size of ordinary \( M^X|G|1 \) queue. The interpretation of the other random variable will be given in **THEOREM 3.4**.

**THEOREM 3.2.** \( \pi_j, j = 0, 1, 2, \ldots, N-1, \) is the probability that the system state (number of customers in the system) visits \( j \) during an idle period.

**PROOF.** Let \( I_j = \begin{cases} 1 & \text{if the state } j \text{ is visited during an idle period,} \\ 0 & \text{o/w.} \end{cases} \)

Conditioning on the first arrival size, we have

\[
Pr(I_j = 1) = g_j + \sum_{k=1}^{j-1} g_k \cdot Pr(I_{j-k} = 1).
\]

Letting \( Pr(I_j = 1) = \pi_j \) and \( \pi_0 = 1 \), we have \( \pi_j = \sum_{k=1}^{j} g_k \cdot \pi_{j-k} \) which satisfies **THEOREM 3.1**.
THEOREM 3.3. \[ \sum_{j=0}^{N-1} \pi_j \] is the mean number of arriving groups during an idle period.

PROOF. From THEOREM 3.2, \[ \frac{N-1}{j=0} I_j \] is the number of states visited during an idle period, or equivalently is the number of arriving groups during an idle period. Then,

\[ \mathbb{E} \left[ \sum_{j=0}^{N-1} I_j \right] = \sum_{j=0}^{N-1} \mathbb{E}(I_j) = \sum_{j=0}^{N-1} \Pr(I_j = 1) = \sum_{j=0}^{N-1} \pi_j. \]

THEOREM 3.4. Let \( \alpha_j = \frac{\pi_j}{\sum_{n=0}^{N-1} \pi_n} \), \( j = 0, 1, 2, \ldots, N-1 \). Then \( \alpha_j \) is the probability that there are \( j \) customers in the system given that the server is idle.

PROOF. By the Poisson process, the average staying time in any state during an idle period is \( 1/\lambda \). Thus we have

\[ \frac{\pi_j}{\sum_{n=0}^{N-1} \pi_n} = \frac{1/\lambda}{\sum_{n=0}^{N-1} \pi_n}. \]

By the renewal reward theorem, the result follows.

4. Other performance measures

In this section we derive some other performance measures.

THEOREM 4.1. Let \( Q_N \) be the queue size at service initiation point under the \( N \)-policy. Then the PGF of \( Q_N \) is given by
\[ Q_N(z) = 1 + [X(z) - 1] \Pi_N(z), \]

(4-1)

where \( \Pi_N(z) = \sum_{j=0}^{N-1} \pi_j z^j \).

The first and second factorial moments become

\[ E(Q_N) = E(X) \sum_{n=0}^{N-1} \pi_n, \]
\[ E(Q_N(Q_N-1)) = E(X^2 - X) \sum_{n=0}^{N-1} \pi_n + 2E(X) \sum_{n=0}^{N-1} n \pi_n. \]

**PROOF.** The proof is by mathematical induction. Conditioning on the first arrival size, we have the recursive equation

\[ Pr(Q_n = k) = g_k + \sum_{j=1}^{N-1} g_j \cdot Pr(Q_{n-j} = k-j), \quad (k \geq n). \]

For \( n = 1 \), \( Q_1(z) = 1 + [X(z) - 1] \pi_0 = X(z) \). Assuming that it works for \( n = 2, \ldots, N-1 \), let us see the case of \( n = N \),

\[ Q_N(z) = \sum_{k=N}^{\infty} Pr(Q_N = k) z^k \]
\[ = \sum_{k=N}^{\infty} \left[ g_k + \sum_{j=1}^{N-1} g_j \cdot Pr(Q_{N-j} = k-j) \right] z^k \]
\[ = X(z) + \sum_{j=1}^{N-1} g_j z^j [Q_{N-j}(z) - 1] \]
\[ = X(z) + [X(z) - 1] \sum_{j=1}^{N-1} g_j z^j \sum_{i=0}^{N-1} \pi_i z^i \]
\[ = X(z) + [X(z) - 1] \sum_{j=1}^{N-1} z^j \sum_{i=1}^{N-1} g_i \pi_{j-i} \]
\[ = X(z) + [X(z) - 1] \sum_{j=1}^{N-1} z^j \pi_j \]
\[ = 1 + [X(z) - 1] \Pi_N(z). \]

