CHAPTER-2

PERFORMANCE ANALYSIS OF AN $M/M/1/N$ QUEUE WITH BALKING AND RENEGING

2.1 INTRODUCTION

Many practical queueing systems especially those with balking and reneging have been widely applied to many real-life problems, such as the situations involving impatient telephone switchboard customers, the hospital emergency rooms handling critical patients, and the inventory systems with storage of perishable goods [77]. In this chapter, we consider an $M/M/1/N$ queueing system with balking and reneging.

Queueing systems with balking, reneging or both have been studied by many researchers. Haight [39] first considered an $M/M/1$ queue with balking. An $M/M/1$ queue with customers reneging was also proposed by Haight [40]. The combined effects of balking and reneging in an $M/M/1/N$ queue have been investigated by Ancker and Gafarian [3], [4]. Abou-EI-Ata and Hariri [1] considered the multiple servers queueing system $M/M/c/N$ with balking and reneging.

The rest of this chapter is organized as follows. In the next part of this chapter, we give a description of the queueing model. In part 2.3 of this chapter, we derive the steady-state equations by the Markov process method. By writing the transition rate matrix as block matrix, we get the matrix form solution of the steady-state probabilities and present a procedure for calculating the steady-state probabilities. In part 2.4 of this chapter, we give some performance measures of the system. Based on the performance analysis, we formulate a cost model to determine the optimal service rate. Some numerical examples are presented to demonstrate how the various parameters of the model influence the behavior of the system. Conclusions are given in last of this chapter.
2.2 SYSTEM MODEL

In this chapter, we consider an $M/M/1/N$ queueing system with balking and reneging. The assumptions of the system model are as follows:

1. Customers arrive at the system one by one according to a Poisson process with rate $\lambda$. On arrival a customer either decides to join the queue with probability $b_n$ or balk with probability $1-b_n$ when $n$ customers are ahead of him ($n = 0, 1, \ldots, N-1$), where $N$ is the maximum numbers of the customers in the system, and

$$0 \leq b_{n-1} \leq b_n < 1, \quad 1 \leq n \leq N-1,$$

$$b_0 = 1, \quad \text{and} \quad b_n = 0, \quad n \geq N.$$

2. After joining the queue each customer will wait a certain length of time $T$ for service to begin. If it has not begun by then, he will get impatient and leave the queue without getting service. This time $T$ is a random variable whose density function is given by

$$d(t) = \alpha e^{-\alpha t}, \quad t \geq 0, \quad \alpha > 0,$$

where $\alpha$ is the rate of time $T$. Since the arrival and departure of the impatient customers without service are independent, the average reneging rate of the customer can be given by $(n-i)\alpha$. Hence, the function of customer’s average reneging rate is given by

$$r(n) = (n-i)\alpha, \quad i \leq n \leq N, \quad i = 0, 1$$

$$r(n) = 0, \quad n > N.$$

3. The customers are served on a first-come, first served (FCFS) discipline. Once service commences it always proceeds to completion. The service times are assumed to be distributed according to an exponential distribution with density function as follows:

$$s(t) = \mu e^{-\mu t}, \quad t \geq 0, \quad \mu > 0,$$

where $\mu$ is the service rate.

2.3 STEADY-STATE PROBABILITY

In this part, we derive the steady-state probabilities by the Markov process method. Let $p(n)$ be the probability that there are $n$ customers in the system. Applying the Markov process theory, we obtain the following set of steady-state equations.

$$\mu P(1) = \lambda P(0), \quad n = 0$$

$$\lambda b_{n-1} P(n-1) + (\mu + n\alpha)P(n+1) = [\lambda b_n + \mu + (n-1)\alpha]P(n), \quad n = 1, 2, \ldots, N-1.$$  

$$\lambda b_{N-1} P(N-1) = [\mu + (n-1)\alpha]P(n), \quad n = N.$$
For \( N = 3 \) steady state difference equations are:

\[
\begin{align*}
\mu P(1) &= \lambda P(0), & n &= 0 \\
\lambda P(0) + (\mu + \alpha)P(2) &= [\lambda b_1 + \mu]P(1), & n &= 1 \\
\lambda b_1 P(1) + (\mu + 2\alpha)P(3) &= [\lambda b_2 + \mu + \alpha]P(2), & n &= 2 \\
\lambda b_2 P(2) &= (\mu + 2\alpha)P(3), & n &= 3
\end{align*}
\]

