Chapter 4
ANALYSIS OF IMPATIENCE IN A MARKOVIAN QUEUING MODEL
WITH NO WAITING SPACE

4.1. Introduction:

In the previous chapter, we have discussed the analysis of a multi-server Markovian queuing system where the system capacity is infinite. In this chapter, we discuss a multiserver Markovian queuing system with finite system capacity with the additional restriction that there is no waiting space. Symbolically it is denoted by $M/M/c/c$. It is also called a c-channel loss system and was first investigated by Erlang (Medhi, 2003). Here we assume that the number of servers in the system is equal to the system capacity. To appreciate the importance of this model, we recall that in the classical $M/M/k$ model, 'it is assumed that the system can accommodate any number of units. In practice this may seldom be the case. We have thus to consider the situation such that the system has limited waiting space and can hold a maximum number of $k$ units (including the one being served)' (Medhi, 1994). In certain queuing systems, waiting space may not be feasible. E.g. in the emergency room of a hospital. Hence the importance of the additional restriction that system size is equal to number of servers. Since the number of servers is equal to the system capacity, a customer arriving at it either goes straight into the service or is turned away without service (as there is no waiting space) if all the servers busy. As waiting is not allowed, balking is not possible in this model. As regards reneging, it is obvious that the reneging rule in the queuing model can only be of R_EOS type.
Thus, we are imposing the additional restriction that customers may renege under R_EOS.

While a queuing system without reneging is a wishful thought, nevertheless it is an inevitable part of such any queuing system. The particular interest in this chapter is reneging behavior of customers and its influence on system parameters. To the best of our knowledge, an explicit analysis providing closed form expressions with reneging of this model has not been carried out. This forms the motivation of our work.

4.2 Assumptions:

The assumptions of this model are:

1) Arrival rate of customer follows exp (λ).
2) Service rate of customer follows exp (μ).
3) There are c servers in parallel that are independent of each other. Further, their efficiencies are same and the service time distribution of each of these servers is exp (μ).
4) As for reneging, each customer joining the system is assumed to have random patience time following exp (ν) which is independent of the position of the customer in the system. Clearly, the reneging rate from the system will be a function of the system state. With system in the state 'n' (n≤c), the reneging rate from the system will be nv. This implies that reneging rate from the system is not constant and is a function of number of
customers in the system. Perhaps this is intuitively appealing, as one would expect higher reneging rate with increase in system size.

5) There can be at most ‘c’ customers in the system at any time i.e. system capacity is restricted to ‘c’.

6) Service discipline is considered as FIFO.

7) The size of calling population is infinite.

4.3. The System State Probabilities:

In this section, the steady state probabilities are derived by the Markov process method. The rate-in-rate-out diagram is given below:

Let $p_n$ denote the probability that there are ‘n’ customers in the system. Applying the Markov process theory, we obtain the following set of steady state equations.

\[
\lambda p_0 = (\mu + \nu) p_1
\]
\[
\lambda p_{n-1} + (n + 1)(\mu + \nu)p_{n+1} = \lambda p_n + n(\mu + \nu)p_n \quad 1 \leq n \leq c - 1
\]
\[
\lambda p_{c-1} = c(\mu + \nu)p_c
\]

Solving recursively, we get
Chapter 4

\[ p_n = \frac{\lambda^n}{n!(\mu + \nu)^n} p_0 \quad ; n=1, 2, \ldots c \]

where \( p_0 \) is obtained from the normalizing condition \( \sum_{n=0}^{c} p_n = 1 \) and is given as

\[
p_0 = \left[ 1 + \frac{\lambda}{\mu + \nu} \right]^{-1} \left[ \sum_{n=0}^{c} \frac{\lambda^n}{n!(\mu + \nu)^n} \right]^{-1}
\]

(4.3.4)

It represents the probability that all the servers are idle. Then

\[ p_n = \frac{\binom{\lambda}{n} / n!}{\sum_{n=0}^{c} \binom{\lambda}{n} / n!} \]

which is the analogue of Erlang's first formula under reneging. Since an arriving unit who finds all channels busy leaves the system, the probability of this event is

\[ p_c = \frac{\lambda^c}{(\mu + \nu)^c c!} p_0 \]

which is modified Erlang's loss formula or blocking formula for reneging. We shall denote it by \( B_2(c, \lambda/\mu+\nu) \).