Then the first and second factorial moments become

\[ E(Q_N) = Q_N(1) = E(X) \sum_{n=0}^{N-1} \pi_n, \]
\[ E(Q_N(Q_N-1)) = Q_N(1) = E(X^2 - X) \sum_{n=0}^{N-1} \pi_n + 2E(X) \sum_{n=0}^{N-1} n \pi_n. \]
THEOREM 4.2. Let $T'_N(\theta)$ be the LST of the idle period. Then we have

$$T'_N(\theta) = \frac{\lambda}{\lambda + \theta} \left[ 1 + \sum_{k=1}^{N-1} g_k \left( T'_{N-k}(\theta) - 1 \right) \right].$$

(4-2)

The mean idle period is given by $E(T_N) = \frac{1}{\lambda} \sum_{j=0}^{N-1} \pi_j$.

**PROOF.** Conditioning on the first arrival size, we have

$$Pr(T_N = t) = \lambda e^{-\lambda t} \sum_{k=N}^\infty g_k + \int_0^t \left\{ \lambda e^{-\lambda x} \sum_{k=1}^{N-1} g_k \cdot Pr(T_{N-k} = t-x) \right\} dx dt.$$ 

Taking the LST of both sides, we have the desired result. The mean idle period easily follows.

THEOREM 4.3. Let $T_c$ be the cycle length. Then

$$E(T_c) = \frac{1}{\lambda(1-p)} \sum_{j=0}^{N-1} \pi_j,$$

where $p = \lambda E(X) E(S)$.

**PROOF.** Let $B_N$ and $B'_N(\theta)$ be the busy period random variable and its LST. Then from

THEOREM 4.1, we have

$$B'_N(\theta) = \sum_{k=N}^\infty Pr(Q_N = k) [B^*(\theta)]^k = Q_N[B^*(\theta)],$$

where $B^*(\theta)$ is the well-known LST of the busy period of ordinary $M^X/G/1$ queue started with one customer which is given by

$$B^*(\theta) = S^* \left[ \theta + \lambda - \lambda X(B^*(\theta)) \right].$$

Then the mean busy period is easily obtained as
\[ E(B_N) = -B^*_N(0) = -Q'_N[B^*(0)]B^*(0) \]
\[ = E(Q'_N) \frac{E(S)}{1 - \rho} \]
\[ = \frac{E(X)E(S)}{1 - \rho} \sum_{j=0}^{N-1} \pi_j \]

Thus
\[ E(T_C) = E(B_N) + E(T_N) = \frac{1}{\lambda(1 - \rho)} \sum_{j=0}^{N-1} \pi_j \]

5. Queue Waiting Time

In this section we derive the queue waiting time of an arbitrary customer at random point of time. In ordinary \( M^X/G/1 \) queue, it is well known that the queue waiting time of the test customer can be obtained by summing the queue waiting time of the first customer in his group and the additional waiting time due to those preceding him in his group. In our system under study, however, obtaining queue waiting time is not so simple because the 'remaining idle period' affects the queue waiting time. We define the 'remaining idle period' as the time period from the moment in an idle period the test customer arrives until the system state reaches or exceeds \( N \) and the server begins to be busy. With these in mind, let us define following \( LSTs \) and events:

- \( W^q_\theta(\theta) \) \( LST \) of the queue waiting time of an arbitrary customer,
- \( W^b_\theta(\theta) \) \( LST \) of the queue waiting time of the customer who arrives while the server is busy,
- \( W^i_\theta(\theta) \) \( LST \) of the queue waiting time of the customer who arrives while the server is idle,
- \( W^a_\theta(\theta) \) \( LST \) of the additional queue waiting time due to those preceding the test customer in the same group,
- \( W^u_\theta(\theta) \) \( LST \) of the queue waiting time of the test customer who arrives while the server is idle and sees \( j \) customers in the queue,
- \( A_j \) Event that a group arrives to the idle system and finds \( j \) customers waiting in the queue, \( j = 0, 1, \cdots, N - 1 \),
$H_r$ Event that the test customer belongs to a group of size $r$;

$E_{i,r}$ Event that the test customer is in the $i^{th}$ position under the condition that the group size is $r$.

