Applied Mathematics, 2015, 6, 908-921
Published Online June 2015 in SciRes. http://www.scirp.org/journal/am
http://dx.doi.org/10.4236/am.2015.66083

Scientific Research Publishing

Stationary Analysis of Geo/Geo/1 Queue with Two-Speed Service and the Optimal Switching Threshold for the Service Rate

Xudong Lin

School of Science, Sichuan University of Science and Engineering, Zigong, China
Email: linxd27@163.com

Received 17 April 2015; accepted 30 May 2015; published 2 June 2015

Abstract

This paper considers a Geo/Geo/1 queueing system with infinite capacity, in which the service rate changes depending on the workload. Initially, when the number of customers in the system is less than a certain threshold *L*, low service rate is provided for cost saving. On the other hand, the high service rate is activated as soon as *L* customers accumulate in the system and such service rate is preserved until the system becomes completely empty even if the number of customers falls below *L*. The steady-state probability distribution and the expected number of customers in the system are derived. Through the first-step argument, a recursive algorithm for computing the first moment of the conditional sojourn time is obtained. Furthermore, employing the results of regeneration cycle analysis, the direct search method is also implemented to determine the optimal value of *L* for minimizing the long-run average cost rate function.

Keywords

Workload-Dependent Service, Switching Threshold, Discrete-Time Queue, Sojourn Time, Regeneration Cycle

1. Introduction

In the classical queueing literature, the server is usually assumed to work at constant speed as long as there is any work present. However, we know that this assumption may not always be appropriate when the system's workload affects the server's efficiency in some real world situations. To better understand this fact, we can cite some practical examples to illustrate this point. In a manufacturing system, the decision-maker is responsible for deciding the service speed of the production equipment according to the level of market demand. If the current

How to cite this paper: Lin, X.D. (2015) Stationary Analysis of Geo/Geo/1 Queue with Two-Speed Service and the Optimal Switching Threshold for the Service Rate. *Applied Mathematics*, **6**, 908-921. http://dx.doi.org/10.4236/am.2015.66083

production capacity is far from meeting market demand, high service rate will be activated to balance the requirements. Nonetheless, once the demand is satisfied and decreases significantly, production will be slowed down to avoid inventory pile up. In addition, the telephone-based directory assistance is another convincing example of service rate depending on the queue length, where as the number of calls increases, the provision of extra attendants is recommended so as to provide better quality of service in terms of reduced waiting time. However, these extra attendants may be removed when the peak time is over and the number of phone calls sharply reduces. Therefore, these real-life applications that mentioned above constitute the main motivation of our study.

Actually, there is a considerable body of queueing literature that deals with workload-dependent service rate. Among some early papers in this area are those by Satty [1] and Gebhard [2], both of whom considered some fundamental queueing problems such as the stationary queue size distribution and the expected queue length for the $M/M/1$ queue. Their work spawned research into modeling a queueing system with adaptable service rate, such as by Gross and Harris [3], Harris and Marchal [4], William and Wang [5], Bekker *et al*. [6] and Zhernovyi [7]. In the past several decades, an important extension to the above model is the multi-server queueing system with queue-dependent servers. Singh [8] respectively analyzed the infinite source $M/M/2$ queueing systems with two homogeneous and heterogeneous servers. A relationship among the system operating costs, traffic intensity and the queue size is obtained. Later, based on the Singh's pioneering work, Garg and Singh [9] revisited the same system and established a cost structure to determine the optimal queue length at which the second server was provided, so that the system may gain the maximum profit. With the assumption that the system capacity was limited, Wang and Tai [10] studied queue-dependent servers in the finite buffer $M/M/3$ queue with three types of service rate. They constructed a relationship among the costs to determine the optimal queue lengths J and K of providing the second server and the third server, respectively. Furthermore, not long ago, Jain [11] investigated the finite capacity $M/M/r$ queueing system with r heterogeneous servers. In particular, the optimal threshold parameters for turning on the servers were obtained in her work. More recently, $M/M/r$ ueueing model with infinite capacity and queue-dependent servers was considered by Lin and Ke [12]. Using the genetic algorithm, they found the best thresholds of queue length in activating servers and their corresponding service rate. These studies greatly enhance the practical value of the multi-server queueing theory since it is realistic to consider the changes in the number of working servers.

However, we may note a common feature existing in the above research works, namely, authors invariably assum that whenever the number of customers or jobs in the system exceeds a certain threshold, the service rate is accelerated to deal with the lengthy queue. Further, if the queue length reduces to less than the threshold, lower service rate is resumed. In fact, such model assumption means that the service rate can be switched countlessly in a regeneration cycle. Here, for the single server queue, the regeneration cycle is the time span between two consecutive starting points of the server's idle period. Obviously, whenever a server is switched from low service rate to high service rate, or *vice-versa*, switching cost is incurred. The more the server switches its service rate, the more additional cost it has to face. In other words, if the switch is reiterated over a long period of time, substantial amount of switching cost will be charged to the system. Therefore, the traditional service rate switching policy has some significant drawbacks in the queueing system with a relatively high arrival rate. In order to prevent switches from occurring too frequently, a modified service rate switching policy is proposed in this paper. Under the control of modified switching policy, the high service rate is activated as soon as L customers accumulate in the system and such service rate is preserved until the system becomes completely empty even if the number of customers falls below L. Hence, for the modified switching policy, the change of service rate can only occur at most once in a regeneration cycle. Undoubtedly, this policy will greatly reduce the switching cost of the system. On the other hand, although a lot of continuous-time queueing models with workload-dependent service rate have been studied extensively in the past years, their discrete-time counterparts received very little attention in the literature. Except the studies done by Chaudhry [13] [14] and Parthasarathy and Lenin [15], no work in this direction has come to our notice. Given that the wide applications of discrete-time queue in digital data networks and flexible manufacturing systems, in this paper, we will develop an analytical model that allows us to extensively analyze and explore the Markovian queueing system with workload-dependent service rate in discrete-time case. Through our work, we wish to develop a computational model that helps decision-makers answer the following important questions: 1) Under a certain cost structure, what is the optimal value of L that minimizes the long-runaverage cost rate function? 2) If the system state information is communicated to the customers upon their arrival, how to evaluate the expected conditional sojourn time of an arriving cus-

tomer?

The rest of this paper is organized as follows. In Section 2, we describe the mathematical model for the problem under consideration. The steady-state analysis of the model is presented in Section 3 and some important system performance measures are derived in this section. Using the first-step argument, we develop an analytical scheme for the customer's sojourn time. Furthermore, we also carried out regeneration cycle analysis to find the expected length of two types of busy periods. In Section 4, a long-run average cost rate function is established based on the system characteristics to determine the optimal switching threshold for the service rate. Section 5 concludes the research and suggests some future topics.

2. Model Formulation

We consider a discrete-time queue with single server or machine, whose service rate may be affected by the number of customers or jobs present in the system. In our model, inter-arrival times $A_1, A_2, \cdots,$ are independent and identically distributed (i.i.d.) random variables with probability mass function (p.m.f.) $\Pr\{A_i = k\} = \bar{\lambda}^{k-1}\lambda, 0 < \lambda < 1, \bar{\lambda} = 1-\lambda$ and $i, k \geq 1$. Arriving customers form a single waiting line based on the order of their arrival. Initially, when the number of customers in the system is less than the given threshold level $L(L \geq 1)$, the server serves customers with low service rate. The high service rate is activated at the instant when the number of customers in the system becomes equal to L, and it will be preserved until the long queue empties. Here, we assume that the two types of service times, namely S_{low} and S_{high}, are independent and geometrically distributed with respective p.m.fs

$$\Pr\{S_{\text{low}} = k\} = \bar{\mu}_1^{k-1}\mu_1, \ k \geq 1, 0 < \mu_1 < 1,$$

$$\Pr\{S_{\text{high}} = k\} = \bar{\mu}_2^{k-1}\mu_2, \ k \geq 1, \ 0 < \mu_2 < 1,$$

where $\bar{\mu}_1 = 1-\mu_1$, $\bar{\mu}_2 = 1-\mu_2$ and $\mu_1 < \mu_2$. λ, μ_1 and μ_2 denote the customer arrival rate, low service rate and high service rate, respectively.

