Unreliable Server M/M/1 Queue with Priority Queueing System

Richa Sharma, Gireesh Kumar

***Abstract*—This investigation deals with M/M/1 queue with priority queueing system and service interruption. There are the two types of customers namely priority and non-priority customers. The service of priority and non-priority customers may be interrupted due to the unavailability of the server. The arrival patterns of two types of the customers are independent Poisson processes wherein the service times are exponentially distributed. Moreover, Runge-Kutta (R-K) method of fourth order has been applied for obtaining the various performance measures of the customers in the system. To examine the effects of various system parameters, the numerical results are performed.**

***Index Terms*– priority queue; unreliable server; queue length; runge-kutta method; queue length.**

I. Introduction

With the advancement of technology, performance modeling is one of the important issues that affect the design, development, configuration and modification of any real time system. Server un-reliability makes negative impact on performance of any system and therefore some measures are to be taken to maintain a desired grade of service. In many practical applications, the server may fail and require repair (renewal). From a practical point of view, several authors have studied different models that describe the failure of a server and its renewal (repairs) as well as the rules of servicing a customer who finds a server in a broken down condition. Reference [1] analyzed an unreliable M/M/1 retrial queue with infinite capacity orbit and normal queue. The Markovian single server queue with breakdowns and repairs was studied by [2]. A single server unreliable queue was considered by [3] References [4-8] dealt with queue under the consideration of service interruption wherein the customers arrive at the service station according to Poisson process in different frameworks.

Priority queueing systems have numerous applications in many real life situations encountered in computer networks, communication systems, transportation, distribution systems and many others. Priority queue with combination of unreliable server has also been studied by a few researchers in different frameworks. The first published paper on preemptive priority discipline with unreliable server was due to [9]. They studied the preemptive priority system with two priority classes by using the method of generating function to obtain the steady state distribution. A single server queue with two classes of customers was considered by [10]. The analysis of unreliable server queue with two class non-preemptive priority subscribers was examined by [11]. A preemptive resume is followed by the server who is subjected to random failures. A single server retrial queue with priority subscribers and server breakdown was studied by [12]. Recently, reference [13] considered an unreliable single server queue with preemptive resume priorities and vacations.

Runge-Kutta (R-K) method is a useful technique for solving the set of differential equations. R-K method has been used to provide transient solution of queueing problems in different frameworks by many researchers. Reference [14] dealt with transient system behavior of the queueing system using the fourth order Runge-Kutta method. The queueing model for the unreliable flexible manufacturing cell (FMC) was developed by [15] which operates under stochastic environment. Using Runge-Kutta method, the probabilities for different system states have been evaluated.

In this investigation, we study two class preemptive priority queueing systems with unreliable server. The service of the priority and non-priority customers may be interrupted due to the failure of the server. The rest of the paper is organized as follows. Section II is devoted to model description. The governing equations of the model are given in section III. Various performance measures are discussed in section IV. The numerical results are carried out in section V. Finally, the conclusions are given in section VI.

II. Model Description

In this paper, we develop a Markovian preemptive priority queueing model wherein the service of both priority and non-priority customers may be interrupted due to unavailability of the server. We consider the two types of customer's namely higher priority and lower priority customers who arrive according to Poisson fashion in the system. A single server serves the customer under preemptive resume priority discipline. The arrival of higher priority customer interrupts the service of a lower priority customer who on its re-entry resumes service from the preempted point. The queue discipline is first come first served (FCFS) within their own queue. During the service of lower priority customer if higher priority customer joins the system then the service of the lower priority customer is interrupted. The service of lower priority customer restarts again when there is no higher priority customer present in the system. The server preempts the lower priority customer whose service is going on and higher priority customer joins the system. This discipline is known as *'preemptive priority'*.

Notations:

To formulate the mathematical model in terms of transient state probabilities, the following notations are used:

Richa Sharma, Department of Mathematics, JK Lakshmipat University, Jaipur -302026, India
Gireesh Kumar, Department of Computer Science & Engineering, JK Lakshmipat University, Jaipur -302026, India

λ_1 Arrival rate for higher priority customer

λ_2 Arrival rate for lower priority customer

μ_1 Ser vice rate for higher priority customer

μ_2 Ser vice rate for lower priority customer

α Failure Rate

β Repair Rate

L Finite capacity for higher priority customer

K Finite capacity for lower priority customer

$P_{i,j,1}(t)$ Probability that there are i $(0 \le i \le L)$ priority customers and j $(0 \le i \le K)$ non-priority customers present in the system when the sever is in busy state at time t

$P_{i,j,2}(t)$ Probability that there are i $(0 \le i \le L)$ priority customers and j $(0 \le i \le K)$ non-priority customers present in the system when the sever is in broken down state at time t.

