Top (2013) 21:542–571
DOI 10.1007/s11750-011-0198-4

ORIGINAL PAPER

A preemptive resume priority retrial queue with state dependent arrivals, unreliable server and negative customers

Ioannis Dimitriou

Received: 10 November 2010 / Accepted: 23 May 2011 / Published online: 11 June 2011

Abstract In this paper we consider an unreliable single server retrial queue accepting two types of customers, with negative arrivals, preemptive resume priorities and vacations. A distinguishing feature of the model is that the rates of the Poisson arrival process depends on the server state. For this model we investigate the stability conditions and the joint queue length distribution in steady state. We also prove that our model satisfies the stochastic decomposition property. Transient, as well as steady state solutions for reliability measures are obtained. Finally, numerical results demonstrate the typical features of the model under consideration.

Keywords Retrial queue · Unreliable server · Negative customers · State dependent arrivals · Preemptive resume priority · Single vacation · Stochastic decomposition · Reliability

Mathematics Subject Classification (2000) 60K25 · 90B22

1 Introduction

Retrial queues have been extensively studied, since they arise in various systems such as telephone switching systems, telecommunication networks with retransmission, call centers and computer networks.

The characteristic feature of the retrial queues is that arriving customers find the servers unavailable, make new attempts to get service after a random time. In call centers, there is a certain number of servers that answer customer calls. When a customer call arrives, it will be served immediately if a server is available. If all servers are busy with other calls, the customer will be put on hold, and will be asked to wait

I. Dimitriou (✉)
Department of Mathematics, University of Ioannina, 45110 Ioannina, Greece
e-mail: idimit@cc.uoi.gr

until a server becomes available. Some customers are patient enough to wait for a server to become available, while others will hang-up or abandon after waiting for some time or immediately. Some calls will automatically be disconnected and the customer will be asked to call back later. A portion of these customers will redial and try to access the call center.

For a complete survey on retrial queues we refer Artalejo (1999, 2010), Kulkarni and Liang (1997) and the monographs by Falin and Templeton (1997) and Artalejo and Gomez-Corral (2008). Queueing models of call centers in which the retrial phenomenon is taken into consideration are given in the reviews of Koole and Mandelbaum (2002) and in the work by Aguir et al. (2004).

The vast majority of papers on retrial queues assumes that the arrival rate remains constant, although it may vary per class of customers (in case when more than one type of customers arrive at the system). The model under consideration, allows the arrival rate of each type of customer to vary depending on the server state. Besides its mathematical interest the model under consideration have practical justification. In practice, the proposed model provides information to the sources of the customers at any time, about the status of the server in order to control the arrival rates that lead to the control of congestion of the system. On the other hand, it is possible the sources to vary the demand process due to unpredictable situations, such as failures, happened in a service procedure (see Shogan 1979).

Several authors have studied retrial queues with priorities. High priority customers are queued and served according to some discipline. In case of blocking, low priority customers leave the system and retry until they find the server free. In related bibliography (Choi and Chang 1999; Falin et al. 1993; Langaris and Moutzoukis 1995), the high priority customers have either preemptive or non-preemptive priority over the low priority customers. Moreover, in a paper by Artalejo et al. (2001) repeated demands appeared to have preemptive priority over the waiting line. To the author's best knowledge, there have been a little attention on research in retrial queues with preemptive resume priorities among two types of customers.

In the last fifteen years, several papers deal with the queueing modeling of systems operating in the presence both of negative customers and repeated attempts. Negative customers can be interpreted as virus in a computer networks, or generally as orders for the customers to leave the system immediately. For a related bibliography we refer the papers by Artalejo and Gomez-Corral (1998, 1997), Anisimov and Artalejo (2001), Shin (2007), Wang et al. (2008). We have to state here that the vast majority of papers on retrial queues with negative customers concerns non-priority retrial queues.

In the most of the queueing literature the server is assumed to be reliable and always available to customers, but it is clear that this assumption in real systems such as communications and manufacturing systems where the machine may be subject to scheduled backups and unpredictable failures, seems to be unrealistic. Retrial queues with server's breakdowns and repairs have been studied in several papers. As a related work we refer Kulkarni and Choi (1990), Aissani and Artalejo (1998), Aissani (1994) and Dimitriou and Langaris (2010). Because of limited ability of repairs and heavy influence of the breakdowns on the performance of the system, it is of essential importance to study the reliability of retrial queues with breakdowns and repairs (see Wang et al. 2001, 2008; Wang 2008).

Concerning state dependent arrival rates, more literature is available for systems consisting of only one queue, often assuming phase-type distributions for vacations and/or service times. A system consisting of a single queue with server breakdowns and arrival rates depending on the server status is studied by Shogan (1979). Shanthikumar (1988) states a stochastic decomposition result for the queue length in an M/G/1 queue with server vacations under less restrictive assumptions than Furhmand and Cooper (1985). One of the relaxations is that the arrival rate of customers may be different during visit periods and vacations. Recently, a polling model with arrival rates that vary depending on the location of the server has been studied in detail by Boon et al. (2010).

There have been a little attention in retrial queueing literature with state dependent parameters. Parthasarathy and Shudesh (2007) consider a Markovian single server retrial queue with parameters depending on the orbit length, while Gomez-Corral and Ramalhoto (1999) study Markovian multiserver retrial queues with parameters depending on the number of servers. Recently, Artalejo and Li (2010), study a discrete time retrial queue with arrivals depend on the state of the system. To author best knowledge the majority of retrial queues with state dependent arrivals deal with non-priority retrial queues with Markovian character.

In this paper we consider a single server retrial queue, accepting two types of customers with the additional features of preemptive resume priorities, breakdowns with repairs, negative arrivals, single vacations and state dependent arrivals. Clearly, the state of the server is not known to the customers that are in the retrial box. The status of the server is known only to the sources of the customers.

The purpose of this paper is to generalize the main model with preemptive resume priorities in several ways. We generalize the main model by allowing server's failures with repairs, introducing the impact of negative customers and mainly by integrating the concept of state dependent arrivals. Specifically, we assume that arrival rates depend both on server's state and on the customer type. We have to state here that the dependence on server's state is generated by internal source and not on an external source such as Markov modulated inputs. To author best knowledge, it is the first time in related literature that the realistic concept of failures with repairs is integrated in retrial queue with preemptive resume priorities among two types of customers. Moreover interesting reliability indices of the server are also obtained and complete the analysis. Another interesting feature which is for the first time integrated in such a model is the presence of negative customers that deletes the customer in service. The proposed model is high abstract and quite complex. Its complexity becomes sharper as we allow the arrival rates to depend both on server's state and the customer type. To author best knowledge is the first time in the retrial queueing literature that such a complex but flexible arrival discipline is integrated in priority retrial queues. Furthermore, we establish for the first time in related literature a stochastic decomposition law for this priority retrial queue with varying arrival rates.

Such arrival process is useful to model many practical situations where the customers aware when the server is available or not and provide a very important flexibility. The dependence of arrival process on the server's state have many practical advantages. Firstly, by providing, information to the customers about the server's state, improve the quality of service and secondly we can model several communication systems where the traffic is able to adapt its rate according to the server condition.