In discrete-time queueing system, the time axis is divided into equal intervals called slots and all queueing activities occur at the slot boundaries. Traditionally, there are two types of systems in the discrete-time case (see [16] and [17]), one is the late arrival with delayed access (LAS-DA) and the other is the early arrival system (EAS). In this paper, we consider the model for the late arrival system with delayed access and therefore, a potential arrival occurs in (t^-, t), and a potential departure takes place in (t, t^+), for $t = 0,1,2,\cdots$. To make it clear, the various time epochs at which events occur are shown in a self-explanatory figure (see **Figure 1**).

3. Steady-State Analysis

In this section, we first apply the Markov process theory to obtain the steady-state difference equations governing the system. Next, the generating function technique and a recursive method are employed to develop the analytical solutions in a neat close-form. Toward this end, we need to define some commonly used notations to analyze the queueing system as follows:

$N(t) \equiv$ the number of customers in the system at time t^-,

$Y(t) \equiv$ the server speed state at time t^-,

where

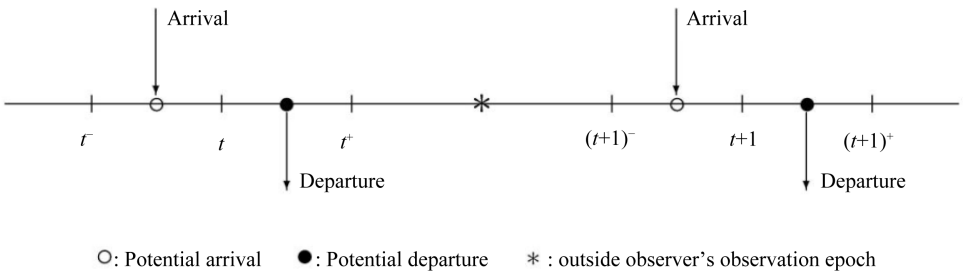

Figure 1. Various time epochs in late arrival system with delayed access (LAS-DA).

$$Y(t)=\begin{cases}0, \text{the single server provides service to the customers at a low speed,}\\ 1, \text{the single server provides service to the customers at a high speed.}\end{cases}$$

Therefore, $\xi(t)=\{(N(t),Y(t)),t=0,1,2,\cdots\}$ is the Markov chain for queueing system with state space

$$\Omega=\{(i,0)|i=0,1,\cdots,L-1\}\cup\{(i,1)|i=1,2,\cdots\}.$$

Furthermore, let us define the following stationary probability distributions for the Markov chain:

$$P_{i,0}=\lim_{t\to\infty}P_{i,0}(t)=\lim_{t\to\infty}\Pr\{N(t)=i,Y(t)=0\},\quad i=1,2,\cdots,L-1,$$

$$P_{i,1}=\lim_{t\to\infty}P_{i,1}(t)=\lim_{t\to\infty}\Pr\{N(t)=i,Y(t)=1\},\quad i=0,1,2,\cdots.$$

3.1. Steady-State Equation

From the state-transition-rate diagram for the Geo/Geo/1 queue with service rate switching threshold (see **Figure 2**), we can set up steady-state equations for $P_{i,0}$ and $P_{i,1}$ in the following:

$$\lambda P_{0,0}=\bar{\lambda}\mu_1P_{1,0}+\bar{\lambda}\mu_2P_{1,1}, \tag{1}$$

$$(1-\bar{\lambda}\bar{\mu}_1-\lambda\mu_1)P_{1,0}=\lambda P_{0,0}+\bar{\lambda}\mu_1P_{2,0}, \tag{2}$$

$$(1-\bar{\lambda}\bar{\mu}_1-\lambda\mu_1)P_{i,0}=\lambda\bar{\mu}_1P_{i-1,0}+\bar{\lambda}\mu_1P_{i+1,0},\ i=2,\cdots,L-2, \tag{3}$$

$$(1-\bar{\lambda}\bar{\mu}_1-\lambda\mu_1)P_{L-1,0}=\lambda\bar{\mu}_1P_{L-2,0}, \tag{4}$$

$$(1-\bar{\lambda}\bar{\mu}_2-\lambda\mu_2)P_{1,1}=\bar{\lambda}\mu_2P_{2,1}, \tag{5}$$

$$(1-\bar{\lambda}\bar{\mu}_2-\lambda\mu_2)P_{i,1}=\lambda\bar{\mu}_2P_{i-1,1}+\bar{\lambda}\mu_2P_{i+1,1},\ i=2,\cdots,L-1, \tag{6}$$

$$(1-\bar{\lambda}\bar{\mu}_2-\lambda\mu_2)P_{L,1}=\lambda\bar{\mu}_1P_{L-1,0}+\bar{\lambda}\mu_2P_{L-1,1}+\bar{\lambda}\mu_2P_{L+1,1}, \tag{7}$$

$$(1-\bar{\lambda}\bar{\mu}_2-\lambda\mu_2)P_{i,1}=\lambda\bar{\mu}_2P_{i-1,1}+\bar{\lambda}\mu_2P_{i+1,1},\ i=L+1,L+2,\cdots. \tag{8}$$

Remark 1. The Markov chain $\xi(t)$ is called stable if it is irreducible and all states are ergodic. As illustrated in **Figure 2**, the state space of this queueing system is a single communicating class. Thus, the Markov chain $\xi(t)$ is irreducible. Under assumption that $\mu_1<\mu_2$, it is clear from the standard Geo/Geo/1 theory that $\lambda/\mu_2<1$ is a necessary and sufficient condition for ergodicity of the system. Intuitively speaking, if, on average, arrivals happen faster than service completions the queue will grow indefinitely long and the system will not have a stationary distribution.

Remark 2. Relating the state probabilities at epochs *t* and $t+1$ or t^+ and $(t+1)^+$, and letting $t\to\infty$, we can easily obtain a set of difference equations, which have exactly the same mathematical form as Equations (1)-(8). So the stationary probabilities of the system state at epochs *t* and t^+ are identical with the one at epoch

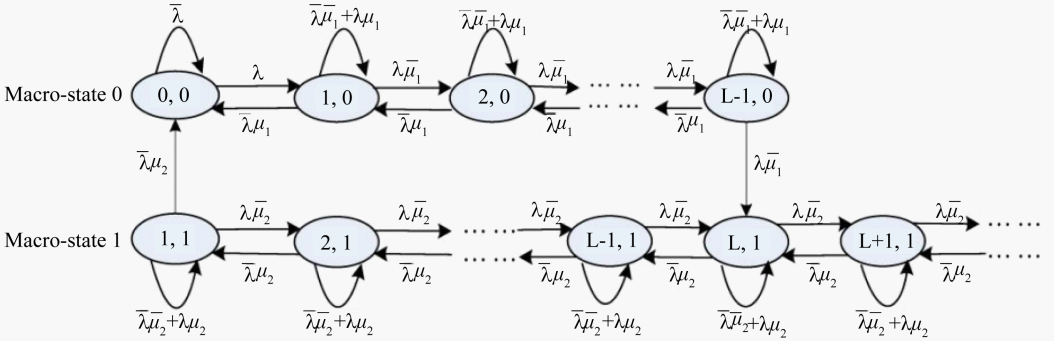

Figure 2. State-transition-rate diagram for the Geo/Geo/1 queue with switching threshold for service rate.

t^-. Furthermore, in LAS-DA, since an outside observer's observation epoch falls in a time interval after a potential departure and before a potential arrival, the probability that outside observer sees i customers in the system and the server in state j is also the same as $P_{i,j}\,(i=0,1,2,\cdots,\ j=0,1)$. For these reasons, only the system state probability at time point t^- is considered in this paper.