III. Governing Equations

In this section, the differential difference equations governing the finite capacity model are constructed using fig. 1 as follows:

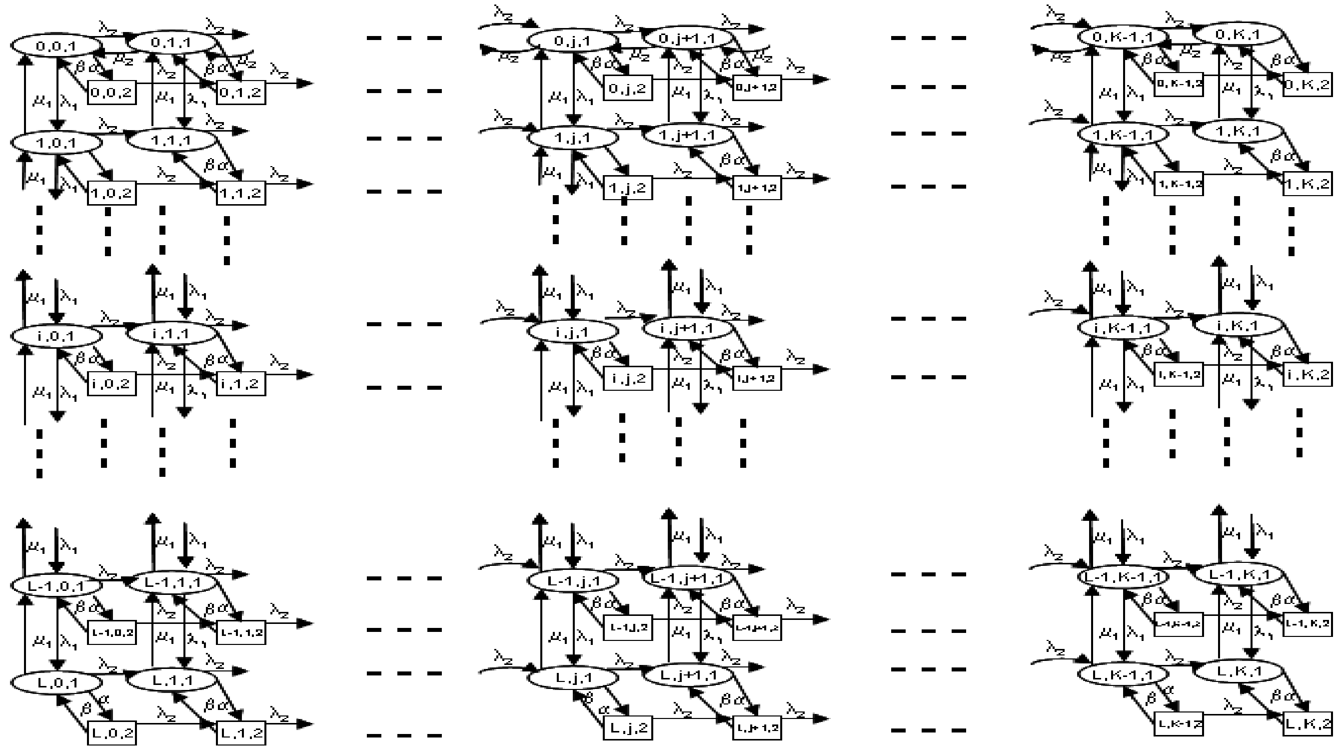

Fig. 1: State transition diagram for finite capacity model.

$$\frac{dP_{0,0,1}(t)}{dt} = -(\lambda_1 + \lambda_2 + \alpha)P_{0,0,1}(t) + \mu_1 P_{1,0,1}(t) + \mu_2 P_{0,1,1}(t) + \beta P_{0,0,2}(t) \quad (1)$$

$$\frac{dP_{0,j,1}(t)}{dt} = -(\lambda_1 + \lambda_2 + \alpha + \mu_2)P_{0,j,1}(t) + \mu_1 P_{1,j,1}(t) + \mu_2 P_{0,j+1,1}(t) + \lambda_2 P_{0,j-1,1}(t) + \beta P_{0,j,2}(t), \quad 1 \le j \le K-1 \quad (2)$$

$$\frac{dP_{0,K,1}(t)}{dt} = -(\lambda_1 + \alpha + \mu_2)P_{0,K,1}(t) + \mu_1 P_{1,K,1}(t) + \lambda_2 P_{0,K-1,1}(t) + \beta P_{0,k,2}(t) \quad (3)$$