In many real-life queueing systems, e.g. in computer systems, manufacturing systems and communication systems etc., the server is subject to breakdown or generally alternates stochastically between different states, such as operational, failed and scheduled interruptions. In such models the arrival rate of jobs may be influenced by the status of the server.

Our retrial queue has applications in a packed-switched network. The router is an interconnection device that attaches two or more networks in a packet-switched network, which takes charge of receiving packets and forwarding them to the next hop, according to the routing information found in its routing table. Two types of IP packets (urgents and regulars) arrives at the router according to a Poisson stream (depending on their type). A packet receives service immediately if the router is idle or it will enter two buffers, according to its importance (priority), which are located inside the router. Some maintenance activities, such as scan virus and routing information backup can be programmed on a regular basis when the router is idle. When these maintenance activities are finished, the router will enter the idle state again and wait for new packets to arrive. The router may subject to breakdowns during service period and receive repair immediately. Such a system is affected by a virus, causing the destruction of the message in transmission (negative customers). The transmission of a regular packet may be interrupted because of an arrival of an urgent packet, while the interrupted packet resumes its transmission from the point of interruption. In this scenario, buffers in the router, router retransmission policy and maintenance activities correspond to the queue and orbit, the server, the retrial discipline, and the vacation policy, respectively. It seems to be realistic to assume that the transmission demands in the router are influenced by the presence of the above mentioned unpredictable random phenomena.

Another relevant application of the proposed model is referred to the single machine production systems. In the context of this production setting, the situation with two priority levels is oftentimes encountered in practice, where production department have to supply both internal and external customers, the latter of which is commonly given a preferential treatment.

More precisely, orders from external customers are queued up and have preemptive resume priority over the orders from internal customers. That is, an external order have to be satisfied immediately upon arrival and if the department is serving an internal order at the moment of the arrival of an external order then the process for internal order is interrupted and resumes whenever the department is free of external orders. An order from an internal customer, if it is not fulfilled upon arrival, has to be retransmitted (retrials) in production department. Any time the facility satisfy all the external orders and a possibly interrupted internal order, accomplish a maintenance (vacation) of the machine. Moreover, mechanical parts of the machine may fail. Then the machine is sent for repair, while the order has to be retransmitted after random period. Furthermore, instantaneous blackouts (negative arrivals), may cause the destruction of the item under production, and as a result the loss of the order.

Thus, it is natural the demand process to be influenced because of this uncertainty and the unpredictable environment. On the other hand the manager of the department must control the flow of newly arriving orders, by providing information about the state of the facility. More precisely, he/she communicates with the customers informing them by a message (possibly on line), about the facility status in order to manage the demand process of the product. In such a case, the demand process (arrival process) depends on the status of the production facility (server). An interesting case is the one where arrival rates become zero during specific periods.

The rest of the paper is organized as follows. In Sect. 2 we describe the mathematical model, while in Sect. 3 some important preliminary results are stated. In Sect. 4 we investigate the stability condition of the model, under which the queue length distribution is investigated in Sect. 5. A stochastic decomposition result is presented in Sect. 6, while in Sect. 7 some important performance measures are obtained. Sect. 8 relates to the reliability results obtained for this model. Finally, numerical results that demonstrate the typical features of our model are presented in Sect. 9.

2 Model description and notations

Consider a single server queueing system accepting two types of customers P_i, $i = 1, 2$. Assume that P_1 customers have preemptive resume priority over P_2 customers. That is, if a P_1 customer arrives during the service of a P_2 customer, he interrupts him and push the server to start serving P_1 customers. The preempted unit remains in the service zone and resumes service, whenever the server become available. The service times of P_i customers follow an arbitrary distribution with distribution function (d.f.) $B_i(x)$, probability distribution function (p.d.f.) $b_i(x)$, Laplace–Stieltjes Transform (LST) $\beta_i^*(s)$, finite mean value $\bar{b}_i$ and second moment about zero $\bar{b}_i^{(2)}$.

The system under consideration suffers from breakdowns, so that the server's lifetime is exponential distributed with parameter μ. If a failure occurs, the server is sent immediately for repair, while the unit being served joins the retrial box from where retries come after an exponential amount of time with parameter α, to connect with the server. Repair times are assumed to be arbitrarily distributed with d.f. $B_3(x)$, and p.d.f. $b_3(x)$, LST $\beta_3^*(s)$, finite mean value $\bar{b}_3$ and second moment about zero $\bar{b}_3^{(2)}$.

Whenever the server becomes free, that is, when there are no customers in the ordinary queue or in the service zone, after a service or repair completion, departs for a single vacation of arbitrarily distributed length with d.f. $B_4(x)$, p.d.f. $b_4(x)$, LST $\beta_4^*(s)$, finite mean value $\bar{b}_4$ and second moment about zero $\bar{b}_4^{(2)}$. Upon returning from the vacation, the server starts serving P_1 customers (if any), or else remains idle awaiting the first unit that request service, either from outside or from the retrial box.

Moreover, a flow of negative arrivals reduces the congestion of our system, by deleting the customer being served. Negative customers, arrive after an exponential amount of time with parameter ν, push out the customer being served and have no other effect on the system.

If an arriving P_1 customer finds the server idle, he occupies him and begins to be served. If an arriving P_1 customer finds the server either busy with a P_1 customer, or under repair, or on vacation, he joins an ordinary queue, waiting to be served, while an arriving P_2 customer who finds the server unavailable join the retrial box.

The feature that distinguishes the model under consideration from commonly studied retrial queues, is the arrival process. This arrival process is a standard Poisson process, but the rate depends on the server's state. The arrival rate of P_i customers is λ_{ij}, where j denotes the state of the server, which is either serving a specific type of customer, or is under repair, or is on vacation, or is in the idle mode. Another novel characteristic of the model is the presence of breakdowns, repairs and negative customers, which are integrated for the first time in related literature with preemptive resume priorities among two types of customers.

Let us denote by $N_i(t), i = 1, 2$, to be the number of P_i customers in the ordinary queue and in the retrial box, respectively, at time t. Denote also

$$\xi_t = \begin{cases} 0, & \text{idle at } t, \\ i, & \text{busy with a } P_i,\ i = 1, 2 \text{ customer at } t, \\ 3, & \text{under repair at } t, \\ 4, & \text{on vacation at } t, \end{cases}$$

$$u_t = \begin{cases} 0, & \text{no } P_2 \text{ customer in limbo (has preempted earlier) at } t, \\ 1, & \text{a } P_2 \text{ customer in limbo at } t. \end{cases}$$

Thus, P_i customers, $i = 1, 2$, arrive according to Poisson process with parameter λ_{ij}, given that the server state is j, $j = 0, 1, 2, 3, 4$. Define also the following state probabilities:

$$\begin{aligned} p_i(k_1, k_2, x, t)\, dx &= P\big(N_1(t) = k_1, N_2(t) = k_2, x < \bar{B}_i(t) \le x + dx, \\ &\qquad \xi_t = i, u_t = 0\big), \quad i = 1, 3, 4, \\ p_2(k_2, x, t)\, dx &= P\big(N_2(t) = k_2, x < \bar{B}_2(t) \le x + dx, \xi_t = 2, u_t = 0\big), \\ p_{i1}(k_1, k_2, x, y, t)\, dx\, dy &= P\big(N_1(t) = k_1, N_2(t) = k_2, x < \bar{B}_i(t) \le x + dx, \\ &\qquad y < \bar{B}_2(t) \le y + dy, \xi = i, u_t = 1\big), \quad i = 1, 3, \\ q(k_2, t) &= P\big(N_2(t) = k_2, \xi_t = 0, u_t = 0\big), \end{aligned}$$

where $\bar{X}$ the elapsed duration of the random variable X.

