International Journal of
Operations Research

International Journal of Operations Research Vol. 5, No. 4, 242–259 (2008)

Bulk Arrival Retrial Queue with Unreliable Server and Priority Subscribers

M. Jain[1] and Charu Bhargava[2,*]

[1,2]Department of Mathematics,Institute of Basic Science,Dr BR Ambedkar University,Agra 282002, India

Received June 2008; *Revised June 2008*; *Accepted October 2008*

Abstract—This paper is concerned with the analysis of unreliable server bulk arrival retrial queue with two class non-preemptive priority subscribers. The two types of subscribers arrive according to Poisson flow in which priority is assigned to class one, and class two subscribers are of non-priority type. The subscribers in each class arrive to the system in batches; the batch sizes follow the geometric process. If the server is free at the time of any batch arrivals, the subscriber of this batch begins to be served immediately and leave the system forever. The priority subscribers that find the server busy are queued and then are served in accordance with FCFS discipline. The arriving non-priority subscribers on finding the server busy cannot be queued and leave the service area and try their chance after some random time. If a subscriber is being served at the instant of the server failure, the service is interrupted and restarted after repair. The life time of the server is assumed to be exponentially distributed. The repair time and service time are also assumed to be i.i.d. general distributed. We obtain the condition of Ergodicity for such a queueing system. The analytical results for queue size distribution as well as some performance characteristics under steady state conditions by applying supplementary variable technique are derived. The waiting time distribution is also discussed for priority and non-priority subscribers. By taking illustration, computational results are provided to facilitate the sensitivity analysis.

Keywords—Bulk arrival, Retrial queue, Non-preemptive priority, Unreliable server, Supplementary variables, Queue size, Waiting time.

Biographical Notes—Madhu Jain is an Associate Professor of Mathematics in the Dr. BR Ambedkar University, Agra. She is a recipient of two Gold medals of the Agra University at MPhil level. There are more than 200 research publications in referred International/National journals and more than 20 books to her credit. She was a recipient of the Young Scientist Award and SERC visiting fellow the department of Science and Technology (DST), India and Career Award of University Grant Commission (UGC), India. She has successfully completed six sponsored major research project of DST, UGC and CSIR. Her current research interest includes the performance modeling, stochastic models, soft computing, bioinformatics, reliability, engineering and queueing theory.

Charu Bhargava did her MSc and MPhil in Mathematics from Dr. BR Ambedkar University, Agra. Currently she is pursuing research for the degree of Ph.D. in the department of Mathematics, Institute of Basic Science, Agra under the supervision of Dr. Madhu Jain. Her research areas of interest include probability theory, stochastic models and queueing theory. She has participated and presented her research papers in seven International/National conferences. She was a recipient of Young Scientist Award for best paper presentation in "9th Conference of International Academy of Physical Sciences" held at Agra during Feb. 3-5, 2007. Several research articles jointly with Dr. Madhu Jain are being published in QTQM, Computers and Operations Research, Journal of KING ABDULAZIZ UNIVERSITY Engineering Sciences and OPSEARCH.

1. INTRODUCTION

The FCFS discipline is a fair procedure in determining the order in which the subscribers are to be served in a service system. However, this is not the case in many real life situations where the jobs are classified according to different priorities. For this reason the priority queue has received considerable attention in the queueing literature. Queueing systems with two classes of subscribers have broad applications in the manufacturing and production systems, distribution and service systems, transportation systems, telecommunication industry, computer and communication systems, etc.. Two well-known priority disciplines in queueing literature are non-preemptive and preemptive discipline. Under non-preemptive discipline if non-priority (class two) subscriber enters service, it cannot be preempted by an arrival of a priority (class one) subscriber, as

* Corresponding author's email: charu_431981@yahoo.co.in madhujain@sancharnet.in

such the priority subscriber has to wait until the non-priority subscriber completes its service. This discipline is known as non-preemptive priority and was introduced by Cobham (1954).

In the existing literature, a few papers appear on bulk arrival priority queueing system. An early overview of some bulk priority queues can be found in Hawkes (1965). Takahashi and Takagi (1990) considered the joint time-dependent processes of the queue size and the elapsed service time in the non-preemptive and preemptive resume priority systems with structured priority batch arrivals. Langaris and Moutzoukis (1995) suggested batch arrival retrial queueing model with priority. Pekoz (2002) used linear programming approach to find an optimal policy for multi server non-preemptive priority queues. Non-preemptive priority queueing model with bulk arrivals was analyzed by Soo and Chung (2003). A queueing system with two-arrival stream was considered by Pla and Giner (2005), where they suggested a number of schemes assigning different priorities to each of the two arrival streams. One of the streams was considered to require a higher priority to access the server than the other. Hassin and Haviv (2006) considered a memoryless single server queue with two classes of customers. Discrete time preemptive priority queueing system was studied by Walraevens et al. (2008). Multiclass non Markovian queueing model with non-preemptive scheme was discussed by Iftikhar et al. (2008).

In computer and communication systems, the arriving subscribers to a busy system are forced to leave and they try back after a random amount of time. In literature such systems are referred as queue with retrials and have been extensively studied under a variety of scenarios for single server case. Falin et al. (1993) considered retrial queueing model with priority customers and obtained the joint distribution in terms of generating function of the queue length of priority and non-priority calls in steady state. Choi et al. (1999) discussed more complicated queueing situations with batch arrivals under retrial attempts. Corral (2002) has developed a model for single server retrial queue with quasi-random input and two priority classes. Dudin et al. (2004) considered single server retrial queueing model with batch Markovian arrival process and general distributed service time. Wu el al. (2005) discussed M/G/1 retrial queue with customer's discouragement. Atencia and Moreno (2006) have given discrete time retrial queueing model with optional service. Roszik et al. (2007) concerned with the performance analysis of finite-source retrial queues with heterogeneous sources operating in random environments. , Aguir et al. (2008) modeled a call centre as a continuous time Markov chain with retrial phenomenon.

In most queueing models, it is not possible to keep the server operational at all times, and service can thus be interrupted. The main reason for this is the breakdown of the server. There can also be scheduled service interruptions such as during weekends or holidays. White and Chrisite (1958) were the first who considered the breakdown events in priority queueing models. Sandhu and Posner (1989) discussed priority model for integrated voice/data transmission where the service medium is subject to breakdowns. Gurukajan and Srinivasan (1995) described a complex two-unit system in which the repair facility is subject to random breakdown. Almasi et al. (2005) studied single server retrial queue with a finite number of homogenous sources of calls and a single nonreliable server. Classical and constant retrial policies in M/G/1 queueing model with active breakdowns of the server were discussed by Atencia et al. (2006). M/G/1 retrial queueing system with unreliable server and setup was given by Jain et al. (2007). Markovian flow of breakdown in which all busy servers are subject to breakdown and repair was considered by Kim et al. (2008) for retrial queueing model with batch arrivals.

In many communication systems, a priority is given to certain classes jobs or calls to improve the grade of service (GoS) over other classes. In wireless communication system which delivers a wide variety of services, priority rule is often followed, e.g. being delay sensitive voice calls are required to have a higher priority than data calls. Our model can be easily implemented in a cellular mobile system; wherein there may be two types of attempts that customers make, one through messages (text or picture or multimedia etc.) and other through voice calls are considered as non priority and priority subscribers, respectively. On finding server busy in some (another call receiving or message reading) work or brokendown the messages are stored in a buffer, form a queue there and retry for service again and again till their service is completed successfully. This is referred as retrial queue in our model while the voice calls do not form any retrial queue and have to wait for their service in the normal queue.

In this investigation, we consider $M^X/G/1$ retrial queueing model with priority subscribers and unreliable server. The generating function method and supplementary variable technique are employed for the analysis of our model. The organization of the paper is as follows. The model is described along with requisite assumptions and notations in section 2. Ergodicity condition for stability of Embedded Markov chain has been obtained in section 3. The queue size distribution is established in section 4. In section 5, some performance characteristics are derived using queue size distribution. Section 6 contains the analysis of waiting time distribution. Computational results and sensitivity analysis are presented in section 7. At last, the conclusion has been drawn in section 8.

2. MODEL DESCRIPTION

For the mathematical formulation of the $M^X/G/1$ retrial queue with unreliable server and two classes of subscribers, we use the following notations:

$N_1(t)$, $N_2(t)$ Number of subscribers in priority (class one) and non-priority (class two); queue at time t

$\lambda_1(\lambda_2)$ Batch arrival rate for priority (non-priority) subscribers
θ Retrial rate for the repeated subscribers of class two