3.2. Two Relationships between $P_{0,0}$ and $P_{L-1,0}$

In this subsection, we first derive two important relationships between $P_{0,0}$ and $P_{L-1,0}$. On this basis, we also give the explicit expression for the stationary probability $P_{0,0}$. In later section, we will see that this quantity is very useful when the cost structure will be introduced in our model. To this end, let us first define the following probability generating functions:

$$\tilde{P}_0(z)=\sum_{k=1}^{L-1}P_{k,0}z^k,\ |z|\le 1,\ \tilde{P}_1(z)=\sum_{k=1}^{\infty}P_{k,1}z^k,\ |z|\le 1,$$

$$\tilde{P}(z)=P_{0,0}+\tilde{P}_0(z)+\tilde{P}_1(z).$$

Multiplying Equations (1)-(4) by z^n and summing over n, $n\ge 0$, we obtain

$$\tilde{P}_0(z)=\frac{\lambda(z-1)P_{0,0}+\bar{\lambda}\mu_2P_{1,1}-\lambda\bar{\mu}_1P_{L-1,0}z^L}{z^{-1}(z-1)\left(\bar{\lambda}\mu_1-z\lambda\bar{\mu}_1\right)}. \tag{9}$$

If we add the right hand sides and left hand sides of the Equations (1)-(4) and cancel the common terms, the following equality holds:

$$\bar{\lambda}\mu_2P_{1,1}=\lambda\bar{\mu}_1P_{L-1,0}. \tag{10}$$

Remark 3. As shown in **Figure 2**, according to the different service rates provided by the server, the state space of the queueing system is divided into two macro-states, namely 0 and 1. They are accessible to each other. Thus, the mean transition rate from state $(L-1,0)$ to state $(L,1)$ is equal to the mean rate from state (1,1) to state (0,0). This fact is properly reflected in the above Equation (10).

Substituting Equation (10) into Equation (9) and after some algebraic manipulation, we have

$$\tilde{P}_0(z)=\frac{\lambda P_{0,0}z+\lambda\bar{\mu}_1P_{L-1,0}\dfrac{z\left(1-z^L\right)}{z-1}}{\bar{\lambda}\mu_1-z\lambda\bar{\mu}_1}=\frac{\lambda P_{0,0}z-\lambda\bar{\mu}_1P_{L-1,0}\sum_{k=1}^{L}z^k}{\bar{\lambda}\mu_1-z\lambda\bar{\mu}_1}. \tag{11}$$

Using a method similar to the derivation of Equation (11), with Equations (5)-(8), we get

$$\tilde{P}_1(z)=\frac{\lambda\bar{\mu}_1P_{L-1,0}\dfrac{z\left(1-z^L\right)}{1-z}}{\bar{\lambda}\mu_2-z\lambda\bar{\mu}_2}=\frac{\lambda\bar{\mu}_1P_{L-1,0}\sum_{k=1}^{L}z^k}{\bar{\lambda}\mu_2-z\lambda\bar{\mu}_2}. \tag{12}$$

Thus, we can rewrite $\tilde{P}(z)$ as follows:

$$\tilde{P}(z)=P_{0,0}+\frac{\lambda P_{0,0}z-\lambda\bar{\mu}_1P_{L-1,0}\sum_{k=1}^{L}z^k}{\bar{\lambda}\mu_1-z\lambda\bar{\mu}_1}+\frac{\lambda\bar{\mu}_1P_{L-1,0}\sum_{k=1}^{L}z^k}{\bar{\lambda}\mu_2-z\lambda\bar{\mu}_2}. \tag{13}$$

Since $\tilde{P}(z)$ is the probability generating function of the queue length distribution, we have $\tilde{P}(1)=1$. By taking into account the normalization condition, and letting $z=1$ in Equation (13), we have

$$P_{0,0}+\frac{\lambda P_{0,0}-L\lambda\bar{\mu}_1P_{L-1,0}}{\mu_1-\lambda}+\frac{L\lambda\bar{\mu}_1P_{L-1,0}}{\mu_2-\lambda}=1. \tag{14}$$

Based on Equation (14) the first relationship between $P_{0,0}$ and $P_{L-1,0}$ is given by

$$P_{L-1,0}=\frac{(\mu_2-\lambda)\left[\lambda-\mu_1\left(1-P_{0,0}\right)\right]}{L\lambda\bar{\mu}_1(\mu_2-\mu_1)}. \tag{15}$$

On the other hand, with the help of Equations (2)-(4), another relationship between $P_{0,0}$ and $P_{L-1,0}$ can be obtained by a backward recursion procedure.

$$P_{L-1,0}=\frac{\left(\lambda\bar{\mu}_1\right)^{L-1}\bar{\lambda}\mu_1-\left(\lambda\bar{\mu}_1\right)^{L}}{\bar{\mu}_1\left[\left(\bar{\lambda}\mu_1\right)^{L}-\left(\lambda\bar{\mu}_1\right)^{L}\right]}P_{0,0}. \tag{16}$$

Substituting Equation (16) into Equation (15), it follows that:

$$\begin{aligned}P_{0,0}=\frac{\left(\mu_2-\lambda\right)\left(\lambda-\mu_1\right)\left[\left(\bar{\lambda}\mu_1\right)^{L}-\left(\lambda\bar{\mu}_1\right)^{L}\right]\bar{\mu}_1}{\left\{\mu_1\left[\left(L-1\right)\left(\lambda\bar{\mu}_1\right)^{L+1}-L\left(\lambda\bar{\mu}_1\right)^{L}\bar{\lambda}\mu_1+\lambda\bar{\mu}_1\left(\bar{\lambda}\mu_1\right)^{L}\right]\right.}\\ \left.-\mu_2\left[L\left(\lambda\bar{\mu}_1\right)^{L+1}-L\left(\lambda\bar{\mu}_1\right)^{L}\bar{\lambda}\mu_1-\bar{\mu}_1\mu_1\left(\lambda\bar{\mu}_1\right)^{L}+\bar{\mu}_1\mu_1\left(\bar{\lambda}\mu_1\right)^{L}\right]\right\}}.\end{aligned} \tag{17}$$

Remark 4. Obviously, the queueing system for $\mu_1=\mu_2$ coincides with the classic Geo/Geo/1 queue studied by Hunter [16]. Additionally, under such an assumption, after some brief algebraic manipulations, the Equation (17) can further be simplified as follows: $P_{0,0}=1-\rho$, where $\rho=\dfrac{\lambda}{\mu_1}$. The formula derived here agrees with the one given by Hunter [16], and it also shows the correctness of our analysis presented above.

Having computed the stationary probabilities $P_{0,0}$ and $P_{L-1,0}$, a recursive algorithm for computing the other steady state probabilities can be established. To demonstrate the working schemes of the recursive method, we describe the solution algorithm in the following **Table 1**.

3.3. Explicit Expression for the Expected Number of Customers in the System

Once the explicit expressions for $P_{0,0}$ and $P_{L-1,0}$ are given, the expected number of customers in the system can be determined from them. Let *N* be the number of customers in the system in steady state. We have

Table 1. Comutation of the stationary distribution $P_{i,j}$.