3 General results

In this section we are going to derive some useful results that are necessary for the following analysis.

Denote by S_i the time interval from the epoch that a P_i customer starts his service, until the epoch the server is ready for a "new service". In this case, a "new service" starts either when the current service terminated successfully, or after a repair completion caused by a breakdown, or after a negative arrival. Denote by $N_j(S_i)$, $i, j = 1, 2$, the number of P_j customers that arrive during S_i. We have

$$\begin{aligned} a_i(k_1, k_2, t)\, dt &= P\big(t < S_i \le t + dt,\ N_j(S_i) = k_j,\ j = 1, 2\big), \quad i = 1, 2, \\ a_i^*(z_1, z_2, s) &= \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \int_0^{\infty} e^{-st} a_i(k_1, k_2, t) z_1^{k_1} z_2^{k_2}\, dt. \end{aligned} \tag{1}$$

Then if $p^{(n)}(k_1,k_2,t)=e^{-\lambda_{1n}t}\frac{(\lambda_{1n}t)^{k_1}}{k_1!}e^{-\lambda_{2n}t}\frac{(\lambda_{2n}t)^{k_2}}{k_2!}$, $n=0,1,2,3,4$,

$$\begin{aligned}a_1(k_1,k_2,t)&=e^{-(\mu+\nu)}p^{(1)}(k_1,k_2,t)b_1(t)+\nu e^{-(\mu+\nu)t}p^{(1)}(k_1,k_2,t)\big(1-B_1(t)\big)\\&\quad+\mu e^{-(\mu+\nu)t}\sum_{i_1=0}^{k_1}\sum_{j_1=0}^{k_2-1}p^{(1)}(i_1,j_1,t)\big(1-B_1(t)\big)\\&\quad*p^{(3)}(k_1-i_1,k_2-1-j_1,t)b_3(t),\end{aligned}$$

where $*$ means convolution.

After manipulations

$$a_1^*(z_1,z_2,s)=\beta_1^*\big(\sigma_1(z_1,z_2,s)\big)+\frac{1-\beta_1^*(\sigma_1(z_1,z_2,s))}{\sigma_1(z_1,z_2,s)}\big[\nu+\mu z_2\beta_3^*\big(\sigma_3(z_1,z_2,s)\big)\big],$$

where

$$\sigma_i(z_1,z_2,s)=s+\lambda_{1i}(1-z_1)+\lambda_{2i}(1-z_2)+(\mu+\nu)\delta_{\{i=1\}},\quad i=1,3,4,$$

where $\delta_{\{\}}$ is Kronecker's delta. Denote

$$\rho_1=\frac{\partial}{\partial z_1}a_1^*(z_1,1,0)|_{z_1=1}=\frac{1-\beta_1^*(\mu+\nu)}{\mu+\nu}(\lambda_{11}+\lambda_{13}\mu\bar{b}_3).$$

Using the above results, the following lemma is a simple extension of Takacs theorem (Takacs 1962).

Lemma 1 *For* (i) $|z_2|<1$, $\mathrm{Re}(s)\geq 0$, *or* (ii) $|z_2|\leq 1$, $\mathrm{Re}(s)>0$, *or* (iii) $|z_2|\leq 1$, $\mathrm{Re}(s)\geq 0$ *and* $\rho_1>1$, *the relation*

$$z_1-a_1^*(z_1,z_2,s),\tag{2}$$

has one and only one root, $z_1=x(s,z_2)$ *say, inside the region* $|z_1|<1$. *Specifically for* $s=0$ *and* $z_2=1$, $x(0,1)$ *is the smallest positive real root of* (2) *with* $x(0,1)<1$ *if* $\rho_1>1$ *and* $x(0,1)=1$ *for* $\rho_1\leq 1$.

Define also by $\mathbf{\Theta}^{(i)}$ the duration of the busy period of P_1 customers, initiated by i P_1 customers and by $N(\mathbf{\Theta}^{(i)})$ the number of the new P_2 customers that arrive during $\mathbf{\Theta}^{(i)}$. If $g_m^{(i)}(t)dt=P(t<\mathbf{\Theta}^{(i)}\leq t+dt, N(\mathbf{\Theta}^{(i)})=m)$, then by following the lines of Langaris and Katsaros (1995) or in Takacs (1962) (pp. 60–63) we obtain

$$g^{(i)}(s,z_2)=\sum_{m=0}^{\infty}z_2^m\int_0^{\infty}e^{-st}g_m^{(i)}(t)\,dt=x^i(s,z_2),$$

where $x(s,z_2)$ is defined in the lemma above.

We have to point out here that in case of a P_2 customer, due to preemptive resume priority, the current service may be interrupted many times by the arrivals of P_1 customer, and resumes each time from the interruption point after the termination of the busy period that initiate the arrival of these P_1 customers.

By applying a similar argument as in (1) we arrive at

$$a_2^*(z_1, z_2, s) = \beta_2^*\big(\sigma_2(z_2, s)\big) + \frac{1 - \beta_2^*(\sigma_2(z_2, s))}{\sigma_2(z_2, s)}\big[\nu + \mu z_2 \beta_3^*\big(\sigma_3(z_1, z_2, s)\big)\big],$$

where

$$\sigma_2(z_2, s) = s + \lambda_{12}\big(1 - x(s, z_2)\big) + \lambda_{22}(1 - z_2) + \mu + \nu.$$

Let V, be the random time from the epoch the server departs for a single vacation until the epoch is for the first time idle. Denote by $N(V)$ the number of new P_2 customers that arrive during V. If $v_m(t)\,dt = P(t < V \le t + dt, N(V) = m)$, then

$$\begin{aligned} v_0(t) &= p^{(4)}(0, 0, t)b_4(t) + \sum_{i=1}^{\infty} p^{(4)}(i, 0, t)b_4(t) * g_0^{(i)}(t) * v_0(t), \\ v_m(t) &= \sum_{i=1}^{\infty}\sum_{j=1}^{m} p^{(4)}(i, j, t)b_4(t) * \sum_{k=0}^{m-j} g_k^{(i)}(t) * v_{m-j-k}(t). \end{aligned} \tag{3}$$

After manipulations

$$\begin{aligned} v^*(s, z_2) &= \sum_{m=0}^{\infty} z_2^m \int_0^{\infty} e^{-st} v_m(t)\,dt \\ &= \frac{\beta_4^*(\sigma_4(0, z_2, s))}{1 - \beta_4^*(\sigma_4(x(s, z_2), z_2, s)) + \beta_4^*(\sigma_4(0, z_2, s))}. \end{aligned}$$

Let us define the completion time C, of a P_2 customer, as the time interval from the epoch the server starts serving a P_2 customer, until the epoch the server is ready to depart for single vacation, and denote by $N(C)$ the number of new P_2 customers that arrive during C. Let $c_m(t)\,dt = P(t < C \le t + dt, N(C) = m)$.