Matrix form of these equations as follows:

\[
\begin{bmatrix}
-\lambda & \mu & 0 & 0 \\
-\lambda & (\lambda b_1 + \mu) & (\mu + \alpha) & 0 \\
0 & b_1 & -(\lambda b_2 + \mu + \alpha) & (\mu + 2\alpha) \\
0 & 0 & \lambda b_2 & -(\mu + 2\alpha)
\end{bmatrix}
\begin{bmatrix}
P(0) \\
P(1) \\
P(2) \\
P(3)
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

On solving these we get,

\[
\begin{align*}
P(0) &= k \\
P(1) &= \frac{\lambda k}{\mu} \\
P(2) &= \frac{\lambda^2 b_1 k}{\mu(\mu + \alpha)} \\
P(3) &= \frac{\lambda^3 b_1 b_2 k}{\mu(\mu + \alpha)(\mu + 2\alpha)}.
\end{align*}
\]

Using

\[
P(0) + P(1) + P(2) + P(3) = 1,
\]

we get,

\[
k = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda^2 b_1}{\mu(\mu + \alpha)} + \frac{\lambda^3 b_1 b_2}{\mu(\mu + \alpha)(\mu + 2\alpha)} \right)^{-1}.
\]
Therefore,

\[ P(0) = \left( 1 + \frac{\lambda}{\mu} + \frac{\lambda^2 b_1}{\mu(\mu + \alpha)} + \frac{\lambda^3 b_1 b_2}{\mu(\mu + \alpha)(\mu + 2\alpha)} \right)^{-1} \]

\[ P(1) = \frac{\lambda}{\mu} \left( 1 + \frac{\lambda}{\mu} + \frac{\lambda^2 b_1}{\mu(\mu + \alpha)} + \frac{\lambda^3 b_1 b_2}{\mu(\mu + \alpha)(\mu + 2\alpha)} \right)^{-1} \]

\[ P(2) = \frac{\lambda^2 b_1}{\mu(\mu + \alpha)} \left( 1 + \frac{\lambda}{\mu} + \frac{\lambda^2 b_1}{\mu(\mu + \alpha)} + \frac{\lambda^3 b_1 b_2}{\mu(\mu + \alpha)(\mu + 2\alpha)} \right)^{-1} \]

\[ P(3) = \frac{\lambda^3 b_1 b_2}{\mu(\mu + \alpha)(\mu + 2\alpha)} \left( 1 + \frac{\lambda}{\mu} + \frac{\lambda^2 b_1}{\mu(\mu + \alpha)} + \frac{\lambda^3 b_1 b_2}{\mu(\mu + \alpha)(\mu + 2\alpha)} \right)^{-1}. \]

In general for \( N = n \)

\[ P(n) = \frac{\lambda^n b_1 b_2 \ldots b_{n-1} k}{\mu(\mu + \alpha)(\mu + 2\alpha) \ldots (\mu + (n-1)\alpha)}, \]

and using

\[ P(0) + P(1) + P(2) + \ldots + P(N) = 1, \]

we get

\[ k = \left( 1 + \frac{\lambda}{\mu} + \frac{\lambda^2 b_1}{\mu(\mu + \alpha)} + \frac{\lambda^3 b_1 b_2}{\mu(\mu + \alpha)(\mu + 2\alpha)} + \ldots + \frac{\lambda^n b_1 b_2 \ldots b_{N-1}}{\mu(\mu + \alpha)(\mu + 2\alpha) \ldots (\mu + (N-1)\alpha)} \right)^{-1}. \]

**2.4 PERFORMANCE MEASURES AND COST MODEL**

In this part of this chapter, we give some performance measures of the system. Based on these performance measures, we develop a cost model to determine the optimal service rate.