Begin algorithm

Input: $\left\{\lambda,\mu_1,\mu_2,L,K\left(\text{the base case stops the recursion}\right)\right\}$.

Output: $P_{i,0}\left(i=1,\cdots,L-2\right)$ and $P_{i,1}\left(i=1,2,\cdots\right)$.

Calculate $P_{0,0}$, $P_{L-1,0}$ and $P_{L-2,0}$ using Equations (17), (16) and (4).

for $i=3:1:L-1$ do

$$P_{L-i,0}=\frac{\left(1-\bar{\lambda}\bar{\mu}_1-\lambda\mu_1\right)P_{L-i+1,0}-\bar{\lambda}\mu_1P_{L-i+2,0}}{\lambda\bar{\mu}_1},$$

end

Calculate $P_{1,1}$ and $P_{2,1}$ using Equations (1) and (5).

for $i=2:1:L-1$ do

$$P_{i+1,1}=\frac{\left(1-\bar{\lambda}\bar{\mu}_2-\lambda\mu_2\right)P_{i,1}-\lambda\bar{\mu}_2P_{i-1,1}}{\bar{\lambda}\mu_2},$$

end

Calculate $P_{L+1,1}$ using Equation (7).

for $i=L+1:1:K$ do

$$P_{i+1,1}=\frac{\left(1-\bar{\lambda}\bar{\mu}_2-\lambda\mu_2\right)P_{i,1}-\lambda\bar{\mu}_2P_{i-1,1}}{\bar{\lambda}\mu_2},$$

end

End algorithm

$$E[N]=\left.\frac{\mathrm{d}\tilde{P}(z)}{\mathrm{d}z}\right|_{z=1}$$
$$=\left.\frac{\lambda P_{0,0}-\lambda\bar{\mu}_1 P_{L-1,0}\dfrac{Lz^{L+1}-(L+1)z^L+1}{(1-z)^2}}{\bar{\lambda}\mu_1-z\lambda\bar{\mu}_1}\right|_{z=1}+\left.\frac{\lambda\bar{\mu}_1\left(\lambda P_{0,0}z-\lambda\bar{\mu}_1 P_{L-1,0}\sum_{k=1}^{L}z^k\right)}{\left(\bar{\lambda}\mu_1-z\lambda\bar{\mu}_1\right)^2}\right|_{z=1}$$
$$+\left.\frac{\lambda\bar{\mu}_1 P_{L-1,0}\dfrac{Lz^{L+1}-(L+1)z^L+1}{(1-z)^2}}{\bar{\lambda}\mu_2-z\lambda\bar{\mu}_2}\right|_{z=1}+\left.\frac{\lambda^2\bar{\mu}_1\bar{\mu}_2 P_{L-1,0}\sum_{k=1}^{L}z^k}{\left(\bar{\lambda}\mu_2-z\lambda\bar{\mu}_2\right)^2}\right|_{z=1}.$$

Using L'Hospital's Rule twice while taking limits $z\to 1$, we get

$$E[N]=\frac{\lambda P_{0,0}-\lambda\bar{\mu}_1 P_{L-1,0}\dfrac{L(L+1)}{2}}{\mu_1-\lambda}+\frac{\lambda^2\bar{\mu}_1 P_{0,0}-L(\lambda\bar{\mu}_1)^2 P_{L-1,0}}{(\mu_1-\lambda)^2}$$
$$+\frac{\lambda\bar{\mu}_1 P_{L-1,0}\dfrac{L(L+1)}{2}}{\mu_2-\lambda}+\frac{L\lambda^2\bar{\mu}_1\bar{\mu}_2 P_{L-1,0}}{(\mu_2-\lambda)^2}. \tag{18}$$

Remark 5. As a matter of fact, the explicit expression for the expected number of customers in the system has been given by Equation (18). Just because the explicit expressions $P_{0,0}$ and $P_{L-1,0}$ are slightly cumbersome to write, we do not indent to substitute Equations (16) and (17) into Equation (18).

3.4. Sojourn Time Performance

In this subsection, we deal with the customer's sojourn time W, defined as the time between the arrival epoch of a customer till the instant at which his service request is satisfied. Here, our aim is to determine the first order moment of the sojourn time. To achieve this goal, we need to introduce some auxiliary random variables.

$W_{i,j}^r$: Customer's sojourn time given that he finds the queueing system at state (i,j) just before his arrival and the residual service time of customer that the server is currently processing is greater than or equal to one time slot. $(i=1,2,\cdots;\ j=0,1)$

$W_{i,j}$: Conditional sojourn time of a customer who arrives at the system when its state is (i,j). $(i=0,1,2,\cdots;\ j=0,1)$

We also denote the corresponding z-transforms of W, $W_{i,j}^r$ and $W_{i,j}$ by $\tilde{W}(z)$, $\tilde{W}_{i,j}^r(z)$ and $\tilde{W}_{i,j}(z)$ respectively. Furthermore, because of the BASTA (*i.e.* Bernoulli arrivals see time averages) property, we have that

$$\tilde{W}(z)=E\left[z^W\right]=\sum_{i=0}^{L-1}P_{i,0}\tilde{W}_{i,0}(z)+\sum_{i=1}^{\infty}P_{i,1}\tilde{W}_{i,1}(z)$$
$$=\sum_{i=0}^{L-1}P_{i,0}E\left[z^{W_{i,0}}\right]+\sum_{i=1}^{\infty}P_{i,1}E\left[z^{W_{i,1}}\right]. \tag{19}$$

By differentiating Equation (19) with respect to z, and evaluating at $z=1$, we arrive at

$$E[W]=\sum_{i=0}^{L-1}P_{i,0}E\left[W_{i,0}\right]+\sum_{i=1}^{\infty}P_{i,1}E\left[W_{i,1}\right]. \tag{20}$$

For determining the unknowns $E\left[W_{i,0}\right]$ and $E\left[W_{i,1}\right]$, we apply a first-step argument and set up the following equations.

$$\begin{aligned}\tilde{W}_{0,0}(z) &= E\left[z^{W_{0,0}} \middle| S_{low} \le \sum_{j=1}^{L-1} A_j\right] \Pr\left\{S_{\text{low}} \le \sum_{j=1}^{L-1} A_j\right\} + E\left[z^{W_{0,0}} \middle| S_{\text{low}} > \sum_{j=1}^{L-1} A_j\right] \Pr\left\{S_{\text{low}} > \sum_{j=1}^{L-1} A_j\right\} \\ &= \sum_{k=1}^{\infty} \sum_{n=\max(k,L-1)}^{\infty} z^k \Pr\{S_{\text{low}} = k\} \Pr\left\{\sum_{j=1}^{L-1} A_j = n\right\} + \sum_{k=L}^{\infty} \sum_{n=L-1}^{k-1} z^n E\left[z^{S_{\text{high}}}\right] \Pr\{S_{\text{low}} = k\} \Pr\left\{\sum_{j=1}^{L-1} A_j = n\right\} \\ &= \sum_{k=1}^{\infty} \sum_{n=\max(k,L-1)}^{\infty} z^k \mu_1 \bar{\mu}_1^{k-1} \binom{n-1}{L-2} \lambda^{L-1} (\bar{\lambda})^{n-L+1} + \sum_{k=L}^{\infty} \sum_{n=L-1}^{k-1} z^n E\left[z^{S_{\text{high}}}\right] \mu_1 \bar{\mu}_1^{k-1} \binom{n-1}{L-2} \lambda^{L-1} (\bar{\lambda})^{n-L+1} \\ &= \frac{\mu_1 z}{1-\bar{\mu}_1 z}\left[1-\left(\frac{\lambda\bar{\mu}_1 z}{1-\bar{\lambda}\bar{\mu}_1 z}\right)^{L-1}\right] + \left(\frac{\lambda\bar{\mu}_1 z}{1-\bar{\lambda}\bar{\mu}_1 z}\right)^{L-1} \frac{\mu_2 z}{1-\bar{\mu}_2 z}.\end{aligned} \tag{21}$$