$$\begin{aligned} c_m(t) &= p^{(2)}(0, m, t)e^{-(\mu+\nu)t}\big[b_2(t) + \nu\big(1 - B_2(t)\big)\big] \\ &\quad + \mu e^{-(\mu+\nu)t} \sum_{j=0}^{m-1} p^{(2)}(0, j, t)\big(1 - B_2(t)\big) * p^{(3)}(0, m - 1 - j, t)b_3(t) \\ &\quad + e^{-(\mu+\nu)t} b_2(t) \sum_{i=1}^{\infty}\sum_{j=0}^{m} p^{(2)}(i, j, t) * g_{m-j}^{(i)}(t) \\ &\quad + \mu e^{-(\mu+\nu)t} \sum_{i_1=0}^{\infty}\sum_{j_1=0}^{m-1} p^{(2)}(i_1, j_1, t)\big(1 - B_2(t)\big) \\ &\quad * \sum_{i_2=0}^{\infty}\sum_{j_2=0}^{m-1-j_1} p^{(3)}(i_2, j_2, t)b_3(t) * g_{m-1-j_1-j_2}^{(i_1+i_2)}(t) \end{aligned}$$

$$+\nu e^{-(\mu+\nu)t}\sum_{i=1}^{\infty}\sum_{j=0}^{m}p^{(2)}(i,j,t)\big(1-B_2(t)\big)*g_{m-j}^{(i)}(t).$$

After manipulations

$$c^*(s,z_2)=\sum_{m=0}^{\infty}z_2^m\int_0^{\infty}e^{-st}c_m(t)\,dt=a_2^*\big(x(s,z_2),z_2,s\big). \tag{4}$$

The concepts of Generalized Busy Period (GBP) of P_1 customer and the Generalized Completion Time (GCT) of a P_2 customer are very important for the analysis of our model. As GBP, say W_1, of a P_1 customer, define the time elapsed from the epoch a P_1 customer arrives in an idle system, until the epoch the server is idle for the first time. Clearly, from the above definitions $W_1=\mathbf{\Theta}^{(1)}+V$. The GCT of a P_2 customer is the time elapsed from the epoch a P_2 customer succeeds to connect with the server, until the epoch the server is idle for the first time. Definitely $W_2=C+V$. Denote

$$w_m^{(i)}(t)\,dt=P\big(t<W_i\le t+dt,N(W_i)=m\big),$$
$$w_i^*(s,z_2)=\sum_{m=0}^{\infty}z_2^m\int_0^{\infty}e^{-st}w_m^{(i)}(t)\,dt,\quad i=1,2.$$

Then

$$w_1^*(s,z_2)=x(s,z_2)v^*(s,z_2),\qquad w_2^*(s,z_2)=c^*(s,z_2)v^*(s,z_2). \tag{5}$$

In the sequel we are going to obtain some useful results. Thus, by differentiating with respect to z_2 at the point $z_2=1,\ s=0$ the above defined relations we arrive at

$$\begin{aligned}
\frac{\partial}{\partial z_2}x(0,z_2)|_{z_2=1}=m_1=\frac{L_1}{1-\rho_1}\\
&=\frac{(\frac{1-\beta_1^*(\mu+\nu)}{\mu+\nu})(\lambda_{21}+\mu(1+\lambda_{23}\bar{b}_3))}{1-\rho_1},\\
\tilde{\rho}_2&=\frac{\partial}{\partial z_2}c^*(0,z_2)|_{z_2=1}=\frac{\rho_2}{1-\rho_1},\\
\tilde{\rho}_4&=\frac{\partial}{\partial z_2}v^*(0,z_2)|_{z_2=1}=\frac{\rho_4}{1-\rho_1}=\frac{\frac{\bar{b}_4}{\beta_4^*(\lambda_{14})}[\lambda_{14}L_1+(1-\rho_1)\lambda_{24}]}{1-\rho_1},\\
\tilde{\rho}_d&=E\big(N(W_1)\big)=\frac{\partial}{\partial z_2}w_1^*(0,z_2)|_{z_2=1}=\frac{\rho_d}{1-\rho_1}=\frac{L_1+\rho_4}{1-\rho_1},\\
\tilde{\rho}_w&=E\big(N(W_2)\big)=\frac{\partial}{\partial z_2}w_2^*(0,z_2)|_{z_2=1}=\frac{\rho_w}{1-\rho_1}=\frac{\rho_2+\rho_4}{1-\rho_1},
\end{aligned} \tag{6}$$

where

$$\rho_2 = \left(\frac{1-\beta_2(\mu+\nu)}{\mu+\nu}\right)\left[L_1(\lambda_{12}+\lambda_{13}\mu\bar{b}_3)+(1-\rho_1)\left(\lambda_{22}+\mu(1+\lambda_{23}\bar{b}_3)\right)\right].$$

Moreover by differentiating (5) with respect to s in the point $z_2 = 1,\ s = 0$, we obtain the mean duration of GBP of P_1 customer and the mean duration of GCT of a P_2 customer. Then

$$\begin{aligned}
E(W_1) &= -\frac{\partial}{\partial s} w_1^*(s,1)|_{s=0} \\
&= \frac{[1-\beta_1^*(\mu+\nu)](1+\mu\bar{b}_3)}{(\mu+\nu)(1-\rho_1)} \\
&\quad + \frac{\bar{b}_4}{\beta_4^*(\lambda_{14})}\left(1+\frac{\lambda_{14}(1+\mu\bar{b}_3)[1-\beta_1^*(\mu+\nu)]}{(1-\rho_1)(\mu+\nu)}\right), \\
E(W_2) &= -\frac{\partial}{\partial s} w_2^*(s,1)|_{s=0} \\
&= \frac{\bar{b}_4}{\beta_4^*(\lambda_{14})}\left(1+\frac{\lambda_{14}(1+\mu\bar{b}_3)[1-\beta_1^*(\mu+\nu)]}{(1-\rho_1)(\mu+\nu)}\right) \\
&\quad + \frac{[1-\beta_2^*(\mu+\nu)](1+\mu\bar{b}_3)}{(\mu+\nu)} \\
&\quad \times \left(1+\frac{(\lambda_{12}+\mu\lambda_{13}\bar{b}_3)(1+\mu\bar{b}_3)[1-\beta_1^*(\mu+\nu)]}{(1-\rho_1)(\mu+\nu)}\right).
\end{aligned} \tag{7}$$

Now we are ready to state the following theorem, which is important for the future analysis.

Theorem 2 *For* (i) $\mathrm{Re}(s) > 0$, (ii) $\mathrm{Re}(s) \geq 0$, *and* $\rho = \rho_1 + \rho_2 + \rho_4 > 1$ *the equation.*

$$z_2 - w_2^*(s, z_2) = 0, \tag{8}$$

has one and only one root, $z_2 = \phi(s)$ *say, inside the region* $|z_2| < 1$. *Specifically for* $s = 0$, $\phi(0)$ *is the smallest positive real root of* (8) *with* $\phi(0) < 1$ *if* $\rho > 1$ *and* $\phi(0) = 1$ *for* $\rho \leq 1$.