$$\frac{dP_{0,0,2}(t)}{dt} = -(\lambda_1 + \lambda_2 + \beta)P_{0,0,2}(t) + \alpha P_{0,0,1}(t) \quad (4)$$

$$\frac{dP_{0,j,2}(t)}{dt} = -(\lambda_1 + \lambda_2 + \beta)P_{0,j,2}(t) + \alpha P_{0,j,1}(t) + \lambda_2 P_{0,j-1,2}(t), \quad 1 \le j \le K-1 \quad (5)$$

$$\frac{dP_{0,K,2}(t)}{dt} = -(\lambda_1 + \beta)P_{0,K,2}(t) + \alpha P_{0,K,1}(t) + \lambda_2 P_{0,K-1,2}(t), \; 1 \le j \le K-1 \quad (6)$$

$$\frac{dP_{i,0,1}(t)}{dt} = -(\lambda_1 + \lambda_2 + \alpha + \mu_1)P_{i,0,1}(t) + \mu_1 P_{i+1,0,1}(t) + \lambda_1 P_{i-1,0,1}(t) + \beta P_{i,0,2}(t), \; 1 \le i \le L-1 \quad (7)$$

$$\frac{dP_{i,j,1}(t)}{dt} = -(\lambda_1 + \lambda_2 + \alpha + \mu_1)P_{i,j,1}(t) + \mu_1 P_{i+1,j,1}(t) + \lambda_1 P_{i-1,j,1}(t) + \lambda_2 P_{i,j-1,1}(t) + \beta P_{i,j,2}(t), \quad 1 \le j \le K-1, 1 \le i \le L-1 \quad (8)$$

$$\frac{dP_{i,K,1}(t)}{dt}=-(\lambda_1+\alpha+\mu_1)P_{i,K,1}(t)+\mu_1P_{i+1,K,1}(t)+\lambda_1P_{i-1,K,1}(t)+\lambda_2P_{i,K-1,1}(t)+\beta P_{i,K,2}(t),\quad 1\le i\le L-1 \tag{9}$$

$$\frac{dP_{L,0,1}(t)}{dt}=-(\lambda_2+\alpha+\mu_1)P_{L,0,1}(t)+\lambda_1P_{L-1,0,1}(t)+\beta P_{L,0,2}(t) \tag{10}$$

$$\frac{dP_{L,j,1}(t)}{dt}=-(\lambda_2+\alpha+\mu_1)P_{L,j,1}(t)+\lambda_1P_{L-1,j,1}(t)+\lambda_2P_{L,j-1,1}(t)+\beta P_{L,j,2}(t),\quad 1\le j\le K-1 \tag{11}$$

$$\frac{dP_{L,K,1}(t)}{dt}=-(\alpha+\mu_1)P_{L,K,1}(t)+\lambda_1P_{L-1,K,1}(t)+\lambda_2P_{L,K-1,1}(t)+\beta P_{L,K,2}(t) \tag{12}$$

$$\frac{dP_{i,0,2}(t)}{dt}=-(\lambda_1+\lambda_2+\beta)P_{i,0,2}(t)+\alpha P_{i,0,1}(t)+\lambda_1P_{i-1,0,2}(t),\ 1\le i\le L-1 \tag{13}$$

$$\frac{dP_{i,j,2}(t)}{dt}=-(\lambda_1+\lambda_2+\beta)P_{i,j,2}(t)+\alpha P_{i,j,1}(t)+\lambda_1P_{i-1,j,2}(t)+\lambda_2P_{i,j-1,2}(t),\quad 1\le j\le K-1, 1\le i\le L-1 \tag{14}$$

$$\frac{dP_{i,K,2}(t)}{dt}=-(\lambda_1+\beta)P_{i,K,2}(t)+\alpha P_{i,K,1}(t)+\lambda_1P_{i-1,K,2}(t)+\lambda_2P_{i,K-1,2}(t),\ \ 1\le i\le L-1 \tag{15}$$

$$\frac{dP_{L,0,2}(t)}{dt}=-(\lambda_2+\beta)P_{L,0,2}(t)+\alpha P_{L,0,1}(t)+\lambda_1P_{L-1,0,2}(t) \tag{16}$$

$$\frac{dP_{L,j,2}(t)}{dt}=-(\lambda_2+\beta)P_{L,j,2}(t)+\alpha P_{L,j,1}(t)+\lambda_1P_{L-1,j,2}(t)+\lambda_2P_{L,j-1,2}(t),\ \ 1\le j\le K-1 \tag{17}$$

$$\frac{dP_{L,K,2}(t)}{dt}=-\beta P_{L,K,2}(t)+\alpha P_{L,K,1}(t)+\lambda_1P_{L-1,K,2}(t)+\lambda_2P_{L,K-1,2}(t) \tag{18}$$