Assume that a customer arrival will occur in (t^-, t). If prior to this arrival there are $i\,(1 \le i \le L-2)$ customers in the system and the server is busy with low service rate, then the departure of the customer that the server is currently processing will take place in (t, t^+) with probability μ_1, thus $\bar{\mu}_1$ is the probability that the above event does not occur. Hence we can easily get the following relationships

$$\tilde{W}_{1,0}(z) = \mu_1 \tilde{W}_{0,0}(z) + \bar{\mu}_1 \tilde{W}_{1,0}^r(z), \tag{22}$$

$$\tilde{W}_{i,0}(z) = \mu_1 \tilde{W}_{i-1,0}^r(z) + \bar{\mu}_1 \tilde{W}_{i,0}^r(z), \quad i = 2,3,\cdots,L-2, \tag{23}$$

$$\tilde{W}_{L-1,0}(z) = \mu_1 \tilde{W}_{L-2,0}^r(z) + \bar{\mu}_1 \left(\frac{\mu_2 z}{1-\bar{\mu}_2 z}\right)^L. \tag{24}$$

For the same reason as mentioned above, when a customer arrives at the system during a busy period with high service rate, we conclude that $\tilde{W}_{i,1}(z)$ satisfies

$$\tilde{W}_{i,1}(z) = \mu_2 \left(\frac{\mu_2 z}{1-\bar{\mu}_2 z}\right)^i + \bar{\mu}_2 \left(\frac{\mu_2 z}{1-\bar{\mu}_2 z}\right)^{i+1}, \quad i = 1,2,\cdots. \tag{25}$$

Alternatively, we can use the memoryless property of the geometric distribution to find the z-transform of $W_{i,j}^r$. For $j = 1$,

$$\tilde{W}_{i,1}^r(z) = \left(\frac{\mu_2 z}{1-\bar{\mu}_2 z}\right)^{i+1}, \quad j = 1,2,\cdots. \tag{26}$$

For $j = 0$, we have

$$\begin{aligned}\tilde{W}_{1,0}^r(z) &= E\left[z^{W_{1,0}^r} \middle| S_{\text{low}} \le \sum_{j=1}^{L-2} A_j\right] \Pr\left\{S_{\text{low}} \le \sum_{j=1}^{L-2} A_j\right\} \\ &\quad + E\left[z^{W_{1,0}^r} \middle| S_{\text{low}} > \sum_{j=1}^{L-2} A_j\right] \Pr\left\{S_{\text{low}} > \sum_{j=1}^{L-2} A_j\right\} \\ &= \sum_{k=1}^{\infty} \sum_{n=\max(k,L-2)}^{\infty} z^k E\left[z^{W_{0,0}}\right] \Pr\{S_{\text{low}} = k\} \Pr\left\{\sum_{j=1}^{L-2} A_j = n\right\} \\ &\quad + \sum_{k=L-1}^{\infty} \sum_{n=L-2}^{k-1} z^n E\left[z^{W_{1,1}^r}\right] \Pr\{S_{\text{low}} = k\} \Pr\left\{\sum_{j=1}^{L-2} A_j = n\right\} \\ &= \sum_{k=1}^{\infty} \sum_{n=\max(k,L-2)}^{\infty} z^k E\left[z^{W_{0,0}}\right] \mu_1 \bar{\mu}_1^{k-1} \binom{n-1}{L-3} \lambda^{L-2} (\bar{\lambda})^{n-L+2} \\ &\quad + \sum_{k=L-1}^{\infty} \sum_{n=L-2}^{k-1} z^n E\left[z^{W_{1,1}^r}\right] \mu_1 \bar{\mu}_1^{k-1} \binom{n-1}{L-3} \lambda^{L-2} (\bar{\lambda})^{n-L+2} \\ &= \frac{\mu_1 z}{1-\bar{\mu}_1 z}\left[1-\left(\frac{\lambda\bar{\mu}_1 z}{1-\bar{\lambda}\bar{\mu}_1 z}\right)^{L-2}\right] \tilde{W}_{0,0}^r(z) + \left(\frac{\lambda\bar{\mu}_1 z}{1-\bar{\lambda}\bar{\mu}_1 z}\right)^{L-2} \tilde{W}_{1,1}^r(z).\end{aligned} \tag{27}$$

Similarly, for $i=2,3,\cdots,L-2$ and $j=0$, we have

$$\begin{aligned}\tilde{W}_{i,0}^{r}(z)&=E\left[z^{W_{i,0}^{r}}\middle|S_{\text{low}}\le\sum_{j=1}^{L-i-1}A_j\right]\Pr\left\{S_{\text{low}}\le\sum_{j=1}^{L-i-1}A_j\right\}\\&\quad+E\left[z^{W_{i,0}^{r}}\middle|S_{\text{low}}>\sum_{j=1}^{L-i-1}A_j\right]\Pr\left\{S_{\text{low}}>\sum_{j=1}^{L-i-1}A_j\right\}\\&=\sum_{k=1}^{\infty}\sum_{n=\max(k,L-i-1)}^{\infty}z^{k}E\left[z^{W_{i-1,0}^{r}}\right]\Pr\{S_{\text{low}}=k\}\Pr\left\{\sum_{j=1}^{L-i-1}A_j=n\right\}\\&\quad+\sum_{k=L-i}^{\infty}\sum_{n=L-i-1}^{k-1}z^{n}E\left[z^{W_{i,1}^{r}}\right]\Pr\{S_{\text{low}}=k\}\Pr\left\{\sum_{j=1}^{L-i-1}A_j=n\right\}\\&=\sum_{k=1}^{\infty}\sum_{n=\max(k,L-i-1)}^{\infty}z^{k}E\left[z^{W_{i-1,0}^{r}}\right]\mu_1\bar{\mu}_1^{k-1}\binom{n-1}{L-i-2}\lambda^{L-i-1}(\bar{\lambda})^{n-L+i+1}\\&\quad+\sum_{k=L-1}^{\infty}\sum_{n=L-2}^{k-1}z^{n}E\left[z^{W_{i,1}^{r}}\right]\mu_1\bar{\mu}_1^{k-1}\binom{n-1}{L-i-2}\lambda^{L-i-1}(\bar{\lambda})^{n-L+i+1}\\&=\frac{\mu_1 z}{1-\bar{\mu}_1 z}\left[1-\left(\frac{\lambda\bar{\mu}_1 z}{1-\bar{\lambda}\bar{\mu}_1 z}\right)^{L-i-1}\right]\tilde{W}_{i-1,0}^{r}(z)+\left(\frac{\lambda\bar{\mu}_1 z}{1-\bar{\lambda}\bar{\mu}_1 z}\right)^{L-i-1}\tilde{W}_{i,1}^{r}(z).\end{aligned}\tag{28}$$

Differentiating both sides of Equation (21) and Equations (25)-(28) with respect to z and evaluating at $z=1$, we can obtain the following equations for the first moment of the conditional sojourn time.