X_ℓ	Random variable denoting the batch size for the ℓ^{th} (ℓ = 1,2) class of subscribers
$C_{\ell,k}$	$\Pr(X_\ell = k)$, $\ell = 1,2$; $k \geq 1$
$C(z_\ell)$	Probability generating function of the batch size X_ℓ for the ℓ^{th} (ℓ = 1,2) class of subscribers
a_ℓ	Mean batch size for the ℓ^{th} (ℓ = 1,2) class of subscribers
$a_\ell^{(2)}$	Second factorial moment of batch size for the ℓ^{th} (ℓ = 1,2) class of subscribers
α	Failure rate of the server
$B_\ell(x)$	Service time distribution function for the ℓ^{th} (ℓ = 1,2) class of subscribers
$G_\ell(y)$	Repair time distribution function of the server when rendering service to the ℓ^{th} (ℓ = 1,2) class of subscribers
$b_\ell(x)$	Instantaneous service rate for the ℓ^{th} (ℓ = 1,2) class of subscribers
$g_\ell(y)$	Instantaneous repair rate of the server when rendering service to ℓ^{th} (ℓ = 1,2) class of subscribers
$b_\ell^*(.), g_\ell^*(.)$	Laplace-Stieltjes transform of $B_\ell(.)$ *and* $G_\ell(.)$
$\xi(t)$	Elapsed service time of the subscribers at time t
$\psi(t)$	Elapsed repair time of the server at time t
τ_d	Time instant of the dth departure
$N_{\ell,d} = N_\ell(\tau_d)$	Number of subscribers in the ℓ^{th} (ℓ = 1,2) class just before the time epoch τ_d
Y_d	Random variable denoting the type of the dth served subscriber
ξ_d, ψ_d	Elapsed service time and elapsed repair time of the dth served subscriber
$v_{\ell,d}$	Number of subscribers, who arrive in the system during the service time (which also includes the down time of the server) of the dth subscriber.
B_d	Number of subscribers, who enter for service from the non-priority queue at time ξ_d
$\beta_{\ell,r}, \gamma_{\ell,r}$	rth moment of $B_\ell(x)$ and $G_\ell(y)$ respectively about origin; $\beta_{\ell,r} = (-1)^r b_\ell^{*(r)}(0)$, $\gamma_{\ell,r} = (-1)^r g_\ell^{*(r)}(0)$
ρ_ℓ	Traffic intensity due to the ℓ^{th} (ℓ = 1,2) class of subscribers $\rho_\ell = \lambda_\ell a_\ell \beta_{\ell 1}$

Here $b_\ell(x) = \dfrac{B'_\ell(x)}{\overline{B}_\ell(x)}$, $g_\ell(y) = \dfrac{G'_\ell(y)}{\overline{G}_\ell(y)}$ $\left[\overline{B}_\ell(x) = 1 - B_\ell(x), \overline{G}_\ell(y) = 1 - G_\ell(y)\right]$

The state of the server at time t is given by

$$Y(t)=\begin{cases}0, & \textit{if the server is in idle state}\\ 1, & \textit{if the server is rendering service to priority subscriber}\\ 2, & \textit{if the server is rendering service to nonpriority subscriber}\\ 3, & \textit{if the server is broken down and under repair while rendering service to priority subscriber}\\ 4, & \textit{if the server is broken down and under repair while rendering service to nonpriority subscriber}\end{cases}$$

There is an interruption in the service due to server failure when rendering service to the ℓ^{th} (ℓ = 1,2) class of subscribers and after repair the characteristic of the server is same as before failure. The life time of the server is exponential distributed. When the server fails, it is repaired immediately. The repair time and service time are assumed to be independent and general distributed. The priority as well as non-priority subscribers arrive in batches with sizes X_1 and X_2, respectively such that $\Pr\{X_1 = k\} = C_{1k}$ and $\Pr\{X_2 = k\} = C_{2k}$, $k \geq 1$. If the server is free at the time of any group arrivals, the subscriber begins to be served immediately and leave the system forever after service completion. On arrival if priority subscriber finds the server busy or brokendown, he queues up and are served according to FCFS rule. The arriving non-priority subscribers on finding server busy or broken, leave the service area and move to a group of blocked subscribers called "orbit", these blocked subscribers are called repeated subscribers and try their chance for service after some random time until they find the server free. The retrial time of any subscriber is independent and exponentially distributed. The priority subscriber is always served before a non-priority subscriber. If a non-priority subscriber enters service, it cannot be preempted by an arrival of a priority subscriber, so that the priority subscriber has to wait until the non-priority subscriber completes its service.

3. ERGODICITY CONDITION

Embedded Markov Chain: The sequence of random vectors $X_d = \{Y_d, N_{1,d}, N_{2,d}\}$ forms a Markov chain with $\{1,2,3,4\}\times Z_+^2$ as state space, which is the Embedded Markov chain corresponding to concerned queueing system. It is easy to see that X_d is irreducible and aperiodic. Now

For priority subscribers: $N_{1,d} = N_{1,d-1} - 1 + \nu_{1,d}$ (1.1)

For non priority subscribers: $N_{2,d} = N_{2,d-1} - B_d + \nu_{2,d}$ (1.2)

where $B_d = 1$ if the d[th] departure is a repeated subscriber and $B_d = 0$ if the d[th] departure is a priority subscriber.

Ergodicity: For the ergodic condition a Markov chain should be irreducible, aperiodic and positive recurrent. Now to prove ergodicity (Gross and Harris;1985) we use Foster's criterion (cf. Pakes, 1969) due to recursive structure of the Eq. (1.1) and (1.2).

Foster's criterion: An irreducible and aperiodic Markov chain X_d with state space S is ergodic if there exist a nonnegative $f(s)$, $s \in S$ called test function and $\varepsilon > 0$ such that the mean drift $x_s \equiv E\{f(X_d) - f(X_{d-1}) / X_{d-1} = s\}$ is finite for all the states and $x_s \leq -\varepsilon$ for all S except for a finite number. [cf. statement 1, pp. 20, Falin and Templeton (1997)]

Theorem 1: *The Embedded Markov chain is ergodic iff* $\rho_1(1+\alpha\gamma_{11}) + \rho_2(1+\alpha\gamma_{21}) < 1$

Proof: In our case, we consider the following test function:

$$f(\ell,i,j) = [\lambda_2 a_2 \beta_{11}(1+\alpha\gamma_{11}) + 1 - \rho_2(1+\alpha\gamma_{21})]\,i + [\lambda_1 a_1 \beta_{21}(1+\alpha\gamma_{21}) + 1 - \rho_1(1+\alpha\gamma_{11})]\,j \tag{1.3}$$

The mean drift is given by

$$x_{\ell,i,j} \equiv E\{f(X_d) - f(X_{d-1}) / X_{d-1} = (\ell,i,j)\} \tag{1.4}$$

$$x_{\ell,i,j} = \begin{cases}\rho_1(1+\alpha\gamma_{11}) + \rho_2(1+\alpha\gamma_{21}) - 1, & i \geq 1\\ \rho_1(1+\alpha\gamma_{11}) + \rho_2(1+\alpha\gamma_{21}) - 1 + \dfrac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+j\theta}, & i = 0\end{cases} \tag{1.5}$$

[For detail see appendix A]

Let $\rho_1(1+\alpha\gamma_{11}) + \rho_2(1+\alpha\gamma_{21}) < 1$ then there exists a positive number $\in$ such that

$$\varepsilon = \frac{1-\rho_1(1+\alpha\gamma_{11})-\rho_2(1+\alpha\gamma_{21})}{2} \tag{1.6}$$

For all states (ℓ, i, j) with $i \geq 1$, we have

$$x_{\ell,i,j} = -2\varepsilon < -\varepsilon \tag{1.7}$$

For $i = 0$, we get

$$x_{\ell,i,j} = -2\varepsilon + \frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+j\theta} \leq -\varepsilon \tag{1.8}$$

So that $x_{\ell,i,j} \leq -\varepsilon$ for all states except for a finite number, hence the chain is ergodic.

This condition can also be evaluated by an alternate approach as used by Kernane and Aissani (2005) to obtain the stability condition of retrial queue with versatile retrial policy.

4. QUEUE SIZE DISTRIBUTION

Since service time and repair time distribution are not exponential, the process $\{Y(t), N_1(t), N_2(t)\}$ is not Markovian. In such a case we introduce supplementary variables corresponding to elapsed service time and elapsed repair time to make it Markovian. Joint distributions of the server state and queue size in steady state are defined as

$$P_{0,i,j} = \lim_{t\to\infty} \Pr\{Y(t)=0, N_1(t)=i, N_2(t)=j\},\ i \geq 0, j \geq 0$$

$$P_{1,i,j}(x)dx = \lim_{t\to\infty} \Pr\{Y(t)=1, x \leq \xi(t) \leq x+dx, N_1(t)=i, N_2(t)=j\},\ i \geq 0, j \geq 0$$

$$P_{2,i,j}(x)dx = \lim_{t\to\infty} \Pr\{Y(t)=2, x \leq \xi(t) \leq x+dx, N_1(t)=i, N_2(t)=j\},\ i \geq 0, j \geq 0$$

$$\mathrm{R}_{1,i,j}(x,y)dy = \lim_{t\to\infty} \Pr\{Y(t)=3, \xi(t)=x, y < \psi(t) \leq y+dy, N_1(t)=i, N_2(t)=j\},\ i \geq 0, j \geq 0$$

$$\mathrm{R}_{2,i,j}(x,y)dy = \lim_{t\to\infty} \Pr\{Y(t)=4, \xi(t)=x, y < \psi(t) \leq y+dy, N_1(t)=i, N_2(t)=j\},\ i \geq 0, j \geq 0$$

From the above-defined probabilities and using appropriate rates we can easily construct the following steady state equations:

$$(\lambda_1+\lambda_2+j\theta)P_{0,0,j} = \int_0^\infty [P_{1,0,j}(x)b_1(x) + P_{2,0,j}(x)b_2(x)]dx,\ j \geq 1 \tag{2.1}$$

$$\left[\frac{d}{dx}+\lambda_1+\lambda_2+b_1(x)+\alpha\right]P_{1,i,j}(x) = \int_0^\infty \mathrm{R}_{1,i,j}(x,y)\,g_1(y)\,dy + \sum_{k=1}^{i}\lambda_1 C_{1k}\,P_{1,i-k,j}(x) + \sum_{k=1}^{j}\lambda_2 C_{2k}\,P_{1,i,j-k}(x), i \geq 0, j \geq 0 \tag{2.2}$$

$$\left[\frac{d}{dx}+\lambda_1+\lambda_2+b_2(x)+\alpha\right]P_{2,i,j}(x) = \int_0^\infty \mathrm{R}_{2,i,j}(x,y)\,g_2(y)\,dy + \sum_{k=1}^{i}\lambda_1 C_{1k}\,P_{2,i-k,j}(x) + \sum_{k=1}^{j}\lambda_2 C_{2k}\,P_{2,i,j-k}(x), i \geq 0, j \geq 0 \tag{2.3}$$

$$\left[\frac{\partial}{\partial y}+\lambda_1+\lambda_2+g_1(y)\right]\mathrm{R}_{1,i,j}(x,y) = \sum_{k=1}^{i}\lambda_1 C_{1k}\,\mathrm{R}_{1,i-k,j}(x,y) + \sum_{k=1}^{j}\lambda_2 C_{2k}\,\mathrm{R}_{1,i,j-k}(x,y), i \geq 0, j \geq 0 \tag{2.4}$$

$$\left[\frac{\partial}{\partial y}+\lambda_1+\lambda_2+g_2(y)\right]\mathrm{R}_{2,i,j}(x,y) = \sum_{k=1}^{i}\lambda_1 C_{1k}\,\mathrm{R}_{2,i-k,j}(x,y) + \sum_{k=1}^{j}\lambda_2 C_{2k}\,\mathrm{R}_{2,i,j-k}(x,y), i \geq 0, j \geq 0 \tag{2.5}$$