It is well known (Burke [1]) that

$$P_r(E_{i,r}) = 1/r,$$

$$P_r(H_r) = 1 \cdot \frac{E_r}{E(X)}.$$

Thus we have

$$W_A(\theta) = \sum_{r=1}^{\infty} \sum_{i=1}^{r} [S^*(\theta)]^{-1} (1/r) \cdot \frac{E_r}{E(X)} = \frac{1}{E(X)[1 - S^*(\theta)]}.$$

(5-1)

First, let us consider the case in which the server is busy. In this case, the arriving test customer who sees $j$ customers in the system must wait "the remaining service time of the customer in service, plus service times of those $j-1$ customers in the queue, plus additional waiting time due to those preceding him in the same group". From equations (2-11), (5-1), THEOREM 1.1, and PASTA (Wolff [5]), we have

$$W_B(\theta) = \sum_{r=1}^{\infty} P_r(\theta)[S^*(\theta)]^{-1} \cdot W_A(\theta)$$

$$= \frac{P^*(S^*(\theta), \theta)}{S^*(\theta)} \cdot \frac{1 - X[S^*(\theta)]}{E(X)[1 - S^*(\theta)]}$$

$$= \frac{(1 - \rho)[\lambda - \lambda X(S^*(\theta))]}{\theta - \lambda + \lambda X(S^*(\theta))} \cdot \sum_{n=0}^{N-1} \pi_n[S^*(\theta)]^n \cdot \frac{1 - X[S^*(\theta)]}{E(X)[1 - S^*(\theta)]}.$$

(5-2)

The most complicated part of the queue waiting time comes from the case where the test customer arrives when the server is idle. If the test customer finds the server idle and sees $j$ customers waiting in the queue ($j \leq N - 1$), he has to wait

(1) those $j$ service times

plus (2) the service times of those preceding him in the same group

plus (3) remaining idle period.

It is to be noted that (2) and (3) are dependent because the test group size affects the
remaining idle period. To evaluate the waiting time due to (2), if the test customer belongs to a
group of size \( r \) and stands \( i^{th} \) within the group (with probability \( 1/r \)), he has to wait those
\((i-1)\) service times and in this case the LST of the remaining idle period becomes
\( T_{N-j-r}(\theta) \) due to the Poisson arrival. But note that remaining idle period becomes zero if the
test customer belongs to the group of size \( N-j \) or more. Thus we have

\[
W^*_{(j)}(\theta) = [S^*(\theta)]^j \left[ \sum_{r=1}^{N-j-1} \frac{r \cdot g_r}{E(X)} \frac{1}{r} [S^*(\theta)]^{i-1} T_{N-j-r}(\theta) \right.
\]

\[
+ \sum_{r=N-j}^{\infty} \frac{r \cdot g_r}{E(X)} \frac{1}{r} [S^*(\theta)]^{i-1} \right] . \tag{5-3}
\]

Then we have, after some algebraic manipulation,

\[
W^*_j(\theta) = \sum_{j=0}^{N-1} \sum_{k=0}^{j} \pi_k [S^*(\theta)]^j \left[ \sum_{r=1}^{N-j-1} \frac{r \cdot g_r}{E(X)[1-S^*(\theta)]} \right]
\]

\[
\left[ T_{N-j-r}(\theta) - 1 \right]
\]

\[
+ \frac{1-X(S^*(\theta))}{E(X)[1-S^*(\theta)]} \right] . \tag{5-4}
\]

Finally the LST of the queue waiting time of an arbitrary customer becomes

\[
W_q(\theta) = W^*_q(\theta) + W^*_q(\theta)
\]

\[
= \frac{\theta(1-\rho)[1-X(S^*(\theta))]}{E(X)[1-S^*(\theta)][\theta - \lambda + \lambda X(S^*(\theta))]} \cdot W^*_q(\theta)
\]

= \( W^*_q(\theta, M^X/G/1 \text{ without threshold}) \cdot W^*_q(\theta) \).
\]

\[
\text{(5-5)}
\]

where

\[
W^*_q(\theta) =
\]

\[
\sum_{j=1}^{N-1} \left[ \frac{\pi_j [S^*(\theta)]^j}{\sum_{j=0}^{N-1} \pi_j} \right]
\]

\[
\left[ 1 + \frac{1-\rho}{W^*_q(\theta, M^X/G/1)} \frac{X(S^*(\theta))}{E(X)[1-S^*(\theta)]} \right] . \tag{5-5}
\]

The mean queue waiting time becomes

\[
W_q = -W^*_q(0, M^X/G/1) - W^*_q(0)
\]

11
Notice that the first two terms are the mean queue waiting time of ordinary $M^X/G/1$ queue without threshold and the third term is the additional waiting time due to the threshold policy. The mean queue size can be obtained from the well-known Little's formula,

\[
L_q = \lambda E(X) W = \frac{\lambda E(X)}{2(1-\rho)} + \sum_{n=0}^{N-1} n \pi_n
\]

and the mean system size is obtained by $L = \lambda E(X) W = L_q + \rho$.