$$E\left[W_{0,0}\right]=\frac{1}{\mu_1}\left[1-\left(\frac{\lambda\bar{\mu}_1}{1-\bar{\lambda}\bar{\mu}_1}\right)^{L-1}\right]+\left(\frac{\lambda\bar{\mu}_1}{1-\bar{\lambda}\bar{\mu}_1}\right)^{L-1}\frac{1}{\mu_2},\tag{29}$$

$$E\left[W_{i,1}\right]=\frac{i+\bar{\mu}_2}{\mu_2},\quad i=1,2,\cdots,\tag{30}$$

$$E\left[W_{i,1}^{r}\right]=\frac{i+1}{\mu_2},\quad i=1,2,\cdots,\tag{31}$$

$$E\left[W_{1,0}^{r}\right]=\left(\frac{1}{\mu_1}+E\left[W_{0,0}\right]\right)\left[1-\left(\frac{\lambda\bar{\mu}_1}{1-\bar{\lambda}\bar{\mu}_1}\right)^{L-2}\right]+\left(\frac{\lambda\bar{\mu}_1}{1-\bar{\lambda}\bar{\mu}_1}\right)^{L-2}E\left[W_{1,1}^{r}\right],\tag{32}$$

$$E\left[W_{i,0}^{r}\right]=\left(\frac{1}{\mu_1}+E\left[W_{i-1,0}^{r}\right]\right)\left[1-\left(\frac{\lambda\bar{\mu}_1}{1-\bar{\lambda}\bar{\mu}_1}\right)^{L-i-1}\right]+\left(\frac{\lambda\bar{\mu}_1}{1-\bar{\lambda}\bar{\mu}_1}\right)^{L-i-1}E\left[W_{i,1}^{r}\right],\ i=2,\cdots,L-2.\tag{33}$$

Therefore, from the above results and Equations (22)-(24), we obtain

$$E\left[W_{1,0}\right]=\mu_1 E\left[W_{0,0}\right]+\bar{\mu}_1 E\left[W_{1,0}^{r}\right],\tag{34}$$

$$E\left[W_{i,0}\right]=\mu_1 E\left[W_{i-1,0}^{r}\right]+\bar{\mu}_1 E\left[W_{i,0}^{r}\right],\ i=2,\cdots,L-2,\tag{35}$$

$$E\left[W_{L-1,0}\right]=\mu_1 E\left[W_{L-2,0}^{r}\right]+\bar{\mu}_1\frac{L}{\mu_2}.\tag{36}$$

Thus, the problem of computing the mean conditional sojourn times $E\left[W_{i,0}\right]$ and $E\left[W_{i,1}\right]$ can be considered solved. Consequently, with the help of stationary probability $P_{i,j}$, we can evaluate the expectation of the unconditional sojourn time by using Equation (20).

To demonstrate the feasibility and efficiency of the proposed algorithm, a numerical experiment is carried out on a personal computer implementing an Intel Core i5 CPU (2.7 GHz) and 4.0 GB RAM. In this example, we

select $\mu_1 = 0.18$, $\mu_2 = 0.24$, $L = 6$ and let λ vary from 0.1 to 0.16. **Figure 3** illustrates the effect of customer's arrival rate on the mean value of the unconditional sojourn time. Also, on putting $\mu_1 = \mu_2 = 0.18$, the queueing system under consideration can be regarded as the classic Geo/Geo/1 queue with constant service rate. From **Figure 3** we can conclude that setting the switching threshold for the service rate can greatly reduce the customer's average sojourn time, for example, when the customer arrival rate is 0.16, the gap between the two average sojourn times is about 25 time units.

3.5. Regeneration Cycle

Regeneration cycles are models of stochastic phenomena in which an event (or combination of events) occurs repeatedly over time, and the times between occurrences are independent and identically distributed. Models of such phenomena typically focus on determining limiting averages for costs or other system parameters. In this paper, the reason for performing regeneration cycle analysis is to determine the optimal switching threshold value *L*, where the high service rate is activated.

A regeneration cycle of our current model consists of a server's idle period and a server's busy period. As regeneration points, we choose the points at which the system becomes empty. There are two types of cycles depending on whether there is a change in service rate during the server's busy period. A cycle is called "type-1" if it does not include switching of the service rate; otherwise it is of "type-2" cycle. To better understand the structure of regeneration cycle, examples of the type-1 and type-2 cycles are shown in **Figure 4** and **Figure 5**, respectively.

We denote $\theta(z)$ as the probability generating function of the busy period for classical Geo/Geo/1 queue. If customer arrival occurs according to a Bernoulli process with parameter λ, and the service times provided by a

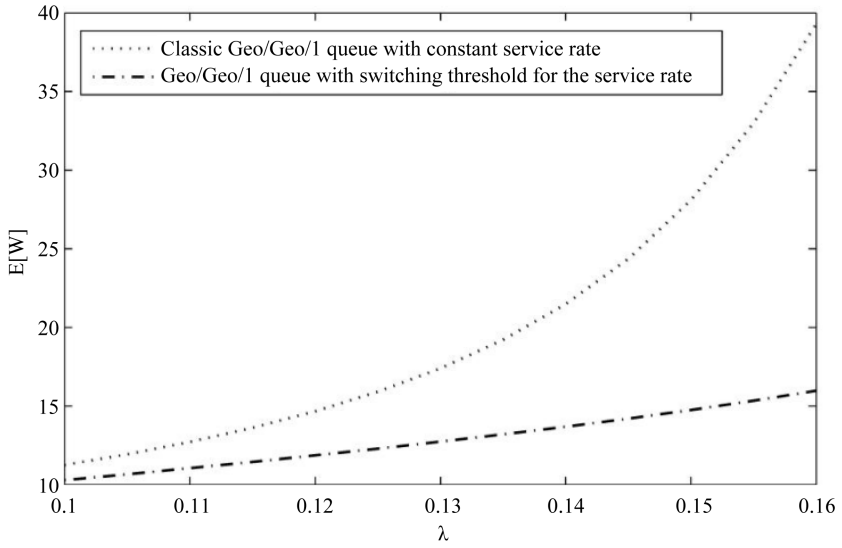

Figure 3. The effect of λ on the mean value of the unconditional sojourn time.

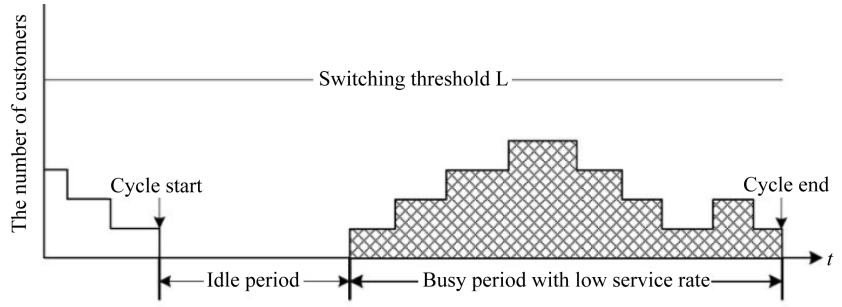

Figure 4. An example of the type-1 cycle.