4.4 Performance Measures:

The mean number of customers in the system is an important measure and it is denoted by 'L'. To obtain an expression for this, we had noted that \( L = P'(1) \) where

\[ P'(1) = \frac{d}{ds} P(s) \bigg|_{s=1} \]
Here \( P(S) \) is the p.g.f. of the steady state probabilities and is given by

\[
P(s) = \sum_{n=0}^{c-1} p_n s^n.
\]

To obtain \( P'(1) \) under R_EOS we proceed as follows:

From equation (4.3.2) we have

\[
\lambda p_{n+1} + (n+1)(\mu + \nu)p_{n+1} = \lambda p_n + n(\mu + \nu)p_n \quad n = 1, 2, 3, ..., c-1
\]

Now multiplying both sides of the equation by \( s^n \) and summing over \( n \)

\[
\lambda s \sum_{n=1}^{c-1} p_{n+1} s^{n-1} + \frac{1}{s} \sum_{n=1}^{c-1} (n+1)(\mu + \nu)p_{n+1} s^{n-1} = \lambda \sum_{n=1}^{c-1} p_n s^n - \frac{1}{s} \sum_{n=1}^{c-1} n(\mu + \nu)p_n s^n
\]

\[
\Rightarrow \lambda s \sum_{n=1}^{c-1} p_{n+1} s^{n-1} - \lambda \sum_{n=1}^{c-1} p_n s^n = \sum_{n=1}^{c-1} n(\mu + \nu)p_n s^n - \frac{1}{s} \sum_{n=1}^{c-1} (n+1)(\mu + \nu)p_{n+1} s^{n-1}
\]

\[
\Rightarrow \lambda s[\mu s^0 + p_1 s^1 + 2p_2 s^2 + 3p_3 s^3 + ... + (c-1)p_{c-1} s^{c-1}] - \lambda [p_1 s^1 + 2p_2 s^2 + 3p_3 s^3 + ... + c p_c s^c]
\]

\[
= [(\mu + \nu)p_1 s^1 + 2(\mu + \nu)p_2 s^2 + ... + (c-1)(\mu + \nu)p_{c-1} s^{c-1}] - \frac{1}{s} [(\mu + \nu)p_2 s^2 + 3(\mu + \nu)p_3 s^3 + ... + c(\mu + \nu)p_c s^c]
\]

\[
\Rightarrow \lambda s[p(s) - p_{c+1} s^c] - \lambda [p(s) - p_0 s^0]
\]

\[
= s(\mu + \nu)[p_1 s^0 + 2p_2 s^1 + 3p_3 s^2 + ... + (c-1)p_{c-1} s^{c-1}] - (\mu + \nu)[2p_2 s^2 + 3p_3 s^3 + ... + c p_c s^c]
\]

\[
\Rightarrow \lambda s[p(s) - p_{c+1} s^c - p_0 s^0] = s(\mu + \nu)[p'(s) - c p_c s^{c-1}] - (\mu + \nu)[p'(s) - p_0]
\]

\[
\Rightarrow (\mu + \nu)p'(s) - s(\mu + \nu)p(s) = (\mu + \nu)p_1 - c(\mu + \nu)p_c s^c - \lambda s[p(s) - p_{c+1} s^c] + \lambda s[p(s) - p_0 s^0]
\]

\[
\Rightarrow (\mu + \nu)(1-s)p'(s) = \lambda (1-s)p(s) - \lambda (1-s)p_c s^c
\]

\[
\Rightarrow p'(s) = \frac{\lambda}{(\mu + \nu)}[p(s) - p_c s^c]
\]

Now

\[
\lim_{s \to 1^-} p'(s) = \lim_{s \to 1^-} \left[ \frac{\lambda}{(\mu + \nu)} \{p(s) - p_c s^c\} \right]
\]

115
Thus, the mean system size is given by

\[ L = \frac{\lambda(1 - p_c)}{(\mu + \nu)} \]

Mean system size for the particular case with no reneging can be similarly derived. In that case we have,

\[ L_{(v=0)} = \frac{\lambda(1 - p_c)}{\mu} \]

Customers arrive into the system at the rate of \( \lambda \). However all the customers who arrive do not join the system because of finite buffer restriction. The effective arrival rate into the system is thus different from the overall arrival rate and is given by

\[ \lambda^e = \lambda p_0 + \lambda p_1 + \ldots + \lambda p_{c-1} + 0.p_c \]

\[ = \lambda \sum_{n=0}^{c-1} p_n \]

\[ = \lambda(1 - p_c) \]

where \( p_c = \frac{\lambda^e}{c!(\mu + \nu)^c} p_0 \)

We have assumed that each customer has a random patience time following \( \text{exp} (v) \). Clearly then, the reneging rate of the system would depend on the state of the system. The average reneging rate (avg rr) is given by
Chapter 4

$Avg \ rr = \sum_{n=1}^{k} n \lambda p_n$

$= \nu p'(1)$

$= \lambda \nu (1 - p_c) / (\mu + \nu)$

In system management, customers who renege represent business lost.