**2.4.1 PERFORMANCE MEASURES**

Using the steady-state probability presented in part 2.3 of this chapter, we can obtain some performance measures of the system, such as the busy probability of the server \( P_B \), the expected number of the waiting customers \( E(N_q) \) and the expected number of the customers in the system \( E(N) \) as follows:
\[ E(N_q) = \sum_{n=1}^{N} (n-1)P(n) \]

\[
= \sum_{n=1}^{N} (n-1) \left\{ \left( \lambda^n b_1 b_2 b_3 \cdots b_{n-1} \right) \left( 1 + \frac{\lambda}{\mu} + \frac{\lambda^2 b_1}{\mu(\mu+\alpha)} + \frac{\lambda^3 b_1 b_2}{\mu(\mu+\alpha)(\mu+2\alpha)} + \cdots \right)^{-1} \right\}
\]

\[
= \sum_{n=1}^{N} \left( n - 1 \right) \left( \frac{\lambda^n b_1 b_2 b_3 \cdots b_{n-1}}{\mu(\mu+\alpha)(\mu+2\alpha)\cdots(\mu+(n-1)\alpha)} \right)^{-1}
\]

\[= \sum_{n=1}^{N} (n-1)P(n) \]

\[E(N) = \sum_{n=1}^{N} nP(n)\]

\[
= \sum_{n=1}^{N} n \left\{ \left( \lambda^n b_1 b_2 b_3 \cdots b_{n-1} \right) \left( 1 + \frac{\lambda}{\mu} + \frac{\lambda^2 b_1}{\mu(\mu+\alpha)} + \frac{\lambda^3 b_1 b_2}{\mu(\mu+\alpha)(\mu+2\alpha)} + \cdots \right)^{-1} \right\}
\]

\[
= \sum_{n=1}^{N} \left( \frac{\lambda^n b_1 b_2 b_3 \cdots b_{n-1}}{\mu(\mu+\alpha)(\mu+2\alpha)\cdots(\mu+(n-1)\alpha)} \right)^{-1}
\]

\[= \sum_{n=1}^{N} n \left( \frac{\lambda^n b_1 b_2 b_3 \cdots b_{n-1}}{\mu(\mu+\alpha)(\mu+2\alpha)\cdots(\mu+(n-1)\alpha)} \right)^{-1}
\]

\[B.R. = \sum_{n=1}^{N} \lambda(1-b_n)P(n)\]

\[
= \sum_{n=1}^{N} \lambda \left( 1 - b_n \right) \left\{ \left( \lambda^n b_1 b_2 b_3 \cdots b_{n-1} \right) \left( 1 + \frac{\lambda}{\mu} + \frac{\lambda^2 b_1}{\mu(\mu+\alpha)} + \frac{\lambda^3 b_1 b_2}{\mu(\mu+\alpha)(\mu+2\alpha)} + \cdots \right)^{-1} \right\}
\]

\[
= \sum_{n=1}^{N} \lambda(1-b_n) \left( \frac{\lambda^n b_1 b_2 b_3 \cdots b_{n-1}}{\mu(\mu+\alpha)(\mu+2\alpha)\cdots(\mu+(n-1)\alpha)} \right)^{-1}
\]

(2.1)

(2.2)