Proof For the closed contour $|z_2| = 1$ and under the assumption (i) we have always

$$\left|w_2^*(s,z_2)\right| \leq w_2^*\left(\mathrm{Re}(s),1\right) < w_2^*(0,1) = 1 \equiv |z_2|,$$

while for $\mathrm{Re}(s) \geq 0$, we need to consider the closed contour $|z_2| = 1-\epsilon$ ($\epsilon > 0$ a small number) in which case

$$\left|w_2^*(s,z_2)\right| \leq w_2^*\left(\mathrm{Re}(s),1-\epsilon\right) < 1-\epsilon \equiv |z_2|, \tag{9}$$

only if in addition

$$\frac{d}{d\epsilon} w_2^*(0, 1-\epsilon)\,|_{\epsilon=0} = -\frac{\rho_2+\rho_4}{1-\rho_1} < \frac{d}{d\epsilon}(1-\epsilon)\,|_{\epsilon=0} = -1,$$

or we need now $\rho > 1$ for the relation (9) to hold. A final reference to Rouché's theorem completes the first part of the proof.

Moreover for $s=0$ the convex function $w_2^*(0, z_2)$ is a monotonically increasing function of z_2, for $0 \le z_2 \le 1$, taking the values $w_2^*(0,0) < 1$ and $w_2^*(0,1) = 1$ and so $0 < \phi(0) < 1$ if $\rho > 1$, while for $\rho \le 1$, $\phi(0)$ becomes equal to 1 and this completes the proof. □

4 Stability conditions

In order to obtain the stability conditions of our model we use results from the theory of Semi-Regenerative processes. We prove that the stochastic process that governs the evolution of our model is a semi-regenerative process and by discovering an embedded Markov renewal process, we follow the lines of Cinlar (1975) (Theorem 6.12, p. 347) to obtain the stability conditions.

Let

$$T_0 = 0 < T_1 < T_2 < \cdots,$$

the time instants at which the server becomes idle for the ith time, $i = 0, 1, \ldots$. Note that the T_i, $i = 0, 1, \ldots$, are the time instants at which either a GBP, or a GCT is terminated and as a result, no P_1 customers are waiting in the ordinary queue (the retrial box is not necessary idle). Define $N_{2i} = N_2(T_i+)$, $i = 0, 1, 2, \ldots$, to be the number of P_2 customers just after T_i. Then $Y = \{N_{2i}, i = 0, 1, \ldots\}$ is an irreducible, aperiodic Markov chain.

Theorem 3 *If $\rho < 1$, then Y is positive recurrent.*

Proof Subject to Pakes (1969) theorem:

An irreducible and aperiodic Markov chain $(Y_n;\ n \ge 0)$, with state space the nonnegative integers, is positive recurrent if $|\delta_k| < \infty$ for all $k = 0, 1, 2, \ldots$ and $\limsup_{k\to\infty} \delta_k < 0$, *where* $\delta_k = E[Y_{n+1} - Y_n \mid Y_n = k]$.

Let us define for $m = -1, 0, 1, \ldots,$

$$h_{km}(t)\,dt = P\big(t < T_{n+1} - T_n \le t + dt, N_{2n+1} - N_{2n} = m | N_{2n} = k\big),$$

$$h_k^*(s,z) = \sum_{k=0}^{\infty} z^k \int_0^\infty e^{-st} h_{km}(t)\,dt.$$

Then

$$h_{km}(t) = \big[\lambda_{10} e^{-(\lambda_0 + k\alpha)t} w_m^{(1)}(t) + \lambda_{20} e^{-(\lambda_0+k\alpha)t} w_m^{(2)}(t) \\ + k\alpha e^{-(\lambda_0+k\alpha)t} w_{m+1}^{(2)}(t)\big]\delta_{\{m=0,1,\ldots\}} + k\alpha e^{-(\lambda_0+k\alpha)t} w_0^{(2)}(t)\delta_{\{m=-1\}},$$

where $\lambda_0 = \lambda_{10} + \lambda_{20}$. After manipulations

$$h_k^*(s,z) = \frac{\lambda_{10} w_1^*(s,z) + \lambda_{20} w_2^*(s,z) + \frac{k\alpha}{z} w_2^*(s,z)}{s + \lambda_0 + k\alpha}. \tag{10}$$

Differentiating (10) with respect to z at the point $z = 1,\ s = 0$ we arrive, for $k = 0, 1, \ldots$ at

$$\delta_k = \frac{\partial}{\partial z} h_k^*(s,z)|_{z=1} = \frac{\lambda_{10} E(N(W_1)) + \lambda_{20} E(N(W_2)) + k\alpha(E(N(W_1)) - 1)}{\lambda_0 + k\alpha},$$

where $E(N(W_1)),\ E(N(W_2))$, have been found in (6).

Thus for $\rho < 1$ we realize that $|\delta_k|$ is finite for all k and also $\limsup_{k\to\infty} \delta_k = E(N(W_2)) - 1 = \frac{\rho_2+\rho_4}{1-\rho_1} - 1 < 0$, and the theorem is satisfied. So, for $\rho < 1$ the steady state probabilities $q_k,\ k = 0, 1, \ldots,$ of the Markov chain Y exists and form a distribution. □

For a stochastic process $(Y(t);\ t \geq 0)$ we will say that it is stable, if its limiting probabilities as $t \to \infty$ exist and form a distribution. Consider the stochastic process

$$\mathbf{X} = \big\{\big(N_1(t), N_2(t), \xi_t\big),\ t \geq 0\big\}.$$

The following theorem gives the sufficient condition for $\mathbf{X}$ to be stable.

Theorem 4 *For $\rho < 1$ the process $\mathbf{X}$ is stable.*

Proof Consider the quantities

$$d_k = E(T_1 |\ N_{20} = k).$$

Differentiating (10) with respect to s (at $z = 1,\ s = 0$) we obtain

$$d_k = \frac{\lambda_{10} E(W_1) + \lambda_{20} E(W_2) + k\alpha E(W_2) + 1}{\lambda_0 + k\alpha}.$$

Then

$$\mathbf{q} \cdot \mathbf{d} = \sum_{k=0}^{\infty} q_k d_k = E(W_2) + \big[1 + \lambda_{10}\big[E(W_1) - E(W_2)\big]\big] \sum_{k=0}^{\infty} \frac{q_k}{\lambda_0 + k\alpha}. \tag{11}$$

Now it is clear that there is always a finite integer k^* such that

$$\frac{1}{\lambda_0 + (k^* - 1)\alpha} > 1 > \frac{1}{\lambda_0 + k^*\alpha},$$

and so

$$\sum_{k=0}^{\infty} \frac{q_k}{\lambda_0 + k\alpha} = \sum_{k=0}^{k^*-1} \frac{q_k}{\lambda_0 + k\alpha} + \sum_{k=k^*}^{\infty} \frac{q_k}{\lambda_0 + k\alpha} < \sum_{k=0}^{k^*-1} \frac{q_k}{\lambda_0 + k\alpha} + \sum_{k=k^*}^{\infty} q_k$$

$$= \sum_{k=0}^{k^*-1} \frac{q_k}{\lambda_0 + k\alpha} + \left(1 - \sum_{k=0}^{k^*-1} q_k\right) < \infty.$$

Substituting now (7) to (11) one can understand that $\mathbf{q} \cdot \mathbf{d} < \infty$.