It is therefore of interest to determine the proportion of customers lost, both out of these joining the system as well as out of those arriving into the system. These are given below

Proportion of customer lost due to reneging out of those arriving and joining the system is

$= \frac{Avgrr}{\lambda}$

$= \frac{\lambda \nu / (\mu + \nu)}$

Proportion of customer lost due to reneging out of total customers arriving in the system is

$= \frac{Avgrr}{\lambda}$

$= \nu (1 - p_c) / (\mu + \nu)$

We know that, in totality, customers are lost to the system in two ways, due to finite buffer and due to reneging. The management would like to know the proportion of total customers lost in order to have an idea about the total business lost. Hence the mean rate at which customers are lost is

Rate of loss due to finite buffer + Avgrr

$= \lambda - \lambda^* + Avgrr$

$= \{\lambda \nu (1 - p_c) / (\mu + \nu)\} + \lambda p_c$
This rate helps in the determination of proportion of customers lost which is

\[ \frac{\lambda - \lambda^e + \text{Avgrr}}{\lambda} \]

\[ = \frac{\nu + \mu P_e}{(\mu + \nu)} \]

The proportion of customers completing service is its complement.

4.5. Sensitivity Analysis:

We presented below the effect of change in the system parameters on server utilization. Let \( p_n(\lambda, \mu, \nu, c) \) denotes the probability that there are ‘n’ customers in a system with parameters \( \lambda, \mu, \nu, c \) in steady state under R_EOS.

It can be shown that

i) Let \( \lambda > \lambda_0 \) then

\[
\frac{P_0(\lambda_1, \mu, \nu, c)}{P_0(\lambda_0, \mu, \nu, c)} < 1
\]

\[
\Rightarrow \frac{(\lambda_0^2 - \lambda_1^2)}{(\mu + \nu)} + \frac{(\lambda_0^2 - \lambda_1^2)}{2!(\mu + \nu)^3} + \cdots + \frac{c!(\mu + \nu)^c}{(\mu + \nu)^c} < 0
\]

which is true and hence \( p_0 \downarrow \) as \( \lambda \uparrow \).

ii) Let \( \mu_1 > \mu_0 \) then

\[
\frac{P_0(\lambda, \mu_1, \nu, c)}{P_0(\lambda, \mu_0, \nu, c)} > 1
\]

\[
\Rightarrow \lambda \left( \frac{1}{(\mu_0 + \nu)} - \frac{1}{(\mu_1 + \nu)} \right) + \frac{\lambda^2}{2!} \left( \frac{1}{(\mu_0 + \nu)^2} - \frac{1}{(\mu_1 + \nu)^2} \right) + \cdots + \frac{\lambda^c}{c!} \left( \frac{1}{(\mu_0 + \nu)^c} - \frac{1}{(\mu_1 + \nu)^c} \right) > 0
\]

which is true and hence \( p_0 \uparrow \) as \( \mu \uparrow \).
Chapter 4

iii) Let \( v_1 > v_0 \) then

\[
\frac{p_0 (\lambda, \mu, v_1, c)}{p_0 (\lambda, \mu, v_0, c)} > 1
\]

\[
= \lambda \left( \frac{1}{(\mu + v_0)} - \frac{1}{(\mu + v_1)} \right) + \frac{\lambda^2}{2!} \left( \frac{1}{(\mu + v_0)^2} - \frac{1}{(\mu + v_1)^2} \right) + \ldots
\]

\[
\frac{\lambda^c}{c!} \left( \frac{1}{(\mu + v_0)^c} - \frac{1}{(\mu + v_1)^c} \right) > 0
\]

which is true and hence \( p_0 \uparrow \text{ as } v \uparrow \).

iv) Let \( c_1 > c_0 \) then

\[
\frac{p_0 (\lambda, \mu, v, c_1)}{p_0 (\lambda, \mu, v, c_0)} < 1
\]

\[
\Rightarrow \sum_{n=1}^{c_0} \frac{\lambda^n}{n!(\mu + v)^n} - \sum_{n=1}^{c_0} \frac{\lambda^n}{n!(\mu + v)^n} < 0
\]

which is true and hence \( p_0 \downarrow \text{ as } c \uparrow \).