(2.3)
\[ R.R. = \sum_{n=1}^{N} (n-1)\alpha P(n) \]

\[ = \sum_{n=1}^{N} (n-1)\alpha \left( \frac{1 + \frac{\lambda}{\mu} + \frac{\lambda^2 b_1}{\mu(\mu + \alpha)} + \frac{\lambda^3 b_2}{\mu(\mu + \alpha)(\mu + 2\alpha)} + \ldots}{\mu(\mu + \alpha)(\mu + 2\alpha)\ldots(\mu + (N-1)\alpha)} \right)^{-1} \]

\[ = \sum_{n=1}^{N} (n-1)\alpha \left( \frac{1 + \frac{\lambda}{\mu} + \frac{\lambda^2 b_1}{\mu(\mu + \alpha)} + \frac{\lambda^3 b_2}{\mu(\mu + \alpha)(\mu + 2\alpha)} + \ldots}{\mu(\mu + \alpha)(\mu + 2\alpha)\ldots(\mu + (n-1)\alpha)} \right)^{-1} \]

\[ L.R. = B.R. + R.R. \]

\[ = \sum_{n=1}^{N} \lambda (1 - b_n) \left( \frac{1 + \frac{\lambda}{\mu} + \frac{\lambda^2 b_1}{\mu(\mu + \alpha)} + \frac{\lambda^3 b_2}{\mu(\mu + \alpha)(\mu + 2\alpha)} + \ldots}{\mu(\mu + \alpha)(\mu + 2\alpha)\ldots(\mu + (N-1)\alpha)} \right)^{-1} \]

\[ + \sum_{n=1}^{N} (n-1)\alpha \left( \frac{1 + \frac{\lambda}{\mu} + \frac{\lambda^2 b_1}{\mu(\mu + \alpha)} + \frac{\lambda^3 b_2}{\mu(\mu + \alpha)(\mu + 2\alpha)} + \ldots}{\mu(\mu + \alpha)(\mu + 2\alpha)\ldots(\mu + (N-1)\alpha)} \right)^{-1} \]

\[ (2.4) \]

\[ (2.5) \]
2.4.2 COST MODEL

In this subpart of this chapter, we develop an expected cost model, in which service rate $\mu$ is the control variable. Our objective is to control the service rate to minimize the system’s total average cost per unit. Let

$C_1 \equiv$ cost per unit time when the server is busy,

$C_2 \equiv$ cost per unit time when a customer joins in the queue and waits for service,

$C_3 \equiv$ cost per unit time when a customer balks or reneges.

Using the definitions of each cost element listed above, the total expected cost function per unit time is given by

$$F(\mu) = C_1 P_B + C_2 E(N_q) + C_3 L.R.$$
where $E(N_q)$, $E(N)$, $B.R.$, $R.R.$ and $L.R.$ are given in Eqs. (2.1) – (2.5). The first item is the cost incurred by the server. The second item $C_2 E(N_q)$ is the cost incurred by the customer’s waiting. The last item $C_3 L.R.$ is the cost incurred by the customer loss.

2.4.3 NUMERICAL RESULTS

In this subpart of chapter, we present some numerical examples to demonstrate how the various parameters of the model influence the optimal service rate $\mu^*$, the optimal expected cost of the system $F(\mu^*)$ and other performance measures of the system. We fix the maximum number of customers in the system $N = 3$, the probability $b_n = 1/(n + 1)$ and the cost elements $C_1 = 15$, $C_2 = 12$, $C_3 = 18$.

Table 2.1: Performance measures for $\alpha = 0.1$.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu^*$</td>
<td>0.9468</td>
<td>0.9089</td>
<td>0.8754</td>
<td>0.8444</td>
<td>0.8155</td>
<td>0.7882</td>
</tr>
<tr>
<td>$F(\mu^*)$</td>
<td>17.2026</td>
<td>18.4446</td>
<td>19.9224</td>
<td>21.6150</td>
<td>23.4906</td>
<td>25.5186</td>
</tr>
<tr>
<td>$E(N_q)$</td>
<td>0.0658</td>
<td>0.1039</td>
<td>0.1495</td>
<td>0.2016</td>
<td>0.2592</td>
<td>0.3211</td>
</tr>
<tr>
<td>$E(N)$</td>
<td>0.4047</td>
<td>0.5180</td>
<td>0.6327</td>
<td>0.7481</td>
<td>0.8627</td>
<td>0.9757</td>
</tr>
<tr>
<td>$L.R.$</td>
<td>0.0785</td>
<td>0.1221</td>
<td>0.1738</td>
<td>0.2331</td>
<td>0.2989</td>
<td>0.3703</td>
</tr>
</tbody>
</table>