Consider finally the irreducible aperiodic and positive recurrent Markov Renewal Process $\{N,\ T\} = \{(N_{2n},\ T_n)\colon n = 0, 1, 2, \ldots\}$. It is easy to see that the stochastic process $\mathbf{X}$ is a Semi-Regenerative Process with embedded Markov Renewal Process $\{N,\ T\}$ and as, for $\rho < 1$, $\mathbf{q} \cdot \mathbf{d} < \infty$ it is clear that $\mathbf{X}$ is, for $\rho < 1$, stable (Cinlar 1975, Theorem 6.12, p. 347). □

5 Steady state analysis

Let us assume that a state for statistical equilibrium exists for our model, so that $\rho < 1$. Define $N_i = \lim_{t\to\infty} N_i(t)$, $i = 1, 2$, $\xi = \lim_{t\to\infty} \xi_t$, $u = \lim_{t\to\infty} u_t$, $\sigma_i(z_1, z_2) = \sigma_i(z_1, z_2, 0)$, $\sigma_2(z_2) = \sigma_2(z_2, 0)$. Let also the generating functions

$$\begin{aligned}
P_i(z_1, z_2, x) &= \sum_{k_1\geq 0}\sum_{k_2\geq 0} p_i(k_1, k_2, x) z_1^{k_1} z_2^{k_2}\, dx,\\
P_2(z_2, x) &= \sum_{k_2\geq 0} p_2(k_2, x) z_2^{k_2}\, dx,\\
P_{i1}(z_1, z_2, x, y) &= \sum_{k_1\geq 0}\sum_{k_2\geq 0} p_{i1}(k_1, k_2, x, y) z_1^{k_1} z_2^{k_2} dx,\\
Q(z_2) &= \sum_{k_2\geq 0} q(k_2) z_2^{k_2}.
\end{aligned}$$

By applying the supplementary variable method we obtain the following equations that govern the dynamics of the system:

$$\begin{aligned}
&\frac{\partial}{\partial x} p_i(k_1, k_2, x) + p_i(k_1, k_2, x)\big[\lambda_{1i} + \lambda_{2i} + \eta_i(x) + \delta_{\{i=1\}}(\mu + \nu)\big]\\
&\quad = \lambda_{1i} p_i(k_1 - 1, k_2, x) + \lambda_{2i} p_i(k_1, k_2 - 1, x), \quad i = 1, 3, 4,\\
&\frac{\partial}{\partial x} p_2(k_2, x) + p_2(k_2, x)\big[\lambda_{12} + \lambda_{22} + \eta_2(x) + \mu + \nu\big] \qquad (12)\\
&\quad = \int_0^\infty p_{11}(0, k_2, y, x)\eta_1(y)\, dy + \nu \int_0^\infty p_{11}(0, k_2, y, x)\, dy\\
&\qquad + \lambda_{22} p_2(k_2 - 1, x) + \int_0^\infty p_{31}(0, k_2, y, x)\eta_3(y)\, dy,\\
&\frac{\partial}{\partial x} p_{i1}(k_1, k_2, x, y) + p_{i1}(k_1, k_2, x, y)\big[\lambda_{1i} + \lambda_{2i} + \eta_i(x) + \delta_{\{i=1\}}(\mu + \nu)\big]\\
&\quad = \lambda_{1i} p_{i1}(k_1 - 1, k_2, x, y) + \lambda_{2i} p_{i1}(k_1, k_2 - 1, x, y), \quad i = 1, 3, \qquad (13)
\end{aligned}$$

$$(\lambda_0 + k_2\alpha)q(k_2) = \int_0^\infty p_4(0, k_2, x)\eta_4(x)\,dx.$$

The boundary conditions are given by

$$\begin{aligned} p_{11}(k_1, k_2, 0, x) &= \lambda_{12} p_2(k_2, x)\delta_{\{k_1=0\}} + \int_0^\infty p_{11}(k_1+1, k_2, y, x)\eta_1(y)\,dy \\ &\quad + \int_0^\infty p_{31}(k_1+1, k_2, y, x)\eta_3(y)\,dy \\ &\quad + \nu\int_0^\infty p_{11}(k_1+1, k_2, y, x)\,dy, \end{aligned} \tag{14}$$

$$\begin{aligned} p_{31}(k_1, k_2, 0, x) &= \mu\int_0^\infty p_{11}(k_1, k_2-1, y, x)\,dy. \\ p_2(k_2, 0) &= \lambda_{20} q(k_2) + \alpha(k_2+1)q(k_2+1), \\ p_3(k_1, k_2, 0) &= \mu\left[\int_0^\infty p_1(k_1, k_2-1, x)\,dx + \int_0^\infty p_2(k_2-1, x)\,dx\,\delta_{\{k_1=0\}}\right], \end{aligned}$$

$$\begin{aligned} p_4(0, k_2, 0) &= \int_0^\infty p_1(0, k_2, x)\eta_1(x)\,dx \\ &\quad + \nu\int_0^\infty p_1(0, k_2, x)\,dx + \int_0^\infty p_2(k_2, x)\eta_2(x)\,dx \\ &\quad + \int_0^\infty p_3(0, k_2, x)\eta_3(x)\,dx + \nu\int_0^\infty p_2(k_2, x)\,dx, \end{aligned} \tag{15}$$

$$\begin{aligned} p_1(k_1, k_2, 0) &= \lambda_{10} q(k_2)\delta_{\{k_1=0\}} + \int_0^\infty p_1(k_1+1, k_2, x)\eta_1(x)\,dx \\ &\quad + \int_0^\infty p_3(k_1+1, k_2, x)\eta_3(x)\,dx \\ &\quad + \nu\int_0^\infty p_1(k_1+1, k_2, x)\,dx \\ &\quad + \int_0^\infty p_4(k_1+1, k_2, x)\eta_4(x)\,dx, \end{aligned} \tag{16}$$

where $\delta_{\{\}}$ is Kronecker's delta. Forming the generating functions, we obtain

$$\begin{aligned} &P_i(z_1, z_2, x) = P_i(z_1, z_2, 0)\big(1 - B_i(x)\big)\exp\big[-\sigma_i(z_1, z_2)x\big], \quad i = 1, 3, 4, \\ &P_{i1}(z_1, z_2, x, y) = P_{i1}(z_1, z_2, 0, y)\big(1 - B_i(x)\big)\exp\big[-\sigma_i(z_1, z_2)x\big], \quad i = 1, 3, \\ &\lambda_0 Q(z_2) + \alpha z_2\frac{\partial}{\partial z_2}Q(z_2) = P_4(0, z_2, 0)\beta_4^*\big(\lambda_{14} + \lambda_{24}(1 - z_2)\big). \end{aligned} \tag{17}$$

From (14),

$$P_{11}(z_1,z_2,0,x)=\frac{\lambda_{12}z_1P_2(z_2,x)-R(z_2,x)}{z_1-a_1^*(z_1,z_2,0)}, \tag{18}$$

where

$$R(z_2,x)=\int_0^\infty P_{11}(0,z_2,y,x)\eta_1(y)\,dy+\int_0^\infty P_{31}(0,z_2,y,x)\eta_3(y)\,dy \\ +\nu\int_0^\infty P_{11}(0,z_2,y,x)\,dy.$$