These equations are to be solved under the following boundary conditions at x=0, y=0

$$P_{1,i,j}(0)=\sum_{k=1}^{i+1}\lambda_1 C_{1k}P_{0,0,i-k+1}\delta_{i,0}+\int_0^\infty P_{1,i+1,j}(x)b_1(x)dx+\int_0^\infty P_{2,i+1,j}(x)b_2(x)dx \tag{2.6}$$

$$P_{2,i,j}(0)=\begin{cases}0, & if \quad i\geq 1\\ \sum_{k=1}^{j+1}\lambda_2 C_{2,k}P_{0,0,j-k+1}+(j+1)\theta P_{0,0,j+1}, & if \quad i=0\end{cases} \tag{2.7}$$

$$R_{1,i,j}(x,0)=\alpha P_{1,i,j}(x) \tag{2.8}$$

$$R_{2,i,j}(x,0)=\alpha P_{2,i,j}(x) \tag{2.9}$$

The normalizing condition is given by

$$\sum_{i=0}^\infty\sum_{j=0}^\infty\left[P_{0,0,j}+\int_0^\infty P_{1,i,j}(x)\,dx+\int_0^\infty P_{2,i,j}(x)\,dx+\int_0^\infty\int_0^\infty R_{1,i,j}(x,y)\,dx\,dy+\int_0^\infty\int_0^\infty R_{2,i,j}(x,y)\,dx\,dy\right]=1 \tag{3}$$

Define the generating functions:

$$P_0(z_1,z_2)=\sum_{i=0}^\infty\sum_{j=0}^\infty P_{0,0,j}\,z_1^i\,z_2^j\ ,\quad P_1(z_1,z_2,x)=\sum_{i=0}^\infty\sum_{j=0}^\infty P_{1,i,j}(x)\,z_1^i\,z_2^j\ ,\quad P_2(z_1,z_2,x)=\sum_{i=0}^\infty\sum_{j=0}^\infty P_{2,i,j}(x)\,z_1^i\,z_2^j\ ,$$

$$R_1(z_1,z_2,x,y)=\sum_{i=0}^\infty\sum_{j=0}^\infty R_{1,i,j}(x,y)\,z_1^i\,z_2^j\ ,\quad R_2(z_1,z_2,x,y)=\sum_{i=0}^\infty\sum_{j=0}^\infty R_{2,i,j}(x,y)\,z_1^i\,z_2^j\ ,$$

$$C_1(z_1)=\sum_{k=1}^\infty C_{1k}\,z_1^k\ ,\qquad C_2(z_2)=\sum_{k=1}^\infty C_{2k}\,z_2^k$$

Using generating functions, Eq. (2.1)-(2.9) become

$$(\lambda_1+\lambda_2)P_0(z_1,z_2)+z_2\theta P_0'(z_1,z_2)=\int_0^\infty[P_1(0,z_2,x)b_1(x)+P_2(0,z_2,x)b_2(x)]\,dx \tag{4.1}$$

$$\left[\frac{\partial}{\partial x}+\lambda_1-\lambda_1C_1(z_1)+\lambda_2-\lambda_2C_2(z_2)+b_1(x)+\alpha\right]P_1(z_1,z_2,x)=\int_0^\infty R_1(z_1,z_2,x,y)\,g_1(y)\,dy \tag{4.2}$$

$$\left[\frac{\partial}{\partial x}+\lambda_1-\lambda_1C_1(z_1)+\lambda_2-\lambda_2C_2(z_2)+b_2(x)+\alpha\right]P_2(z_1,z_2,x)=\int_0^\infty R_2(z_1,z_2,x,y)\,g_2(y)\,dy \tag{4.3}$$

$$\left[\frac{\partial}{\partial y}+\lambda_1-\lambda_1C_1(z_1)+\lambda_2-\lambda_2C_2(z_2)+g_1(y)\right]R_1(z_1,z_2,x,y)=0 \tag{4.4}$$

$$\left[\frac{\partial}{\partial y}+\lambda_1-\lambda_1C_1(z_1)+\lambda_2-\lambda_2C_2(z_2)+g_2(y)\right]R_2(z_1,z_2,x,y)=0 \tag{4.5}$$

$$z_1P_1(z_1,z_2,0)=\lambda_1C_1(z_1)+\int_0^\infty[P_1(z_1,z_2,x)-P_1(0,z_2,x)]\,b_1(x)\,dx+\int_0^\infty[P_2(z_1,z_2,x)-P_2(0,z_2,x)]\,b_2(x)\,dx \tag{4.6}$$

$$P_2(z_1,z_2,0)=\frac{\lambda_2C_2(z_2)P_0(z_1,z_2)}{z_2}+\theta\frac{dP_0(z_2)}{z_2} \tag{4.7}$$

$$R_1(z_1,z_2,x,0)=\alpha P_1(z_1,z_2,x) \tag{4.8}$$

$$R_2(z_1, z_2, x, 0) = \alpha P_2(z_1, z_2, x) \tag{4.9}$$

$$\lim_{z_1, z_2 \to 1} \left[P_0(z_2) + \int_0^\infty P_1(z_1, z_2, x)\,dx + \int_0^\infty P_2(z_1, z_2, x)\,dx + \int_0^\infty\int_0^\infty R_1(z_1, z_2, x, y)\,dx\,dy + \int_0^\infty\int_0^\infty R_2(z_1, z_2, x, y)\,dx\,dy \right] = 1 \tag{5}$$

In order to obtain probability generating function, we denote

$$z_1 = h(z_2)$$

$$k_i(z_1, z_2) = b_i^*\{(\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2)) + \alpha \overline{g}_i^*(\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2))\}$$

Here $k_i(z_1,z_2)$ denotes Laplace Stieltjes transform for the completion time; the time interval from when the server begins to serve the arbitrary customer until the customer's service ends, which includes the repair time. Further, $h(z_2)$ is defined as the generating function of the number of class two jobs that arrive during the busy period formed by class one jobs.

Theorem 2: The probability generating functions for the idle state, busy state and under repair state of the server are given as

$$P_0(z_2) = [1 - \rho_1(1 + \alpha\gamma_{11}) - \rho_2(1 + \alpha\gamma_{21})] \exp\left\{ \frac{1}{\theta} \int_1^{z_2} \frac{[\lambda_1 + \lambda_2 - \lambda_1 C_1(h(u)) - \frac{\lambda_2 C_2(u) k_2(h(u), u)}{u}]}{[k_2(h(u), u) - u]} du \right\} \tag{6.1}$$

$$\begin{aligned} P_1(z_1, z_2, x) = & \left\{ \left(\lambda_1 - \lambda_1 C_1(h(z_2)) + \lambda_2 - \frac{\lambda_2 C_2(z_2) k_2(h(z_2), z_2)}{z_2} \right) \left(k_2(z_1, z_2) - z_2 \right) \right\} \\ & - \left\{ \left(\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \frac{\lambda_2 C_2(z_2) k_2(z_1, z_2)}{z_2} \right) \left(k_2(h(z_2), z_2) - z_2 \right) \right\} \\ & \times \exp\{-[\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2) + \alpha \overline{g}_1^*(\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2))]x\} \overline{B}_1(x) \\ & \times \left\{ \left(k_2(h(z_2), z_2) - z_2 \right) \left(z_1 - k_1(z_1, z_2) \right) \right\}^{-1} P_0(z_2) \end{aligned} \tag{6.2}$$

$$\begin{aligned} P_2(z_1, z_2, x) = & \frac{\lambda_1 + \lambda_2 - \lambda_1 C_1(h(z_2)) - \lambda_2 C_2(z_2)}{[k_2(h(z_2), z_2] - z_2} P_0(z_2) \\ & \times \exp\{-[\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2) + \alpha \overline{g}_2^*(\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2))]x\} \overline{B}_2(x) \end{aligned} \tag{6.3}$$

$$R_1(z_1, z_2, x, y) = \alpha P_1(z_1, z_2, x) \exp\{-[\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2)]y\} \overline{G}_1(y) \tag{6.4}$$

$$R_2(z_1, z_2, x, y) = \alpha P_2(z_1, z_2, x) \exp\{-[\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2)]y\} \overline{G}_2(y) \tag{6.5}$$

Proof: For proof see appendix B.