IV. Performance Measures

Various performance indices in terms of transient probabilities have been obtained using Runge-Kutta (R-K) method. R-K technique is implemented by developing program in MATLAB software. After obtaining transient probabilities, some performance indices are calculated as:

- The probability of the server being idle at time t is

$$P_I(t)=P_{0,0,1}(t) \tag{37}$$

- The probability of the server being busy at time t is

$$P_B(t)=\sum_{j=1}^{K}P_{0,j,1}(t)+\sum_{i=1}^{L}\sum_{j=0}^{K}P_{i,j,1}(t) \tag{38}$$

- The probability of the server being broken down at time t is

$$P_D(t)=\sum_{i=0}^{L}\sum_{j=0}^{K}P_{i,j,2}(t) \tag{39}$$

- The expected number of the customers in the system at time t is given by

$$EN_S(t)=\sum_{i=0}^{L}\sum_{j=0}^{K}j\{P_{i,j,1}(t)+P_{i,j,2}(t)\} \tag{40}$$

V. Numerical Results

To solve the set of transient state equations (1)-(18), the coding of the computer program has been done in the 'MATLAB' software. Further, the numerical results are obtained for various performance indices viz. the probability of the server being idle, busy, broken down and expected number of the customers in the system at time 't'. For the computation purpose, we fix the default parameters as L=6, K=6, λ_1=0.3, λ_2=0.2, μ_1=2, μ_2=2, α=0.1, β=0.1 .

Tables I-II show the effect of arrival rate and service rate on various performance indices for varying values of time t. We notice that $P_B(t)$ and $P_D(t)$ increase whereas $P_I(t)$ decreases as time t grows. It is clear from tables I-II that $P_B(t)$ and $P_D(t)$ increase (decrease) with the increasing values of $\lambda_1(\mu_1)$ but the reverse trend is observed for $P_I(t)$.

Figs 2-3 show that $EN_S(t)$ increases sharply as time t grows. Also, an increasing trend is seen for $EN_S(t)$ on increasing the values of λ_1. It is also observed form fig. 3 that initially there is a gradual decrement in the queue length for higher values of μ_1.

TABLE I. Effect Of Arrival Rate On Various Performance Measures.

λ_1	t	$P_I(t)$	$P_B(t)$	$P_D(t)$
0.1	1.0	0.872	0.118	0.009
	1.3	0.852	0.135	0.012
	1.6	0.837	0.148	0.014
0.3	1.0	0.765	0.226	0.010
	1.3	0.731	0.262	0.013
	1.6	0.702	0.296	0.015

TABLE II. EFFECT OF SERVICE RATE ON VARIOUS PERFORMANCE MEASURES.

μ_1	t	$P_I(t)$	$P_B(t)$	$P_D(t)$
0.5	1.0	0.723	0.268	0.005
	1.3	0.665	0.326	0.007
	1.6	0.607	0.386	0.008
0.9	1.0	0.756	0.235	0.001
	1.3	0.716	0.277	0.006
	1.6	0.680	0.317	0.007

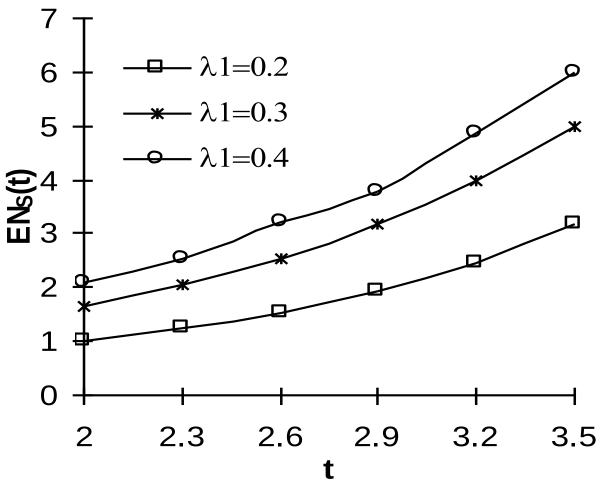

Fig. 2. Average queue length vs. time by varying arrival rate.