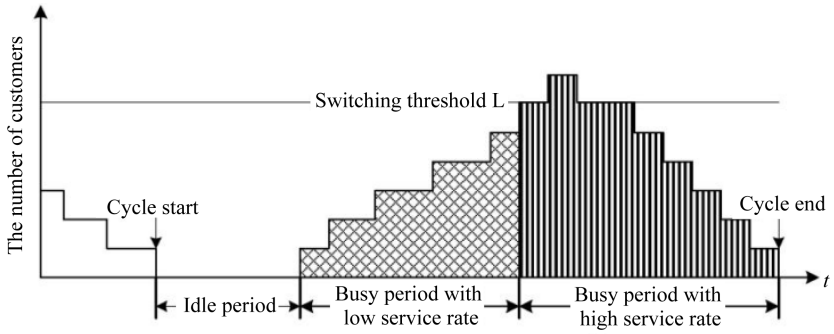

Figure 5. An example of the type-2 cycle.

single server follow geometric distribution with parameter μ_2, then from Takagi [17], we have

$$\theta(z)=\frac{\mu_2 z\left(\lambda\theta(z)+\bar{\lambda}\right)}{1-\bar{\mu}_2 z\left(\lambda\theta(z)+\bar{\lambda}\right)},$$

$$E\left[B_{\text{Geo/Geo/1}}\right]=\frac{1}{\mu_2-\lambda},$$

where $E\left[B_{\text{Geo/Geo/1}}\right]$ is the mean value of the busy period for classic Geo/Geo/1 queue. Furthermore, let I, B_{low} and B_{high} respectively denote the length of server's idle period and the length of busy periods with low and high service rates in a regeneration cycle. It is obvious that I follows geometric distribution, thus $E[I]=1/\lambda$. Next, we derive the probability generating function of busy period with high service rate $\tilde{B}_{\text{high}}(z)$. According to the model assumptions, the busy period with high service rate is only activated by L customers waiting in the queue (including the one in service). By conditioning on the duration of the remaining service time for the customer currently being served, we get

$$\begin{aligned}\tilde{B}_{\text{high}}(z)&=P_{\text{switch}}\sum_{j=1}^{\infty}\mu_2\bar{\mu}_2^{j-1}z^j\sum_{r=0}^{j}\binom{j}{r}\lambda^r\left(\bar{\lambda}\right)^{j-r}\left(\theta(z)\right)^{r+L-1}\\&=P_{\text{switch}}\sum_{j=1}^{\infty}\mu_2\bar{\mu}_2^{j-1}z^j\left(\theta(z)\right)^{L-1}\left(\lambda\theta(z)+\bar{\lambda}\right)^j\\&=P_{\text{switch}}\left(\frac{\mu_2 z\left(\lambda\theta(z)+\bar{\lambda}\right)}{1-\bar{\mu}_2 z\left(\lambda\theta(z)+\bar{\lambda}\right)}\right)^L,\end{aligned}\tag{37}$$

where P_{swtich} denotes the probability that the service rate does switch in a regeneration cycle. Thus, from Equation (37), we have

$$E\left[B_{\text{high}}\right]=\frac{\mathrm{d}}{\mathrm{d}z}\tilde{B}_{\text{high}}(z)\Big|_{z=1}=\frac{P_{\text{switch}}L\left(1+\lambda E\left[B_{\text{Geo/Geo/1}}\right]\right)}{\mu_2}=\frac{P_{\text{switch}}L}{\mu_2-\lambda}.\tag{38}$$

On the other hand, let $E[C]$ be the unconditional expected length of the regeneration cycle, the mean duration of busy period with high service rate can also be obtained from a result of renewal theory. Using

$$\frac{E[I]}{E[C]}=P_{0,0},$$

$$\frac{E\left[B_{\text{low}}\right]}{E[C]}=\tilde{P}_0(z)\Big|_{z=1}=\frac{\lambda-\dfrac{L\lambda\left[\left(\lambda\bar{\mu}_1\right)^{L-1}\bar{\lambda}\mu_1-\left(\lambda\bar{\mu}_1\right)^L\right]}{\left[\left(\bar{\lambda}\mu_1\right)^L-\left(\lambda\bar{\mu}_1\right)^L\right]}}{\mu_1-\lambda}P_{0,0},$$

$$\frac{E\left[B_{\text{high}}\right]}{E[C]}=\tilde{P}_1(z)\big|_{z=1}=\frac{\dfrac{L\lambda\left[\left(\lambda\bar{\mu}_1\right)^{L-1}\bar{\lambda}\mu_1-\left(\lambda\bar{\mu}_1\right)^{L}\right]}{\left[\left(\bar{\lambda}\mu_1\right)^{L}-\left(\lambda\bar{\mu}_1\right)^{L}\right]}}{\mu_2-\lambda}P_{0,0},$$

we can get

$$E[C]=\frac{1}{\lambda P_{0,0}}, \tag{39}$$

$$E\left[B_{\text{low}}\right]=\frac{1-\dfrac{L\left[\left(\lambda\bar{\mu}_1\right)^{L-1}\bar{\lambda}\mu_1-\left(\lambda\bar{\mu}_1\right)^{L}\right]}{\left[\left(\bar{\lambda}\mu_1\right)^{L}-\left(\lambda\bar{\mu}_1\right)^{L}\right]}}{\mu_1-\lambda}, \tag{40}$$

$$E\left[B_{\text{high}}\right]=\frac{\dfrac{L\left[\left(\lambda\bar{\mu}_1\right)^{L-1}\bar{\lambda}\mu_1-\left(\lambda\bar{\mu}_1\right)^{L}\right]}{\left[\left(\bar{\lambda}\mu_1\right)^{L}-\left(\lambda\bar{\mu}_1\right)^{L}\right]}}{\mu_2-\lambda}. \tag{41}$$

Comparing the right hand sides of Equations (38) and (41), we see that

$$P_{\text{switch}}=\frac{\left[\left(\lambda\bar{\mu}_1\right)^{L-1}\bar{\lambda}\mu_1-\left(\lambda\bar{\mu}_1\right)^{L}\right]}{\left[\left(\bar{\lambda}\mu_1\right)^{L}-\left(\lambda\bar{\mu}_1\right)^{L}\right]}.$$

Once we have found the expressions of $E[C]$, $E\left[B_{\text{low}}\right]$, $E\left[B_{\text{high}}\right]$, P_{switch} and $E[N]$, we can try to construct the cost structure of this queueing system in the next section.

4. Optimal Switching Threshold for the Service Rate and Numerical Examples

In manufacturing process management, managers are always interested in minimizing the long-run average cost per unit time of the system. In this section, based on the performance measures that we obtained in the previous section and the renewal reward theorem, we first construct an expected cost rate function $TC(L)$ for the Geo/Geo/1 queue with switching threshold for the service rate, in which a key decision variable L is considered. Here, our objective is to determine the optimal threshold value L^* under some cost structure, so as to minimize the long-run average cost rate.

Let us consider the following cost elements:

$C_s \equiv$ setup cost per cycle;

$C_{\text{switch}} \equiv$ switching cost for changing the service rate in a regeneration cycle;

$C_h \equiv$ holding cost per customer per unit time;

$C_{\text{low}} \equiv$ running cost per unit time when the service provides low speed service;

$C_{\text{high}} \equiv$ running cost per unit time when the service provides high speed service.

Utilizing the definition of each cost element listed above, the long-run average cost rate minimization problem can be illustrated mathematically as

$$\min_{L} TC(L)=\frac{C_s+C_{\text{switch}}P_{\text{switch}}+C_{\text{low}}E\left[B_{\text{low}}\right]+C_{\text{high}}E\left[B_{\text{high}}\right]}{E[C]}+C_hE[N].$$

As shown in **Figure 4** and **Figure 5**, the switching cost is incurred at most only once in a regeneration cycle, and the switching occurs with probability P_{switch}. This is the reason why we multiply C_{switch} by P_{switch} in our cost structure. On the other hand, we also note that it is rather difficult to develop analytic results for the optimal value of L because the long-run average cost rate function is highly non-linear and complex. In spite of that, since L is a discrete variable, the optimal value L^* may be found by using direct substitution of successive val-

ues of L into the long-run average cost rate function until the minimum value $TC(L)$ is achieved.