Theorem 3: Queue size distribution under steady state is $(z_1, z_2) = P_0(z_2) + P_1(z_1, z_2) + P_2(z_1, z_2) + R_1(z_1, z_2) + R(z_1, z_2)$ where

$$\begin{aligned} P_1(z_1, z_2) = & \left\{ \left(\lambda_1 - \lambda_1 C_1(h(z_2)) + \lambda_2 - \frac{\lambda_2 C_2(z_2) k_2(h(z_2), z_2)}{z_2} \right) \left(k_2(z_1, z_2) - z_2 \right) \right\} \\ & - \left\{ \left(\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \frac{\lambda_2 C_2(z_2) k_2(z_1, z_2)}{z_2} \right) \left(k_2(h(z_2), z_2) - z_2 \right) \right\} \\ & \times \frac{1 - k_1(z_1, z_2)}{\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2) + \alpha \overline{g}_1^*(\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2))} \\ & \times \left\{ \left(k_2(h(z_2), z_2) - z_2 \right) \left(z_1 - k_1(z_1, z_2) \right) \right\}^{-1} P_0(z_2) \end{aligned} \tag{7.1}$$

$$P_2(z_1,z_2)=\frac{\lambda_1+\lambda_2-\lambda_1C_1(h(z_2))-\lambda_2C_2(z_2)}{[k_2(h(z_2),z_2]-z_2}P_0(z_2)\times\frac{1-k_2(z_1,z_2)}{\lambda_1-\lambda_1C_1(z_1)+\lambda_2-\lambda_2C_2(z_2)+\alpha\overline{g}_2^*(\lambda_1-\lambda_1C_1(z_1)+\lambda_2-\lambda_2C_2(z_2))} \tag{7.2}$$

$$R_1(z_1,z_2)=\alpha P_1(z_1,z_2)\frac{1-g_1^*(\lambda_1-\lambda_1C_1(z_1)+\lambda_2-\lambda_2C_2(z_2))}{\lambda_1-\lambda_1C_1(z_1)+\lambda_2-\lambda_2C_2(z_2)} \tag{7.3}$$

$$R_2(z_1,z_2)=\alpha P_2(z_1,z_2)\frac{1-g_2^*(\lambda_1-\lambda_1C_1(z_1)+\lambda_2-\lambda_2C_2(z_2))}{\lambda_1-\lambda_1C_1(z_1)+\lambda_2-\lambda_2C_2(z_2)} \tag{7.4}$$

Proof: For proof see appendix C.

5. PERFORMANCE CHARECTERISTICS

In this section, we derive the expressions for some performance characteristics to predict the behaviour of system as follows.

(I) The long run probabilities of the server states:

- The long run probability that the server being idle is

$$P(I)=1-\rho_1(1+\alpha\gamma_{11})-\rho_2(1+\alpha\gamma_{21}) \tag{8.1}$$

- The long run probability that the server is rendering service to a priority subscriber is

$$P(B_1)=\rho_1 \tag{8.2}$$

- The long run probability that the server is rendering service to a non-priority subscriber is

$$P(B_2)=\rho_2 \tag{8.3}$$

- The long run probability that the server is broken down and under repair while rendering service to the priority subscriber is

$$P(R_1)=\rho_1\alpha\gamma_{11} \tag{8.4}$$

- The long run probability that the server is broken down and under repair while rendering service to the non-priority subscriber is

$$P(R_2)=\rho_2\alpha\gamma_{21} \tag{8.5}$$

(II) Average queue length

- Expected number of subscribers in the priority queue is

$$E(N_1)=\frac{\lambda_2a_2\lambda_1a_1\beta_{22}(1+\alpha\gamma_{21})^2+\lambda_1\beta_{11}(a_1^{(2)}-a_1)(1+\alpha\gamma_{11})+\beta_{12}(\lambda_1a_1(1+\alpha\gamma_{11}))^2+\alpha\lambda_1a_1(\rho_2\gamma_{22}+\rho_1\gamma_{12})}{2[1-\rho_1(1+\alpha\gamma_{11})]} \tag{8.6}$$

- Expected number of subscribers in the non-priority queue is

$$\begin{aligned}E(N_2)=&\frac{[\lambda_2\rho_1(1+\alpha\gamma_{11})+\lambda_2\rho_2(1+\alpha\gamma_{21})+\lambda_2a_2-\lambda_2]}{\theta[1-\rho_1(1+\alpha\gamma_{11})-\rho_2(1+\alpha\gamma_{21})]}+\frac{1}{2[1-\rho_1(1+\alpha\gamma_{11})][1-\rho_1(1+\alpha\gamma_{11})-\rho_2(1+\alpha\gamma_{21})]}\\&\times[\lambda_1a_1\lambda_2a_2\,\beta_{12}\,(1+\alpha\gamma_{11})^2+(\lambda_2a_2)^2\beta_{22}\,(1+\alpha\gamma_{21})^2+2\alpha\,\rho_1\,\rho_2\,\gamma_{22}\,\lambda_2a_2(1+\alpha\gamma_{11})\\&+\lambda_1(a_1^{(2)}-a_1)\lambda_2a_2\,\beta_{11}^2(1+\alpha\gamma_{11})^2+\lambda_2\,\beta_{21}\,(a_2^{(2)}-a_2)(1+\alpha\gamma_{21})\{1-\rho_1(1+\alpha\gamma_{11})\}+\alpha\,\lambda_2\,\rho_2\,\gamma_{22}\\&-\alpha\,\lambda_2a_2\gamma_{22}\,\rho_1^2\,\rho_2\,(1+\alpha\gamma_{11})^2-\alpha\,\lambda_2a_2\,\gamma_{22}\,\rho_1\,\rho_2^2\,(1+\alpha\gamma_{11})(1+\alpha\gamma_{21})-\alpha\,\lambda_2a_2\,\gamma_{12}\,\rho_1]\end{aligned} \tag{8.7}$$

(III) Expected Waiting Time

Using Little formula, we obtain the expected waiting time for priority and non- priority subscribers as

$$E(W_1)=\frac{E(N_1)}{\lambda_1 a_1} \text{ and } E(W_2)=\frac{E(N_2)}{\lambda_2 a_2} \tag{8.8}$$

6. WAITING TIME DISTRIBUTION

In this section, we derive LST of waiting time for an arbitrary subscriber in the priority and non-priority queue.

Theorem 4: Waiting time distribution in the queue for an arbitrary priority subscriber is given by

$$W_{q_1}^{*}(s)=W_{q_1}^{*+}(s)\frac{1-C_1[b_1^{*}\{s+\alpha\overline{g}_1^{*}(s)\}]}{a_1[1-b_1^{*}\{s+\alpha\overline{g}_1^{*}(s)\}]} \tag{9.1}$$

where $W_{q_1}^{*+}(s)$ is the Laplace Stieltjes transform of the distribution of queueing time for first subscriber of priority group (+ notation is used for 'immediately after a departure of a subscriber').

Proof: Stationary distribution shows the long run behaviour of embedded Markov chain, which is defined as $\pi_{\ell,i,j}=\lim_{d\to\infty}P(Y_d=\ell,N_{1,d}=i,N_{2,d}=j)$.

Let the joint generating function for the stationary distribution of the embedded Markov chain is defined as $\Psi(z_1,z_2)=\Psi_1(z_1,z_2)+\Psi_2(z_1,z_2)$ where $\Psi_1(z_1,z_2)$ and $\Psi_2(z_1,z_2)$ are the generating functions corresponding to busy state of the server for priority and non-priority subscribers, respectively. The generating function corresponding to idle state of the server is denoted by $\chi(z_2)$.

Following Falin and Templeton (1997), we get

$$z_1\Psi_1(z_1,z_2)=\mathrm{k}_1(\mathrm{z}_1,\mathrm{z}_2)[\lambda_1\mathrm{C}_1(\mathrm{z}_1)\,\chi(\mathrm{z}_2)+\Psi(\mathrm{z}_1,\mathrm{z}_2)-\Psi(0,\mathrm{z}_2)] \tag{9.2}$$

$$\Psi_2(z_1,z_2)=k_2(z_1,z_2)\left[\frac{\lambda_2C_2(z_2)}{z_2}\chi(z_2)+\theta\frac{d\chi(z_2)}{dz_2}\right] \tag{9.3}$$

$$\begin{aligned}\Psi_1(z_1,z_2)=&\left\{\left(\lambda_1-\lambda_1C_1(b(z_2))+\lambda_2-\frac{\lambda_2C_2(z_2)k_2(b(z_2),z_2)}{z_2}\right)\left(k_2(z_1,z_2)-z_2\right)\right\}\\&-\left\{\left(\lambda_1-\lambda_1C_1(z_1)+\lambda_2-\frac{\lambda_2C_2(z_2)k_2(z_1,z_2)}{z_2}\right)\left(k_2(b(z_2),z_2)-z_2\right)\right\}k_1(z_1,z_2)\chi(z_2)\\&\times\left\{\left(k_2(b(z_2),z_2)-z_2\right)\left(z_1-k_1(z_1,z_2)\right)\right\}^{-1}\end{aligned} \tag{9.4}$$

$$\Psi_2(z_1,z_2)=\frac{\lambda_1+\lambda_2-\lambda_1C_1(b(z_2))-\lambda_2C_2(z_2)}{[k_2(b(z_2),z_2]-z_2}k_2(z_1,z_2)\chi(z_2) \tag{9.5}$$

$$\chi(z_2)=\frac{1-\rho_1(1+\alpha\gamma_{11})-\rho_2(1+\alpha\gamma_{21})}{\lambda_1a_1+\lambda_2a_2}\exp\left\{\frac{1}{\theta}\int_1^{z_2}\frac{[\lambda_1+\lambda_2-\lambda_1C_1(b(u))-\frac{\lambda_2C_2(z_2)k_2(b(u),u)}{u}]}{[k_2(b(u),u)-u]}du\right\} \tag{9.6}$$

We identify a priority batch with a single priority subscriber, and then as its service time is the total service time of the subscribers constituting the batch, the Laplace transform of this service time is denoted by $C_1(b_1^{*}(.))=b_1^{*+}(.)$. Let us denote $N_1^{(n)}$ as the number of priority subscribers in the system at the time of n[th] departure. Waiting time of the first subscriber in his batch and completion time (i.e. service time including repair time) of the subscribers ahead of him in the batch, are independent to each other. It implies that

$$E\{z^{N_1(n)}\}=W_{q_1}^{*+}(\lambda_1-\lambda_1z_1)\,b_1^{*+}\{\lambda_1-\lambda_1z_1+\alpha\overline{g}_1^{*}(\lambda_1-\lambda_1z_1)\} \tag{9.7}$$

where $b_1^{*+}\{(\lambda_1-\lambda_1 z_1)+\alpha\overline{g}_1^*(\lambda_1-\lambda_1 z_1)\}$ is the Laplace Stieltjes transform of the distribution of completion time for the required subscriber.