6. Optimal design of the $N$-policy

In this section, we find the optimal $N^*$ that minimizes the long-run average cost. We consider following costs,

- $C_s$: start-up cost / cycle,
- $C_h$: holding cost / unit time,
- $C_o$: operating cost / unit time.

Yadin and Naor[4] and Lee and Srinivasan[2] considered similar cost structure. The average cost per unit time, $TC(N)$, is obtained as:

\[
TC(N) = \frac{C_s}{E(T_C)} + C_h \cdot L + \rho \cdot C_o
\]

\[
= \frac{\lambda C_s(1-\rho) + C_h \sum_{j=0}^{N-1} j \pi_j}{\sum_{j=0}^{N-1} \pi_j} + C_h \cdot L(M^X/G/1) + \rho \cdot C_o.
\]
It is hard to prove that $TC(N)$ is convex. But we now present a procedure that makes it possible to calculate the optimal threshold $N^*$.

**Theorem 6.1.** Under the long-run expected average cost criterion, the optimal threshold for the $MX|G|1/N-$ policy queue is given by

$$N^* = \min \left\{ k \geq 1 \mid \sum_{j=0}^{N} (k-j) \pi_j > \lambda C_s \frac{(1-\rho)}{C_h} \right\}.$$

**Proof.** Let $J_k = \sum_{j=0}^{k} \pi_j$ and $M_k = \sum_{j=0}^{k} j \pi_j$. Then, we have

$$TC(k+1) - TC(k) = \lambda C_s (1-\rho) \left\{ \frac{1}{J_k} - \frac{1}{J_{k-1}} \right\} + C_s \left\{ \frac{M_k - M_{k-1}}{J_k} \right\}$$

$$= \frac{\pi_k}{J_k - J_{k-1}} \left[ C_h (k J_k - M_k) - \lambda C_s (1-\rho) \right].$$

Since $C_h (k J_k - M_k) > 0$ and $\frac{\pi_k}{J_k - J_{k-1}} > 0$, the sign of $h(k) = C_h (k J_k - M_k) - \lambda C_s (1-\rho)$ determines whether $TC(k)$ increases or decreases. Let $m$ be the first $k$ such that $h(k) > 0$. Then

$$h(m+1) = C_h \left\{ (m+1) J_{m+1} - M_{m+1} \right\} - \lambda C_s (1-\rho)$$

$$= C_h \left\{ (m+1) J_m - M_m \right\} - \lambda C_s (1-\rho)$$

$$= C_h \left\{ m J_m - M_m \right\} - \lambda C_s (1-\rho) + C_h J_m > 0.$$

Thus we see that $TC(n) > TC(m)$ for $n > m$. Finally we have

$$N^* = \text{first } k \text{ such that } h(k) > 0$$

$$= \min \left\{ k \geq 1 \mid \sum_{j=0}^{N} (k-j) \pi_j > \lambda C_s \frac{(1-\rho)}{C_h} \right\}.$$


13
**REMARK 6.1.** It is worthwhile to note that if \( \frac{C_h}{C_z} > \frac{\lambda(1 - \rho)}{1} \), the optimal threshold value \( N^* \) is always equal to 1. This implies that if the holding cost per unit time is greater than the start-up cost per unit time, it is not beneficial to have a control policy (note that \( \frac{1}{\lambda(1 - \rho)} \) is the expected cycle length for the \( M^X/G/1 \) queue without threshold). For single unit arrival queues, the optimal threshold \( N^* \) is the same as the result of Yadin and Naor\[4\], which becomes

\[
N^* = \sqrt{\frac{2\lambda(1 - \rho)C_z/C_h}.}
\]

**REMARK 6.2.** This procedure is basically the same as the one developed by Lee and Srinivasan \[2\]. But their cost function is expressed by the ratio of the first and second factorial moments of the system size at a service initiation point, which makes it hard to provide meaningful interpretation.

Acknowledgement

The authors would like to thank Dr. Takagi of IBM Japan for his helpful comments. The authors also would like to thank the anonymous referee for improving the readability of this paper.

REFERENCES


14