Table 2.2: Performance measures for $\lambda = 0.5$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.05</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu^*$</td>
<td>0.9652</td>
<td>0.9089</td>
<td>0.8010</td>
<td>0.6993</td>
<td>0.6041</td>
<td>0.5160</td>
</tr>
<tr>
<td>$E(N_q)$</td>
<td>0.1006</td>
<td>0.1039</td>
<td>0.1105</td>
<td>0.1173</td>
<td>0.1239</td>
<td>0.1302</td>
</tr>
<tr>
<td>$E(N)$</td>
<td>0.5002</td>
<td>0.5180</td>
<td>0.5550</td>
<td>0.5956</td>
<td>0.6830</td>
<td>0.6825</td>
</tr>
<tr>
<td>$L.R.$</td>
<td>0.1131</td>
<td>0.1228</td>
<td>0.1440</td>
<td>0.1674</td>
<td>0.1937</td>
<td>0.2196</td>
</tr>
</tbody>
</table>

First, we select the rate of the waiting time $\alpha = 0.1$ and change values of arrival rate of customers $\lambda$. The numerical results are summarized in Table 2.1. This shows that: (i) the optimal service rate $\mu$ decreases with the increasing $\lambda$ and its minimum expected cost $F(\mu^*)$ increases with the increasing
λ; (ii) the expected number of the waiting customers \( E(N_q) \), the expected number of customers in the system \( E(N) \) and the average rate of customer loss \( L.R. \) all increases with the increase of \( \lambda \). This is because the number of the customers in the system increases with the increase of \( \lambda \). Thus, \( E(N_q) \) and \( L.R. \) all increases which result in the increase of the optimal cost.

Finally, we select \( \lambda = 0.5 \), and change values of \( \alpha \). The numerical results are summarized in Table 2.2. This shows that: (i) the optimal service rate \( \mu^* \) decreases with the increasing \( \alpha \), and its minimum expected cost \( F(\mu^*) \) increases with the increasing \( \alpha \); (ii) \( E(N_q) \), \( E(N) \) and \( L.R. \) all increases with the increasing \( \alpha \). Thus \( E(N_q) \) and \( L.R. \) all increases which result in the increase of the optimal cost. The following graph shows the performance of the model at \( \alpha = 0.1 \) and \( \lambda = 0.5 \). Figures 2.1 and 2.5 have been drawn for \( \alpha = 0.1 \) and they depict the performance of the model. Similarly, Figure 2.6 and 2.10 are for \( \lambda = 0.5 \).

![Graph between arrival rate and optimal service rate](image)

**Figure 2.1**: Graph between arrival rate and optimal service rate
Figure 2.2: Graph between arrival rate and expected optimal cost function

Figure 2.3: Graph between arrival rate and expected queue length
Figure 2.4: Graph between arrival rate and expected length of system

Figure 2.5: Graph between arrival rate and average rate loss of customers
Figure 2.6: Graph between reneging rate and optimal service rate

Figure 2.7: Graph between reneging rate and expected optimal cost function
Figure 2.8: Graph between reneging rate and expected queue length

Figure 2.9: Graph between reneging rate and expected length of system
2.5 CONCLUSION

In this chapter, we considered an $M/M/1/N$ queueing system with balking and reneging. We developed the equations of the steady state probabilities and derived the matrix form solution of the steady-state probabilities. We also gave some performance measures of the system, and formulated a cost model to determine the optimal service rate. Although the function of the cost is too complicated to derive the explicit expression of the optimal service rate, the performance measures and the optimal service rate can be numerically evaluated by the formula in part 2.4 of this chapter. Some numerical examples were presented to demonstrate how the various parameters of the model influence the behavior of the system.