Now substituting $z_1=z$, $z_2=1$ and replacing $C_1(z)$ by z and $C_1(b_1^*(.))$ *by* $b_1^{*+}(.)$ in Eq. (9.4), we get

$$\begin{aligned}\Psi_1(z,1)=&\left[b_1^{*+}\left\{\lambda_1-\lambda_1 z+\alpha\overline{g}_1^*(\lambda_1-\lambda_1 z)\right\}\right]\left\{\left[\lambda_2 a_2(1-b_2^*\{\lambda_1-\lambda_1 z+\alpha\overline{g}_2^*(\lambda_1-\lambda_1 z)\})\right]\right\}\\&+\left[(\lambda_1-\lambda_1 z)\{1-\rho_1(1+\alpha\gamma_{11})-\rho_2(1+\alpha\gamma_{21})\}\right]\times\left\{[b_1^{*+}\{\lambda_1-\lambda_1 z+\alpha\overline{g}_1^*(\lambda_1-\lambda_1 z)\}-z][\lambda_1 a_1+\lambda_2 a_2]\right\}^{-1}\end{aligned}\tag{9.8}$$

$$\Psi_1(1,1)=\frac{\lambda_1}{\lambda_1 a_1+\lambda_2 a_2}\tag{9.9}$$

Under steady state $E\{z^{N_1(n)}\}=\dfrac{\Psi_1(z,1)}{\Psi_1(1,1)}$ (9.10)

Replacing $\lambda_1(1-z_1)$ by 's', Eq. (9.7) and (9.10) yield

$$W_{q_1}^{*+}(s)=\frac{\lambda_2 a_2[1-b_2^*\{s+\alpha\overline{g}_2^*(s)\}]+s\{1-\rho_1(1+\alpha\gamma_{11})-\rho_2(1+\alpha\gamma_{21})\}}{s-\lambda_1+\lambda_1 C_1[b_1^*\{s+\alpha\overline{g}_1^*(s)\}]}\tag{9.11}$$

Laplace Stieltjes transform of the distribution of queueing time for an arbitrary subscriber of priority group (cf. Takagi,1991) is given by

$$W_{q_1}^*(s)=W_{q_1}^{*+}(s)\frac{1-C_1[b_1^*\{s+\alpha\overline{g}_1^*(s)\}]}{a_1[1-b_1^*\{s+\alpha\overline{g}_1^*(s)\}]}\tag{9.12}$$

Now $-\left.\dfrac{dW_{q_1}^*(s)}{ds}\right|_{s=0}$ gives the result, which coincides with that of Eq. (8.8) of section 5.

Theorem 5: Waiting time distribution in the queue for an arbitrary non-priority subscriber is given by

$$\mathbf{E}e^{-s\mathbf{T}_{mn}}=\int_1^{g(s)}\frac{\{h(s,u)\}^m u^{n-1}}{\theta[k_2(s,h(s,u),u)-u]}\times\exp\left\{\frac{1}{\theta}\int_u^1\frac{[\theta+s+\lambda_1+\lambda_2-\lambda_1 C_1(h(s,v))-\dfrac{\lambda_2 C_2(s,v)k_2(h(s,v),v)}{u}]}{[k_2(s,h(s,v),v)-v]}dv\right\}du\tag{10}$$

where g(s) is the root with the smallest absolute value of the equation $z=k_2\left(s,h\left(s,z\right),z\right)$.

Proof: In case of non-priority subscribers, we randomly select a tag subscriber from non-priority group and its waiting time is denoted by $T_{m\,n}$, when there are m priority subscribers and $n\geq 1$ non-priority subscribers in the system at the moment of departure of tagged subscriber.

Now $E_{\ell,i,j}^{(d)}(s)$ defines that d^{th} served subscriber of class ℓ and at the time of the d^{th} departure, there are i priority and $j\geq 1$ non-priority subscribers in the system that also involves the tagged subscriber. The analysis can be done in the similar manner as described in Falin and Templeton (1997); for detail see appendix D.

7. COMPUTATIONAL RESULTS

To illustrate the analytical results derived in the earlier sections, we provide numerical results for which program is coded in software MATLAB using Pentium IV. The effect of variation of various parameters on various performance indices are displayed in Tables 1-3 and Figures 1-4. The service time and repair time are general distributed therefore second moment of service time for the ℓ^{th} (ℓ = 1,2) class of subscribers and repair time of the server while rendering service to the ℓ^{th} (ℓ = 1,2) class of subscribers for different distributions are taken as

(i) M/E$_k$(E$_k$)/1 model : $\beta_{\ell,2}=\dfrac{(k+1)}{k\mu_\ell^2}$, $\gamma_{\ell,2}=\dfrac{(k+1)}{k\gamma_\ell^2}$ (11.1)

(ii) M/D(D)/1 model : $\beta_{\ell,2}=\dfrac{1}{\mu_\ell^2}$, $\gamma_{\ell,2}=\dfrac{1}{\gamma_\ell^2}$ (taking $k\to\infty$, in Eq. (11.1)) (11.2)

(iii) M/M(M)/1 model : $\beta_{\ell,2}=\dfrac{2}{\mu_\ell^2}$, $\gamma_{\ell,2}=\dfrac{2}{\gamma_\ell^2}$ (taking $k=1$, in Eq. (11.1)) (11.3)

(iv) M/E$_5$(E$_5$)/1 model : $\beta_{\ell,2}=\dfrac{6}{5\mu_\ell^2}$, $\gamma_{\ell,2}=\dfrac{6}{5\gamma_\ell^2}$ (taking $k=5$, in Eq. (11.1)) (11.4)

Here μ_ℓ is the service rate for the ℓ^{th} (ℓ = 1,2) class of subscribers and k denotes the number of phases in service time and repair time. γ_ℓ is the repair rate of the server when rendering service to the ℓ^{th} (ℓ = 1,2) class of subscribers.

The impact of mean batch size on various performance indices is examined by assuming that the batch size follows a geometrical distribution with parameter p=.2; so that the mean batch size and second factorial moment of batch size are given as a_i=q/p and $a_i^{(2)}=2q^2/p^2$, respectively, where q=1-p. The trends for the average queue length for priority and non priority subscribers have been shown by continuous and discrete lines, respectively in figs 1 and 2 for M/D/1, M/M/1 and M/E$_5$/1 models by varying arrival rates (λ_1,λ_2), mean batch sizes (a_1,a_2) of priority as well as non priority subscribers, failure rate (α) of the server and retrial rate (θ) of non priority subscribers. The results for the expected waiting time have been shown through bar graphs in Figures 3 and 4 by varying number of phases (k) in service time and repair time distributions for different sets of λ_1, λ_2, a_1, a_2, α and θ. For both tables and graphs, we choose the default parameters as $\lambda_1=\lambda_2=2$, $\alpha=1$, $\theta=3$, $\mu_1=\mu_2=3$, $\gamma_1=\gamma_2=5$.

Table 1. Effect of service rate (μ_1) on the long run probabilities of server states.

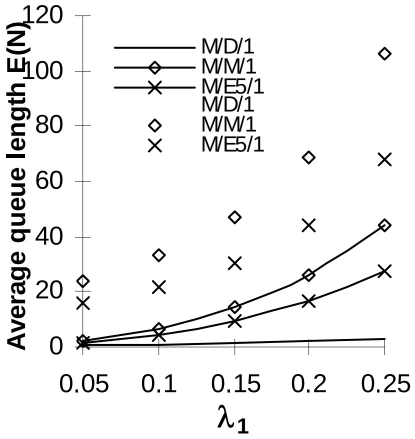
μ_1	**P(I)**		**P(B$_1$)**		**P(B$_2$)**		**P(R$_1$)**		**P(R$_2$)**	
	$(\lambda_1,\lambda_2)=(.1,.2)$	$(\lambda_1,\lambda_2)=(.3,.1)$	$(\lambda_1,\lambda_2)=(.1,.2)$	$(\lambda_1,\lambda_2)=(.3,.1)$	$(\lambda_1,\lambda_2)=(.1,.2)$	$(\lambda_1,\lambda_2)=(.3,.1)$	$(\lambda_1,\lambda_2)=(.1,.2)$	$(\lambda_1,\lambda_2)=(.3,.1)$	$(\lambda_1,\lambda_2)=(.1,.2)$	$(\lambda_1,\lambda_2)=(.3,.1)$
2	0.440	0.120	0.200	0.600	0.267	0.133	0.040	0.120	0.053	0.027
4	0.560	0.480	0.100	0.300	0.267	0.133	0.020	0.060	0.053	0.027
6	0.600	0.600	0.067	0.200	0.267	0.133	0.013	0.040	0.053	0.027
8	0.620	0.660	0.050	0.150	0.267	0.133	0.010	0.030	0.053	0.027
10	0.632	0.696	0.040	0.120	0.267	0.133	0.008	0.024	0.053	0.027

Table 2. Effect of service rate (μ_2) on the long run probabilities of server states.