To illustrate the direct search algorithm described above, a numerical example is provided by considering the following cost parameters:

$$C_s = \$270/\text{time},\ C_{\text{switch}} = \$150/\text{time},\ C_h = \$4/\text{customer/unit time}$$
$$C_{\text{low}} = \$7/\text{unit time},\ C_{\text{high}} = \$12/\text{unit time},$$

and other system parameters are taken as $\lambda = 0.14$, $\mu_1 = 0.18$, $\mu_2 = 0.24$. Substituting these values into $TC(L)$, we can obtain the results presented in **Table 2** and **Figure 6**. The curve representing the long-run average cost rate function $TC(L)$ is plotted in **Figure 6** for different values of L. As can be seen in **Figure 6**, we observe that this function is convex and a single relative minimum exists. The optimal value L^* and the corresponding long-run average cost rate $TC(L^*)$ are tabulated in **Table 2**. From **Table 2**, it appears that the minimum average cost per unit time of 24.2347 is obtained with $L^* = 7$.

5. Conclusion

In this paper, we have carried out an analysis of a discrete-time infinite-buffer Geo/Geo/1 queuing system under

Table 2. The long-run average cost rate against the values of L.

L	$TC(L)$	L	$TC(L)$	L	$TC(L)$
2	29.1306	11	24.6055	20	25.5883
3	26.5196	12	24.7529	21	25.6414
4	25.2645	13	24.8970	22	25.6858
5	24.6309	14	25.0325	23	25.7228
6	24.3345	15	25.1567	24	25.7535
7	24.2347	16	25.2680	25	25.7787
8	24.2522	17	25.3663	26	25.7993
9	24.3386	18	25.4519	27	25.8162
10	24.4630	19	25.5256	28	25.8298

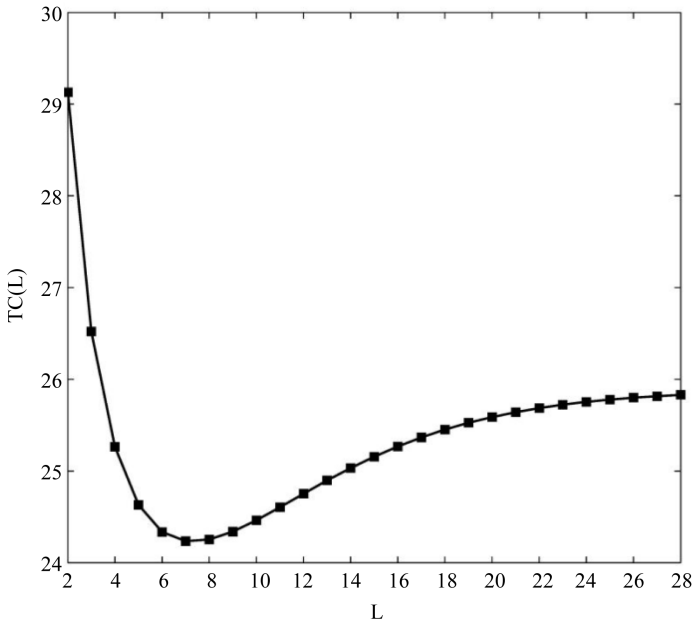

Figure 6. The plot of $TC(L)$ against the switching threshold L.

a modified service rate switching policy that has potential applications in modeling manufacturing and telecommunication systems. We have developed a recursive method to find the steady-state queue size distribution. The recursive method is powerful and easy to implement. Further, we obtain the analytically explicit expressions for the expected number of customers in the system. Using the first-step argument, a simple algorithm for calculating the customer's mean sojourn time has been proposed. Moreover, we also performed regeneration cycle analysis of the queue to find the optimal service rate switching threshold L. Our current model is useful and significant to engineers or managers who design an efficient system with economic management. It should be pointed out that the economic importance of this model resides in the multiple applications to manufacturing processes, since most of them operate on a discrete time basis. Furthermore, the optimal control of service rate switching policy is also a main objective from the enterprise point of view. For future studies, the present investigation can be extended by incorporating bulk input or bulk service. Another area of interest may be expanding our model into Geo/G/1 type, because there will be a significant improvement inapplicability to real world system.

Acknowledgements

The work described in this paper is supported by Sichuan Provincial Department of Education (14ZB0221).

References

[1] Bekker, R., Borst, S., Boxma, O. and Kella, O. (2004) Queues with Workload-Dependent Arrival and Service Rates. *Queueing Systems*, **46**, 537-556. http://dx.doi.org/10.1023/B:QUES.0000027998.95375.ee

[2] Chaudhry, M.L. and Gupta, U.C. (1996) On the Analysis of the Discrete-Time Geo$m(n)$/$G(n)$/1/N Queue. *Probability in the Engineering and Informational Sciences*, **10**, 415-428. http://dx.doi.org/10.1017/S0269964800004447

[3] Chaudhry, M.L., Templeton, J.G.C. and Gupta, U.C. (1996) Analysis of the Discrete-Time $GI(n)$/Geo$m(n)$/1/N Queue. *Computers & Mathematics with Applications*, **31**, 59-68. http://dx.doi.org/10.1016/0898-1221(95)00182-X

[4] Garg, R.L. and Singh, P. (1993) Queue-Dependent servers Queueing System. *Microelectronics Reliability*, **33**, 2289-2295. http://dx.doi.org/10.1016/0026-2714(93)90072-7

[5] Gebhard, R.F. (1967) A Queueing Process with Bilevel Hysteretic Service-Rate Control. *Naval Research Logistics Quarterly*, **14**, 55-67. http://dx.doi.org/10.1002/nav.3800140106

[6] Gross, D. and Harris, C.M. (1985) Fundamentals of Queueing Theory. 2nd Edition, John Wiley, New York.

[7] Harris, C.M. and Marchal, W.G. (1988) State Dependence in M/G/1 Server-Vacation Models. *Operations Research*, **36**, 560-565. http://dx.doi.org/10.1287/opre.36.4.560

[8] Hunter, J.J. (1983) Mathematical Techniques of Applied Probability, Discrete-Time Models: Techniques and Applications. Vol. II, Academic Press, New York.

[9] Jain, M. (2005) Finite Capacity M/M/r Queueing System with Queue Dependent Servers. *Computers & Mathematics with Applications*, **50**, 187-199. http://dx.doi.org/10.1016/j.camwa.2004.11.018

[10] Lin, C.H. and Ke, J.C. (2011) Optimization Analysis for an Infinite Capacity Queueing System with Multiple Queue-Dependent Servers: Genetic Algorithm. *International Journal of Computer Mathematics*, **88**, 1430-1442. http://dx.doi.org/10.1080/00207160.2010.509791

[11] Parthasarathy, P.R. and Lenin, R.B. (1999) Exact Busy Period Distribution of a Discrete Queue with Quadratic Rates. *International Journal of Computer Mathematics*, **71**, 427-436. http://dx.doi.org/10.1080/00207169908804819

[12] Saaty, T.L. (1961) Elementary of Queueing Theory with Applications. McGraw-Hill, New York.

[13] Singh, V.P. (1973) Queue-Dependent Servers. *Journal of Engineering Mathematics*, **7**, 123-126. http://dx.doi.org/10.1007/BF01535357

[14] Takagi, H. (1993) Queueing Analysis: A Foundation of Performance Evaluation. Vol. 3, North-Holland, New York.

[15] Wang, K.H. and Tai, K.Y. (2000) A Queueing System with Queue-Dependent Servers and Finite Capacity. *Applied Mathematical Modelling*, **24**, 807-814. http://dx.doi.org/10.1016/S0307-904X(00)00013-5

[16] William, J.G. and Wang, P. (1992) An M/G/1-Type Queueing Model with Service Times Depending on Queue Length. *Applied Mathematical Modelling*, **16**, 652-658. http://dx.doi.org/10.1016/0307-904X(92)90098-N

[17] Zhernovyi, Y.V. (2012) Stationary Characteristics of M^X/M/1 Systems with Two-Speed Service. *Journal of Communications Technology and Electronics*, **57**, 920-931. http://dx.doi.org/10.1134/S1064226912080074