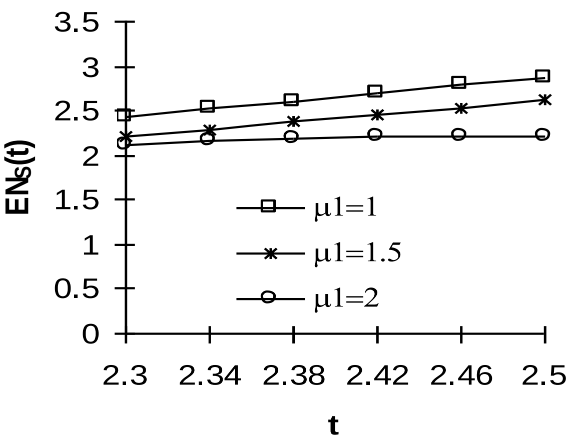

Fig. 3. Average queue length vs. time by varying service rate.

Overall based on numerical experiment performed by taking an illustration, we conclude that better service can be expected by the server up to a certain level if we increase the service rate. As expected, the average system size can also be controlled by improving the service rate.

VI. CONCLUSION

In this investigation, two class preemptive priority queueing system with service interruption has been studied. The provision of service interruption makes the model more versatile in depicting the many real time congestion situations encountered in industrial as well as day-to-day problems. The transient model discussed in the present investigation may provide a fruitful insight to the system developers and queue analysts for assessing the actual congestion problems based upon the performance measures so that the blocking and delay can be controlled.

REFERENCES

[1] N. P. Sherman, and J. P. Kharoufeh, "An M/M/1 retrial queue with unreliable server," Oper. Res. Lett., vol. 34, no. 6, 2006, pp. 697-705.

[2] A. Economou, and S. Kanta, "Equilibrium balking strategies in the observable single-server queue with breakdowns and repairs," Oper. Res. Lett., vol. 36, no. 6, 2008, pp. 696-699.

[3] A. Economou, and S. Kapodistria, "Synchronized abandonments in a single server unreliable queue," Euro. J. Oper. Res., vol. 203, no. 1, 2010, pp. 143-155.

[4] M. Jain, G. C. Sharma, and R. Sharma, "Working vacation queue with service interruption and multi-optional repair," Int. J. Inform. Mange. Sci., vol. 22, 2011, pp. 157-175.

[5] M. Jain, G. C. Sharma, and R. Sharma, "Optimal control of (N,F) policy for unreliable server queue with multi optional phase repair and start up," Int. J. Math. Oper. Res., vol. 4, no. 2, 2012, pp. 152-174.

[6] M. Jain, G. C. Sharma, and R. Sharma, "A batch arrival retrial queueing system with essential and optional services with server breakdown and bernoulli vacation," Int. J. Int. Enter. Manag., vol. 8, no. 1, 2012, pp. 16-45.

[7] M. Jain, G. C. Sharma, and R. Sharma, "Unreliable server M/G/1 queue with multi-optional services and multi-optional vacations," Int. J. Math. Oper. Res., vol. 5, no. 2, 2013, pp. 145-169.

[8] M. Jain, R. Sharma, and G. C. Sharma, "Multiple vacation policy for $M^X/H_k/1$ queue with un-reliable server," J. Indus. Engg. Int., vol. 9, no. 36, 2013, pp. 1-11.

[9] H. White, and L. Christie, "Queueing with preemptive priorities or with breakdown," Oper. Res., vol. 7, 1958, pp. 86-103.

[10] D. Towsley, and S. K. Tripathi, "A single server priority queue with server failures and queue flushing," Oper. Res. Lett., vol. 10, no. 6, 1991, pp. 253-362.

[11] M. Jain, and C. Bhargava, "Bulk arrival retrial queue with unreliable server and priority subscribers," Int. J. Oper. Res., vol. 5, no. 4, 2008, pp. 242-259.

[12] J. Wang, "On the single server retrial queue with priority subscribers and server breakdowns," J. Syst. Sci. Comp., vol. 21, 2008, pp. 304-315.

[13] L. Dimitriou, "A preemptive resume priority retrial queue with state dependent arrivals, unreliable server and negative customers," TOP, vol. 21, 2013, pp. 542-571.

[14] H. Van As, "Transient analysis of Markov queueing systems and its application to congestion-control modeling," IEEE J. Selected Area. Comm., vol. 4, no. 6, 2003, pp. 891-904.

[15] S. Maheshwari, P. Sharma, and M. Jain, "Unreliable flexible manufacturing cell with common cause failure," Int. J. Engg. Sci. Tech., vol. 2, no. 9, 2010, pp. 4071-4716.