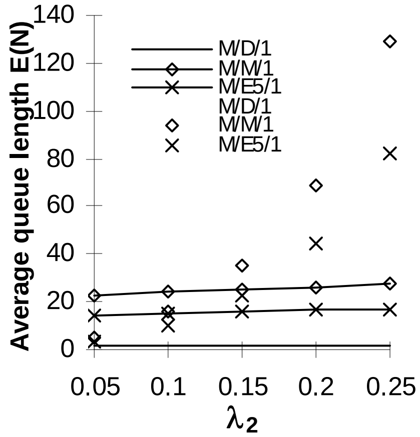
μ_2	**P(I)**		**P(B$_1$)**		**P(B$_2$)**		**P(R$_1$)**		**P(R$_2$)**	
	$(\lambda_1,\lambda_2)=(.1,.2)$	$(\lambda_1,\lambda_2)=(.3,.1)$	$(\lambda_1,\lambda_2)=(.1,.2)$	$(\lambda_1,\lambda_2)=(.3,.1)$	$(\lambda_1,\lambda_2)=(.1,.2)$	$(\lambda_1,\lambda_2)=(.3,.1)$	$(\lambda_1,\lambda_2)=(.1,.2)$	$(\lambda_1,\lambda_2)=(.3,.1)$	$(\lambda_1,\lambda_2)=(.1,.2)$	$(\lambda_1,\lambda_2)=(.3,.1)$
2	0.360	0.280	0.133	0.400	0.400	0.200	0.027	0.080	0.080	0.040
4	0.600	0.400	0.133	0.400	0.200	0.100	0.027	0.080	0.040	0.020
6	0.680	0.440	0.133	0.400	0.133	0.067	0.027	0.080	0.027	0.013
8	0.720	0.460	0.133	0.400	0.100	0.050	0.027	0.080	0.020	0.010
10	0.744	0.472	0.133	0.400	0.080	0.040	0.027	0.080	0.016	0.008

Table 3. Effect of failure rate (α) on the long run probabilities of server states.

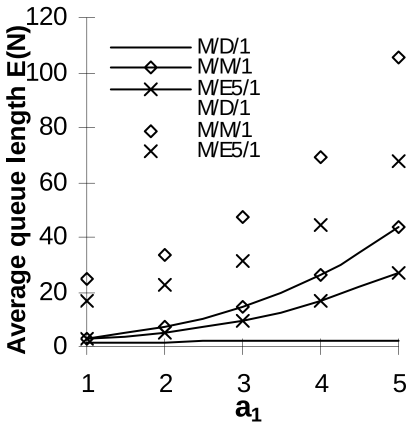
α	**P(I)**		**P(B_1)**		**P(B_2)**		**P(R_1)**		**P(R_2)**	
	$(a_1,a_2)=(1,3)$	$(a_1,a_2)=(4,2)$	$(a_1,a_2)=(4,2)$	$(a_1,a_2)=(4,2)$	$(a_1,a_2)=(4,2)$	$(a_1,a_2)=(4,2)$	$(a_1,a_2)=(4,2)$	$(a_1,a_2)=(4,2)$	$(a_1,a_2)=(4,2)$	$(a_1,a_2)=(4,2)$
0	0.733	0.600	0.067	0.267	0.200	0.133	0.000	0.000	0.000	0.000
1	0.680	0.520	0.067	0.267	0.200	0.133	0.013	0.053	0.040	0.027
2	0.627	0.440	0.067	0.267	0.200	0.133	0.027	0.107	0.080	0.053
3	0.573	0.360	0.067	0.267	0.200	0.133	0.040	0.160	0.120	0.080
4	0.520	0.280	0.067	0.267	0.200	0.133	0.053	0.213	0.160	0.107
5	0.467	0.200	0.067	0.267	0.200	0.133	0.067	0.267	0.200	0.133



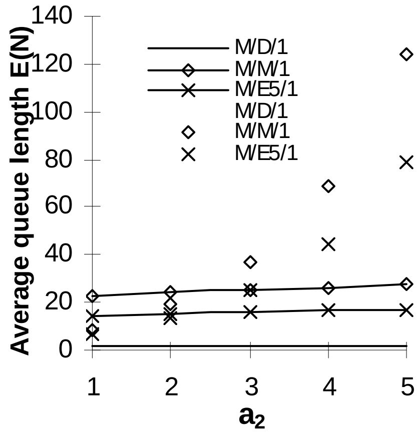
(a)



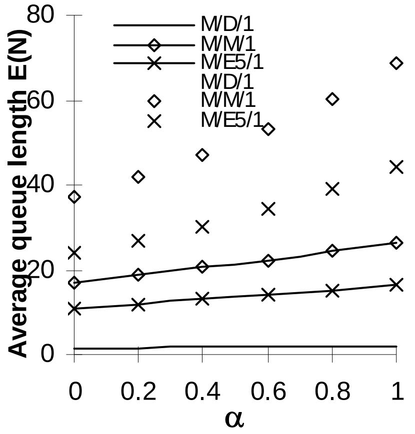
(a)



(b)

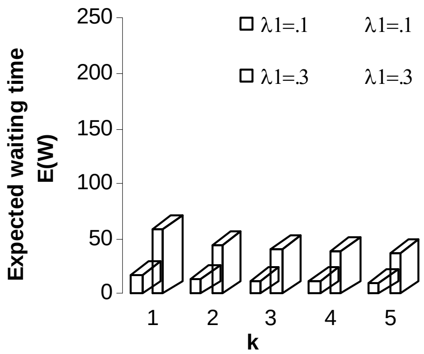


(b)

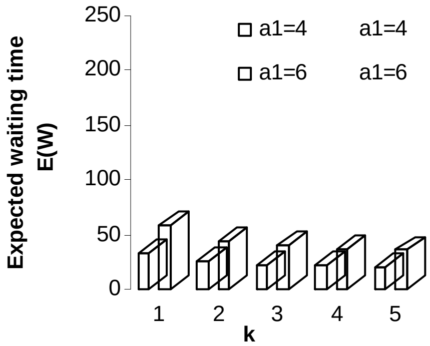


(c)

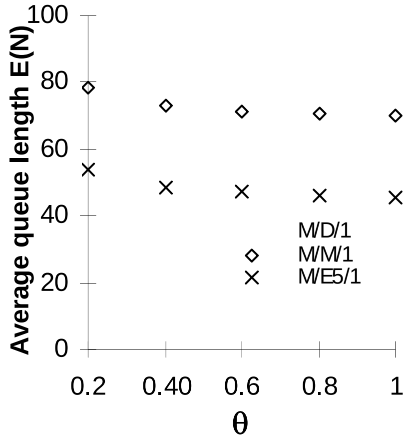
Figure 1. Effect of (a) λ_1 (b) a_1 (c) α on E(N) for various service and repair time distributions



(a)

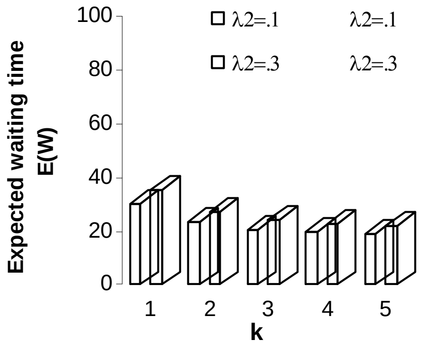


(b)

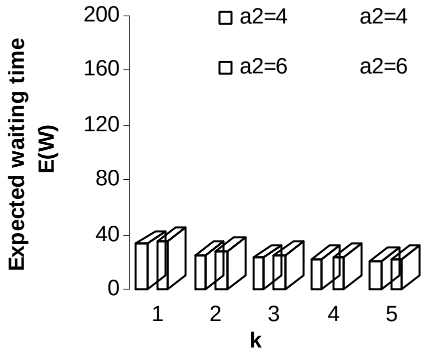


(c)

Figure 2.Effect of (a) λ_2 (b) a_2 (c) θ on E(N) for various service and repair time distributions



(a)



(b)

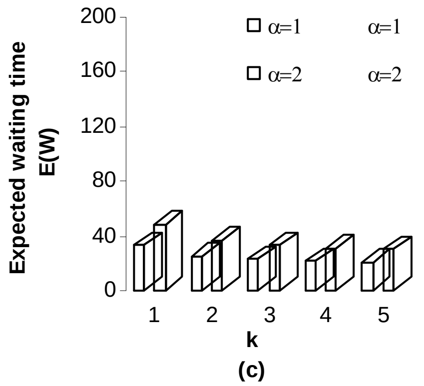


Figure 3. Effect of different sets of (a) λ_1 (b) a_1 (c) α on E(W) by varying k

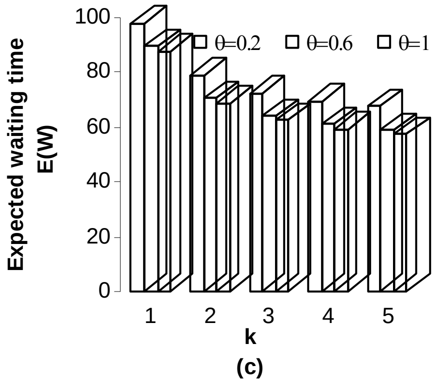


Figure 4. Effect of different sets of (a) λ_2 (b) a_2 (c) θ on E(W) by varying k

Tables 1 and 2 demonstrate the effect of service rates μ_1 and μ_2 respectively on the long run probabilities of server states for different sets of arrival rates (λ_1, λ_2) of priority and non priority subscribers. From Table 1, we observe that as μ_1 increases, P(I) increases, P(B_1) & P(R_1) decrease while P(B_2) and P(R_2) remain almost constant. Table 2 demonstrates that P(I) increases, P(B_1) and P(R_1) remain constant whereas at the same time P(B_2) and P(R_2) decrease with the increase in μ_2.The variation of the failure rate (α) of the server on the long run probabilities of the server states for different sets of mean batch sizes (a_1, a_2) has been summarized in Table 3. On increasing α, we see that P(I) decreases, P(B_1) and P(B_2) remain almost constant whereas P(R_1) and P(R_2) increase.