Then by substituting the only root, say $x(z_2)\equiv x(0,z_2)$, in $|z_1|\leq 1$ of the equation $z_1-a_1^*(z_1,z_2,0)$, in the numerator of (18), we obtain

$$R(z_2,x)=\lambda_{12}x(z_2)P(z_2,x). \tag{19}$$

Using (19), in (12), we obtain

$$P_2(z_2,x)=P_2(z_2,0)\bigl(1-B_2(x)\bigr)\exp\bigl(-\sigma_2(z_2)x\bigr). \tag{20}$$

The generating functions for the boundary conditions are given by

$$\begin{aligned} P_{11}(z_1,z_2,0,x)&=\frac{\lambda_{12}(z_1-x(z_2))}{z_1-a_1^*(z_1,z_2,0)}P_2(z_2,x),\\ P_2(z_2,0)&=\lambda_{20}Q(z_2)+\alpha\frac{\partial}{\partial z_2}Q(z_2),\\ P_3(z_1,z_2,0)&=\mu z_2\bigl[P_1(z_1,z_2)+P_2(z_2)\bigr],\\ P_{31}(z_1,z_2,0,x)&=\mu z_2P_{11}(z_1,z_2,0,x)\frac{1-\beta_1^*(\sigma_1(z_1,z_2))}{\sigma_1(z_1,z_2)}. \end{aligned} \tag{21}$$

Moreover

$$\begin{aligned} P_1(z_1,z_2,0)=&\bigl\{\lambda_{10}z_1Q(z_2)+P_2(z_2,0)a_2^*(z_1,z_2,0)\\ &-P_4(0,z_2,0)\bigl[1-\beta_4^*\bigl(\sigma_4\bigl(z_1,z_2\bigr)\bigr)+\beta_4^*\bigl(\sigma_4(0,z_2)\bigr)\bigr]\bigr\}\\ &\times\bigl\{z_1-a_1^*(z_1,z_2,0)\bigr\}^{-1}. \end{aligned} \tag{22}$$

Using the root $x(z_2)$, in $|z_1|\leq 1$, of the equation $z_1-a_1^*(z_1,z_2,0)$ we arrive at

$$P_4(0,z_2,0)=\frac{(\lambda_{10}x(z_2)+\lambda_{20}c^*(0,z_2))Q(z_2)+\alpha\frac{\partial}{\partial z_2}Q(z_2)c^*(0,z_2)}{1-\beta_4^*(\sigma_4(x(z_2),z_2))+\beta_4^*(\sigma_4(0,z_2))}. \tag{23}$$

Substitute (23), in the last of (17) we obtain

$$\alpha\bigl(z_2-w_2^*(0,z_2)\bigr)\frac{\partial}{\partial z_2}Q(z_2)+\bigl[\lambda_0-\lambda_{10}w_1^*(0,z_2)-\lambda_{20}w_2^*(0,z_2)\bigr]Q(z_2)=0. \tag{24}$$

In order to solve the differential equation (24) we have to use Theorem 2. According to Theorem 2, for $\rho < 1$ equation $z_2 - w_2^*(0, z_2) = 0$ never become zero in the unit disk $|z_2| < 1$. Let now,

$$\psi(z_2) = \frac{\lambda_0 - \lambda_{10} w_1^*(0, z_2) - \lambda_{20} w_2^*(0, z_2)}{z_2 - w_2^*(0, z_2)}.$$

Thus, $\psi(z_2)$ is analytic function in $|z_2| < 1$, and also a continuous one on the boundary because of

$$\lim_{z_2 \to 1} \psi(z_2) = \frac{-\lambda_{10}\rho_d - \lambda_{20}\rho_w}{1 - \rho} < \infty.$$

As a result, (24), can be solved for $|z_2| \le 1$ and

$$Q(z_2) = Q(1) \exp\left\{ - \int_{z_2}^{1} \frac{\lambda_0 - \lambda_{10} w_1^*(0, u) - \lambda_{20} w_2^*(0, u)}{\alpha[w_2^*(0, u) - u]} du \right\}. \tag{25}$$

By substituting (25), all above defined probability generating functions in steady state, are completely known. Setting $z_2 = 1$, to all probability generating functions and asking the total probability to sum in unity we obtain

$$Q(1) = \frac{1 - \rho}{A}, \tag{26}$$

where

$$A = (1 - \rho)\big[1 + \lambda_{10} E(W_1)\big] + \big[\lambda_{10}\rho_d + \lambda_{20}(1 - \rho_1)\big] E(W_2).$$

The following theorem shows that the condition $\rho < 1$ is also necessary for a stable system.

Theorem 5 *If the system is stable, then* $\rho < 1$.

Proof Suppose that the system is stable and $\rho > 1$. Then from Theorem 2 the equation $z_2 - w_2^*(0, z_2) = 0$ has a root strictly less than one ($\phi(0) < 1$) and so $\lambda_0 - \lambda_{10} w_1^*(0, \phi(0)) - \lambda_{20} w_2^*(0, \phi(0)) \neq 0$. By putting now $\phi(0)$ instead of z_2 in (24) we obtain

$$\big[\lambda_0 - \lambda_{10} w_1^*\big(0, \phi(0)\big) - \lambda_{20} w_2^*\big(0, \phi(0)\big)\big] Q\big(\phi(0)\big) = 0,$$

and so $Q(\phi(0)) = \sum q(j)\phi^j(0) = 0$ with $0 < \phi(0) < 1$. Thus $q(j) = 0\ \forall j$ and also from the generating functions in (17)–(23) it is clear that all probabilities become zero. This of course contradicts the hypothesis that the system is stable.

Suppose finally that the system is stable and $\rho = 1$. Differentiating (24) with respect to z_2 (at $z_2 = 1$) we arrive (for $\rho = 1$) at

$$\begin{aligned} &\frac{d}{dz_2}\big[\lambda_0 - \lambda_{10} w_1^*(0, z_2) - \lambda_{20} w_2^*(0, z_2)\big]\big|_{z_2=1} Q(1) \\ &\quad = -\big[\lambda_{10} E\big(N(W_1)\big) + \lambda_{20} E\big(N(W_2)\big) + \lambda_0 E\big(N(V)\big)\big] Q(1) = 0, \end{aligned}$$

and so $Q(1)=\sum q(j)=0$ and this again contradicts the hypothesis that the system is stable. $\square$

6 Stochastic decomposition law

The stochastic decomposition property was first established by Furhmand and Cooper (1985) for the M/G/1 queue with generalized vacations, while an analogous result for retrial queues was presented by Falin et al. (1993). Shanthikumar (1988), generalized the result by Furhmand and Cooper (1985) on the decomposition property in M/G/1 queue by assuming the arrival process to be different at certain random periods. In this Section, we establish a decomposition property for our retrial model with varying arrival rates according to server state.

Let us allow the retrial rate $\alpha\to\infty$. In this case, the modified model consist of two parallel queues, one of which have preemptive resume priority over the other one. A control on the arrivals has been made, by varying the arrival rates according to server's state.

When a P_1customer arrives during the service of a P_2 customer, he immediately interrupts him and forces the server to start serving him. The service of the interrupted P_2 customer is resumed from the point of interruption as soon as the server become available. The server departs for single vacation each time, upon finishing the service of a P_1 or a P_2 customer or after a repair completion or after a negative arrival, he faces an empty priority queue (the non-priority queue is not necessary idle). The server starts serving in non-priority queue in a preemptive resume basis only when upon returning from a vacation he finds the priority queue empty. The model is subject to breakdowns and repairs. When a breakdown occurs, the server is sent for repair and the customer being served joins the non-priority queue. Upon returning from repair, the server start serving in the priority queue (if not empty). If priority queue is empty it departs for single vacation.