These results state that an increase in arrival rate would result in lowering of the fraction of time the server is idle. An increase in service rate would mean the server is able to work efficiently so that it can process same amount of work quickly. This translates to higher server idle time. An increase in reneging rate would mean the server has fewer work to do and hence higher fraction of idle time. If the number of server and the system capacity increased which may increase the arrival rate of customer would result in lowering of the fraction of time the server is idle.
4.6. Numerical Example:

To illustrate the use of our results, we apply them to a queuing problem. The problem we describe below has been suitably adopted from an example from Ravindran, Phillips and Solberg (1987, page 338). While we have not modified the system parameters, the set up has been changed to make it more relevant to the model in hand.

Consider the operation theatre (O. T.) of the emergency unit of a small hospital. The O.T. of this unit contains two beds, which are manned, by doctors and paramedical staff round the clock. Since this is an emergency unit, there is no space for patients to wait. The state of the system is the number of patients in the O.T. 0, 1 and 2. If there is an empty bed when a customer arrives, he enters the O. T. and his treatment begins. If both the beds are occupied when he arrives, he does not enter the O. T. and leaves for another hospital. As soon as a patient's emergency treatment is over, he is shifted to the ward instantaneously. On the average, a patient arrives every 10 minutes and each patient takes an average of 15 minutes of time on the O. T. bed.

Notice that in the above problem, some potential patients are turned away. If the O. T. had another bed, it might be able to profit from additional paying patients. On the other hand, the additional facility (bed) would have to be paid. The hospital management would be interested to know if this additional investment would be worthwhile.

We assume markovian arrival and service distribution. Further, since arriving patients who do not find the bed instantaneously leave the hospital, this is a finite buffer queuing system with no waiting space. Since the emergency patients
arrive, it is possible that some patients would expire while being treated. This phenomenon would be a case fit for analysis using the concept of reneging till end of service. We assume two alternative scenarios. In the 1st scenario mean reneging rate is assumed to be 1 hour (v=1/hr) and in the 2nd scenario the assumption is 30 minutes (v=2/hr).

Various performance measures of interest computed under the two reneging scenarios are given in Table 4.1 and Table 4.2. These measures were arrived at using a FORTRAN 77 program coded by the authors. The whole objective is to examine how the performance measures vary with the increase in bed capacity.

**Table 4.1: Performance Measures assuming λ=6/hr, μ=4/hr, v=1/hr.**

<table>
<thead>
<tr>
<th>Performance Measure</th>
<th>Number of beds in O.T.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>c=2</td>
<td>c=3</td>
</tr>
<tr>
<td>Proportion of customers completing service.</td>
<td>0.60274</td>
<td>0.72817</td>
</tr>
<tr>
<td>Fraction of time that all server is idle (p₀)</td>
<td>0.34246</td>
<td>0.31172</td>
</tr>
<tr>
<td>Average length of system</td>
<td>0.9041</td>
<td>1.09227</td>
</tr>
<tr>
<td>Mean reneging rate</td>
<td>0.9041</td>
<td>1.09227</td>
</tr>
<tr>
<td>Rate of loss due to finite buffer.</td>
<td>1.47945</td>
<td>0.53865</td>
</tr>
<tr>
<td>Effective arrival rate</td>
<td>4.52055</td>
<td>5.46135</td>
</tr>
<tr>
<td>Proportion of customers lost (due to reneging and finite buffer)</td>
<td>0.39726</td>
<td>0.27182</td>
</tr>
</tbody>
</table>
Table 4.2: Performance Measures assuming $\lambda=6/hr$, $\mu=4/hr$, $v=2/hr$.

<table>
<thead>
<tr>
<th>Performance Measure</th>
<th>Number of beds in O.T.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c=2$</td>
</tr>
<tr>
<td>Proportion of customers completing service.</td>
<td>0.53333</td>
</tr>
<tr>
<td>Fraction of time that all server is idle ($p_0$)</td>
<td>0.4</td>
</tr>
<tr>
<td>Average length of system</td>
<td>0.8</td>
</tr>
<tr>
<td>Mean reneging rate</td>
<td>1.6</td>
</tr>
<tr>
<td>Rate of loss due to finite buffer.</td>
<td>1.2</td>
</tr>
<tr>
<td>Effective arrival rate</td>
<td>4.8</td>
</tr>
<tr>
<td>Proportion of customers lost (due to reneging and finite buffer)</td>
<td>0.46667</td>
</tr>
</tbody>
</table>

In both the scenarios, we observe that with increase in number of beds, the proportion of customers who complete receiving services go up. It goes up by 20.8% (Table 4.1) and 17.2% (Table 4.2). The average number of patients in the system also goes up. This is important from the revenue point of view.

In order to decide if the addition of an extra bed would be viable, this increase in revenue would have to be compared against the cost of setting up a new bed as well as increase in other recurring expenditure (eg. salary). The other performance measures also have shown substantial improvement with the increase in bed capacity.

*******