In Figures 1(a) and 1(b), we examine the effect of arrival rate (λ_1) and mean batch size (a_1) of priority subscribers on the average queue length E(N), respectively. Initially E(N) for both priority and non-priority subscribers increases gradually and then after increases sharply. Figures 2(a) and 2(b) reveal almost linearly increment in case of average queue length E(N_1) for priority subscribers as we increase λ_2 and a_2, respectively; however there is a sharp increment in the average queue length E(N_2) for non priority subscribers. Figure. 1(c) exhibits the trend for average queue length E(N) by varying failure rate (α) of the server. We notice that there is prominent increasing trend of E(N) with the increase in α. Figure. 2(c) illustrates the effect of retrial rate (θ) on the average queue length E(N_2) for non priority subscribers; initially decreasing trend is quite visible with the increase in θ, however after some time it becomes almost constant.

In Figures 3 and 4, we note that there is a decreasing trend in the expected waiting time for priority as well as non-priority subscribers with increasing value of k which is what we expect in real world situations. Similar trend has been found for E(W) as observed in Figures 1(a-c) and 2(a-c) with E(N) when we vary λ_1, λ_2, a_1, a_2, α and θ.

From the tables and graphs, overall we conclude that:

- The average number of subscribers and expected waiting time in priority and non priority queue increase as arrival rates (λ_1, λ_2), mean batch sizes (a_1, a_2) of priority as well as non priority subscribers and failure rate (α) of the server increase.
- On increasing retrial rate (θ), the average queue length and expected waiting time both decrease.
- By increasing the phases (k) of the service time and the repair time distributions, the average queue length and the expected waiting time for both priority and non priority subscribers, decrease which is in agreement with the physical situation.

8. CONCLUSION

The performance analysis of bulk arrival retrial queue with random service interruption under non-preemptive priority rule is studied. Such system is frequently encountered in practice and in particular, in service oriented operations. PGF of queue size distribution at an arbitrary time is obtained. Explicit formulae for the queue length and expected waiting time have been found for both priority and non-priority subscribers, which can be computed easily as shown by numerical illustration. The LST of waiting time distribution has also been obtained. Through numerical experiments, we have exhibited the impact of bulk size, retrial rate, failure rate, variation of service time and repair time distribution on the system behaviour. The model developed incorporates many features simultaneously including (i) bulk input (ii) retrial (iii) unreliable server and (iv) priority, which makes our results applicable to more versatile congestion situations encountered in computer and communication systems, distribution and service sectors, production and manufacturing systems, etc..

ACKNOWLEDGEMENTS

We are thankful to both learned referees as well as Editor in Chief for their constructive comments and suggestions which have helped us a lot in improving our paper.

REFERENCES

1. Aguir, M.S., Akşin, O.Z., Karaesmen, F. and Dallery, Y. (2008).On the interaction between retrials and sizing of call centers. *European Journal of Operational Research*, 191(2): 398-408.
2. Almasi, B., Roszik, J. and Sztrik, J. (2005). Homogenous finite source retrial queue with server subject to breakdowns and repairs. *Mathematical and Computer Modelling*, 42: 673-682.
3. Atencia, I. and Moreno, P. (2006). Geo/G/1 retrial queue with 2nd optional service. *International Journal of Operational Research*, 1(4): 340-362.
4. Atencia, I., Fortes, I., Moreno, P. and Sanchez, S. (2006). An M/G/1 retrial queue with active breakdowns and Bernoulli schedule in the server. *Information and Management Science*, 17(1): 1-17.
5. Choi, B.D., Chung, Y.H. and Dudin, A.N. (1999). The BMAP/SM/1 model with controllable operations modes. *European Journal of Operational Research*, 131:16-30.
6. Cobham, A. (1954). Priority assignment in waiting line problems. *Operations Research*, 2: 70-76.
7. Corral, A. G. (2002). Analysis of a single server retrial queue with quasi-random input and nonpreemptive priority. *Computers and Mathematics with Applications*, 43: 767-782.
8. Dudin, A.N., Krishnamoorthy, A., Joshua, V.C. and Tsarenkov, G.V. (2004). Analysis of the BMAP/G/1 retrial system with search of customers from the orbit. *European Journal of Operational Research*, 157: 169-179.
9. Falin, G.I. and Templeton, J.G.C. (1997). *Retrial Queues,* Chapman and Hall, London.
10. Falin, G.I., Artalejo, J.R. and Martin, M. (1993). On the single server retrial queue with priority customers. *Queueing Systems*, 14: 439-455.
11. Gross, D. and Harris, C.M. (1985). Fundamentals of Queueing Theory. John Wiley and Sons, New York
12. Gurukajan, M. and Srinivasan, B. (1995). A complex two unit system with random breakdown of repair facility. *Microelectronics and Reliability*, 35(2): 299-322.
13. Hassin, R. and Haviv, M. (2006). Who should be given priority in a queue?. *Operations Research Letters*, 34: 191-198.
14. Hawkes, A.G. (1965). Time dependent solution of a priority queue with bulk arrivals. *Operations Research*, 13: 586-596.
15. Iftikhar, M., Singh, T., Landfeldt, B. and Caglar, M. (2008): Multiclass G/M/1 queueing system with self-similar input and non-preemptive priority. *Computers and Communications*, 31(5): 1012-1027.
16. Jain, M., Sharma, G.C. and Pundhir, R.S. (2007). The M/G/1 retrial queueing system with setup, server breakdown and repair. *Ganita*, 58(2):137-155.
17. Kernane, T. and Aissani, A. (2006). Stability of retrial queues with versatile retrial policy. *Journal of Applied Math. Stochastic Anal.*, 1-16.
18. Kim, C., Klimenok, V.I. and Orlovsky, D.S. (2008). The BMAP/PH/N retrial queue with Markovian flow of breakdown. *European Journal of Operational Research*, 189(3): 1057-1072.
19. Langaris, C. and Moutzoukis, E.A. (1995). A retrial queue with structured batch arrivals. priorities and server vacations, *Queueing Systems*, 20: 341-368.
20. Pakes, A.G. (1969). Some conditions for ergoticity and recurrence of Markoc chains. Operations Research, 17: 1058-1061.
21. Pekoz, E.A. (2002). Optimal policy for multi server non-preemptive priority queues. *Queueing Systems*, 42: 91-101.
22. Pla, V. and Giner, V.C. (2005). Analysis of priority channel assignment schemes in mobile cellular communication system: a spectral theory approach. *Performance Evaluation*, 59(2-3): 199-224.
23. Roszik, J., Sztrik, J. and Virtamo, J. (2007). Performance analysis of finite-source retrial queues operating in random environments. *International Journal of Operational Research*, 2(3): 254-268.
24. Sandhu, D. and Posner, M.J.M. (1989). A priority M/G/1 queue with the application to data/voice communication. *European Journal of Operational Research*, 40: 99-108.
25. Soo, H.M. and Chung, J.M. (2003). Analysis of non-preemptive priority queueing of Diffserv Network with bulk arrivals, AEU- *International Journal of Electronics and Communications*, 57: 409-414.
26. Takagi, H. (1991). Queueing Analysis: A Foundation of Performance Evaluation, Vol. 1: Vacation and Priority Systems, Part I, Elsevier Science Publishers B.V., Amsterdam.
27. Takahashi, Y. and Takagi, H. (1990). Structured priority queue with batch arrivals, *Journal of Operational Research Society of Japan*, 33(3): 242-261.
28. Walraevens, J., Steyaert, B. and Bruneel, H. (2008). Analysis of a discrete-time preemptive resume priority buffer. European Journal of Operational Research, 186(1): 182-201.
29. White, H. C. and Christie, L. S. (1958). Queueing with preemptive priorities or with breakdowns, *Operations Research*, 6: 79-95.

30. Wu, X., Brill, P. Hlynka, M. and Wang, J. (2005). An M/G/1 retrial queue with balking and retrials during service. International Journal of Operational Research, 1(1/2): 30-51.

APPENDIX A

Proof of Theorem 1:

We know that expected number of subscribers who arrives in the system during completion time i.e. service time that includes repair time also. Thus

$$E(\nu_{\ell,d}) = \rho_\ell(1+\alpha\gamma_{\ell 1}); \qquad \ell = 1,2$$

Let $i \geq 1$, then for Y_d=1, we have

$$E\left(v_{1,d} / Y_d = 1\right) = \rho_1\left(1+\alpha\gamma_{11}\right) \tag{A.1}$$

$$E\left(v_{2,d} / Y_d = 1\right) = \lambda_2{}^a 2\beta_{11}\left(1+\alpha\gamma_{11}\right) \tag{A.2}$$

Now

$$\begin{aligned} x_{\ell,i,j} &= [\lambda_2 a_2 \beta_{11}(1+\alpha\gamma_{11}) + 1 - \rho_2(1+\alpha\gamma_{21})][-1 + E(\nu_{1,d} / Y_d = 1))] \\ &\quad + [\lambda_1 a_1 \beta_{21}(1+\alpha\gamma_{21}) + 1 - \rho_1(1+\alpha\gamma_{11})][E(\nu_{2,d} / Y_d = 1)] \\ &= \rho_1(1+\alpha\gamma_{11}) + \rho_2(1+\alpha\gamma_{21}) - 1 \end{aligned} \tag{A.3}$$

For $i = 0$, we obtain

$$N_{1,d} = v_{1,d} \tag{A.4}$$

$$N_{2,d} = N_{2,d-1} - B_d + v_2, d \tag{A.5}$$

Now

$$E\left(v_{1,d} / N_{1,d-1} = 0, N_{2,d-1} = j\right) = \frac{\lambda_1}{\lambda_1+\lambda_2+j\theta}\lambda_1 a_1\beta_{11}(1+\alpha\gamma_{11}) + \frac{\lambda_2+j\theta}{\lambda_1+\lambda_2+j\theta}\lambda_1 a_1 \beta_{21}(1+\alpha\gamma_{21}) \tag{A.6}$$

$$E\left(B_d / N_{1,d-1} = 0, N_{2,d-1} = j\right) = \frac{j\theta}{\lambda_1+\lambda_2+j\theta} \tag{A.7}$$

$$E\left(v_{2,d} / N_{1,d-1} = 0, N_{2,d-1} = j\right) = \frac{\lambda_1}{\lambda_1+\lambda_2+j\theta}\lambda_2 a_2\beta_{11}(1+\alpha\gamma_{11}) + \frac{\lambda_2+j\theta}{\lambda_1+\lambda_2+j\theta}\lambda_2 a_2 \beta_{21}(1+\alpha\gamma_{21}) \tag{A.8}$$

$$\begin{aligned} x_{\ell,i,j} &= [\lambda_2 a_2 \beta_{11}(1+\alpha\gamma_{11}) + 1 - \rho_2(1+\alpha\gamma_{21})] E(\nu_{1,d} / N_{1,d-1} = 0, N_{2,d-1} = j)] \\ &\quad + [\lambda_1 a_1 \beta_{21}(1+\alpha\gamma_{21}) + 1 - \rho_1(1+\alpha\gamma_{11})][-E(B_d / N_{1,d-1} = 0, N_{2,d-1} = j) \\ &\quad + E(\nu_{2,d} / N_{1,d-1} = 0, N_{2,d-1} = j)] \end{aligned} \tag{A.9}$$

In particular, we get

$$x_{\ell,0,j} = \rho_1(1+\alpha\gamma_{11}) + \rho_2(1+\alpha\gamma_{21}) - 1 + \frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+j\theta} \tag{A.10}$$

APPENDIX B

Proof of Theorem 2:

On solving Eqs (4.4) and (4.5) and using Eqs (4.8) and (4.9), we get

$$R_1(z_1, z_2, x, y) = \alpha P_1(z_1, z_2, x)\exp\{-[\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2)]y\}\overline{G}_1(y) \quad \text{B.1}$$

$$R_2(z_1, z_2, x, y) = \alpha P_2(z_1, z_2, x)\exp\{-[\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2)]y\}\overline{G}_2(y) \quad \text{B.2}$$

On solving Eqs (4.2) and (4.3) and using Eqs (B.1) and (B.2), we find

$$P_1(z_1, z_2, x) = P_1(z_1, z_2, 0)\times\exp\{-[\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2) + \alpha\overline{g}_1^*(\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2))]x\}\overline{B}_1(x) \quad \text{B.3}$$

$$P_1(z_1, z_2, x) = P_1(z_1, z_2, 0)\times\exp\{-[\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2) + \alpha\overline{g}_2^*(\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2))]x\}\overline{B}_2(x) \quad \text{B.4}$$

After some algebraic manipulation, Eq. (4.6) becomes

$$P_1(z_1, z_2, 0) = \frac{\theta[k_2(z_1, z_2) - z_2]\dfrac{dP_0(z_2)}{dz_2} - [\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \dfrac{\lambda_2 C_2(z_2)}{z_2}k_2(z_1, z_2)]}{[z_1 - k_1(z_1, z_2)]}P_0(z_2) \quad \text{B.5}$$

where $k_i(z_1, z_2) = b_i^*\{(\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2)) + \alpha\overline{g}_i^*(\lambda_1 - \lambda_1 C_1(z_1) + \lambda_2 - \lambda_2 C_2(z_2))\}$ is the Laplace Stieltjes transform for the completion time.

Here $P_0(z_2) = P_0(z_1, z_2)$ since $N_1(t) = 0$ if $Y(t) = 0$ then $P_{0,i,j} = 0$ for $i \geq 1$.

On taking $z_1 - k_1(z_1, z_2) = 0$ (in Eq. B.5) B.6

Note that equation B.6 has a unique root $z_1 = h(z_2)$ in the unit disk $|z_1| \leq 1$ [cf. Falin and Templeton, 1997]. The function $h(z_2)$ can be thought of as the generating function of the number of class two jobs that arrive during the busy period formed by class one jobs.

Solution of Eq. (B.5) using Eq. (B.6) gives Eq. (6.1).

APPENDIX C

Proof of Theorem 3:

Using

$$P_1(z_1, z_2) = \int_0^\infty P_1(z_1, z_2, x)\,dx\,,\quad P_2(z_1, z_2) = \int_0^\infty P_2(z_1, z_2, x)\,dx \quad \text{C.1}$$

$$R_1(z_1, z_2) = \int_0^\infty\int_0^\infty R_1(z_1, z_2, x, y)dxdy, R_2(z_1, z_2) = \int_0^\infty\int_0^\infty R_2(z_1, z_2, x, y)dxdy \quad \text{C.2}$$

we get Eqs (7.1)-(7.4)

Using the normalizing condition $P_0(1) + P_1(1,1) + P_2(1,1) + R_1(1,1) + R_2(1,1) = 1$, we determine the normalizing constant.

$$P(I) = \lim_{z_2 \to 1} P_0(z_2) = 1 - \rho_1(1 + \alpha\gamma_{11}) - \rho_2(1 + \alpha\gamma_{21}) \quad \text{C.3}$$

$$P(B_1) = \lim_{z_1, z_2 \to 1}\int_0^\infty P_1(z_1, z_2, x)dx\,,\quad P(B_2) = \lim_{z_1, z_2 \to 1}\int_0^\infty P_2(z_1, z_2, x)dx \quad \text{C.4}$$

$$P(R_1) = \lim_{z_1, z_2 \to 1}\int_0^\infty\int_0^\infty R_1(z_1, z_2, x, y)dxdy\,, P(R_2) = \lim_{z_1, z_2 \to 1}\int_0^\infty\int_0^\infty R_2(z_1, z_2, x, y)dxdy \quad \text{C.5}$$

$$E(N_1) = \frac{\partial}{\partial z_1} \Pi(z_1, z_2)\bigg|_{z_1, z_2 \to 1}, \quad E(N_2) = \frac{\partial}{\partial z_2} \Pi(z_1, z_2)\bigg|_{z_1, z_2 \to 1} \qquad \text{C.6}$$

These Eqs are used to establish results given in Eq. (8.1)-(8.7)

APPENDIX D

Proof of theorem 5:

$$P(s, z_1, z_2) = P_1(s, z_1, z_2) + P_2(s, z_1, z_2) + P^{(0)}(s, z_1, z_2) \qquad \text{D.1}$$

$$P_1(s, z_1, z_2) = \sum_{d=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=1}^{\infty} z_1^i z_2^j P_{1,i,j}^{(d)}(s), \quad P_2(s, z_1, z_2) = \sum_{d=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=1}^{\infty} z_1^i z_2^j P_{2,i,j}^{(d)}(s) \qquad \text{D.2}$$

$$P^{(0)}(s, z_1, z_2) = z_1^m z_2^n, \quad \Psi(s, z_2) = \sum_{d=0}^{\infty} \Psi^{(d)}(s, z_2) \qquad \text{D.3}$$

$$P_{1,i,j}^{(d)}(s) = P\{E_{1,i,j}^{(d)}(s)\}, \qquad P_{2,i,j}^{(d)}(s) = P\{E_{2,i,j}^{(d)}(s)\}, \qquad P_{i,j}^{(0)}(s) = \delta_{i,m}\delta_{j,n} \qquad \text{D.4}$$

$$E_{1,i,j}^{(d)}(s) = \{Y_d = 1, N_{1,d} = i, N_{2,d} = j, T_{m.n} > \tau_d\} \qquad \text{D.5}$$

where τ_d denotes time instant of the d^{th} departure customer

$$E_{2,i,j}^{(d)}(s) = \{Y_d = 2, N_{1,d} = i, N_{2,d} = j, T_{m.n} > \tau_d\} \qquad \text{D.6}$$

The equation and its solution for the above-defined generating functions can be written as [cf. Falin and Templeton, 1997]

$$\begin{aligned}
&[z_1 - k_1(s, z_1, z_2)]P(s, z_1, z_2) \\
&= \left[-(s + \lambda_1 + \lambda_2)k_1(s, z_1, z_2) + \lambda_1 C_1(z_1)k_1(s, z_1, z_2) + \frac{\lambda_2 C_2(z_2)}{z_2} k_2(s, z_1, z_2) - \frac{\theta z_1}{z_2} k_2(s, z_1, z_2) \right] \Psi(s, z_2) \\
&+ \theta[z_1 k_2(s, z_1, z_2) - z_2 k_1(s, z_1, z_2)] \frac{\partial \Psi(s, z_2)}{\partial z_2} + z_1^{m+1} z_2^n
\end{aligned} \qquad \text{D.7}$$

$$\Psi(s, z) = \int_z^{g(s)} \frac{z(h(s,u))^m u^{n-1}}{\theta[k_2(s, h(s,u), u) - u]} \times \exp\left\{ \frac{1}{\theta} \int_u^z \frac{[\theta + s + \lambda_1 + \lambda_2 - \lambda_1 C_1(h(s,v)) - \frac{\lambda_2 C_2(s,v) k_2(h(s,v), v)}{u}]}{[k_2(s, h(s,v), v) - v]} dv \right\} du \qquad \text{D.8}$$

Laplace transform of the waiting time in queue of tagged subscriber is obtained as

$$\mathbf{E}\mathbf{e}^{-\mathbf{s}\mathbf{T}_{\mathbf{mn}}} = \theta \Psi(s, 1).$$