Let us denote by $\hat{N}_i$, $i=1,2$, to be the number of P_i, customers in steady state for the modified model and by $\hat{P}_i(.)$, $i=1,2,3,4$, $\hat{P}_{i1}$, $i=1,3$, the corresponding probability generating functions. Clearly,

$$\begin{aligned}
\hat{P}_1(z_1,z_2)=&\frac{(1-\beta_1^*(\sigma_1(z_1,z_2)))q_0}{\sigma_1(z_1,z_2)(z_1-a_1^*(z_1,z_2,0))}\\
&\times\left\{z_1-x(z_2)\frac{1-\beta_4^*(\sigma_4(z_1,z_2))+\beta_4^*(\sigma_4(0,z_2))}{1-\beta_4^*(\sigma_4(x(z_2),z_2))+\beta_4^*(\sigma_4(0,z_2))}\right\}\\
&+\frac{\lambda_0-\lambda_{10}w_1^*(0,z_2)-\lambda_{20}z_2}{w_2^*(0,z_2)-z_2}\\
&\times\left[a_2^*(z_1,z_2,0)-\frac{c^*(0,z_2)[1-\beta_4^*(\sigma_4(z_1,z_2))+\beta_4^*(\sigma_4(0,z_2))]}{1-\beta_4^*(\sigma_4(x(z_2),z_2))+\beta_4^*(\sigma_4(0,z_2))}\right],
\end{aligned}$$

$$\hat{P}_2(z_2)=\frac{(\lambda_0-\lambda_{10}w_1^*(0,z_2)-\lambda_{20}z_2)(1-\beta_2^*(\sigma_2(z_2)))}{(w_2^*(0,z_2)-z_2)\sigma_2(z_2)}q_0,$$

$$\hat{P}_3(z_1,z_2) = \mu z_2\big[\hat{P}_1(z_1,z_2) + \hat{P}_2(z_2)\big]\frac{1-\beta_3^*(\sigma_3(z_1,z_2))}{\sigma_3(z_1,z_2)},$$

$$\hat{P}_{11}(z_1,z_2) = \frac{\lambda_{12}(z_1 - x(z_2))\hat{P}_2(z_2)}{z_1 - a_1^*(z_1,z_2,0)}\left(\frac{1-\beta_1^*(\sigma_1(z_1,z_2))}{\sigma_1(z_1,z_2)}\right),$$

$$\hat{P}_{31}(z_1,z_2) = \mu z_2\hat{P}_{11}(z_1,z_2)\frac{1-\beta_3^*(\sigma_3(z_1,z_2))}{\sigma_3(z_1,z_2)},$$

$$\hat{P}_4(z_1,z_2) = \frac{(1-\beta_4^*(\sigma_4(z_1,z_2)))q_0[\lambda_{10}x(z_2) + \frac{\lambda_0-\lambda_{10}w_1^*(0,z_2)-\lambda_{20}z_2}{w_2^*(0,z_2)-z_2}c^*(0,z_2)]}{\sigma_4(z_1,z_2)[1-\beta_4^*(\sigma_4(x(z_2),z_2)) + \beta_4^*(\sigma_4(0,z_2))]},$$

where $q_0 = Q(1)$. Note that the probability generating functions of our retrial model are given by

$$Q(z_2) = q_0 Q_\alpha(z_2), \qquad P_2(z_2) = \hat{P}_i(z_2)Q_\alpha(z_2),$$
$$P_{i1}(z_1,z_2) = \hat{P}_{i1}(z_1,z_2)Q_\alpha(z_2), \quad i = 1,3,$$
$$P_i(z_1,z_2) = \hat{P}_i(z_1,z_2)Q_\alpha(z_2), \quad i = 1,3,4,$$

where

$$Q_\alpha(z_2) = \exp\left\{-\int_{z_2}^1 \frac{\lambda_0 - \lambda_{10}w_1^*(0,u) - \lambda_{20}w_2^*(0,u)}{\alpha[w_2^*(0,u)-u]}\,du\right\},$$

the generating function of the number of customers in orbit given that the server is idle. According to the above result our model satisfy the stochastic decomposition property. This outcome can be summed up in the following theorem.

Theorem 6 *The number of customers in priority and non-priority queue and the server's state* (N_1, N_2, ξ) *for the system under study can be represented as the sum of two independent random variables, one of which is the number of customers in priority and non-priority queue and the server's state in the non retrial model with failures, repairs and negative customers where arrival rates depend on the server state* $(\hat{N}_1, \hat{N}_2, \hat{\xi})$ *and the other is the number of the customers in orbit given that the server is idle* $(0, Q_\alpha, 0)$. *That is* $(N_1, N_2, \xi) = (\hat{N}_1, \hat{N}_2, \hat{\xi}) + (0, Q_\alpha, 0)$.

7 Performance measures

In this sections we are going to obtain some measures that describes the evolution of or model. Firstly, by setting $z_1 = z_2 = 1$ in the probability generating functions (17)–(23), we derive the probabilities of server's state. After the manipulations

$$P(\xi = 0) = Q(1) = \frac{1-\rho}{A},$$

$$P(\xi = 1, u = 0) = P_1(1,1) = \frac{1-\beta_1^*(\mu+\nu)}{(\mu+\nu)A(1-\rho_1)}\left\{\lambda_{10}(1-\rho)\left(1+\frac{\lambda_{14}\bar{b}_4}{\beta_4^*(\lambda_{14})}\right)\right.$$

$$+\big(\lambda_{10}\rho_d+\lambda_{20}(1-\rho_1)\big)$$
$$\times\left[\frac{\lambda_{13}\mu\bar{b}_3(1-\beta_2^*(\mu+\nu))}{\mu+\nu}+\frac{\lambda_{14}\bar{b}_4}{\beta_4^*(\lambda_{14})}\right]\Bigg\},$$

$$P(\xi=2)=P_2(1)=\frac{\lambda_{10}\rho_d+\lambda_{20}(1-\rho_1)}{A}\left(\frac{1-\beta_2^*(\mu+\nu)}{\mu+\nu}\right),$$

$$P(\xi=4)=P_4(1,1)=\left(\frac{\lambda_{10}(1-\rho+\rho_d)+\lambda_{20}(1-\rho_1)}{A}\right)\frac{\bar{b}_4}{\beta_4^*(\lambda_{14})},$$

$$P(\xi=1,u=1)=P_{11}(1,1)=\frac{\lambda_{12}}{1-\rho_1}P_2(1)\frac{1-\beta_1^*(\mu+\nu)}{\mu+\nu},$$

$$P(\xi=3,u=1)=P_{31}(1,1)=\mu\bar{b}_3P_{11}(1,1),$$

$$P(\xi=3,u=0)=P_3(1,1)=\mu\bar{b}_3\big[P_1(1,1)+P_2(1)\big].$$

In the sequel we give in closed form the mean number of P_2 customers in steady state. These expressions can be derived after heavy manipulations, by differentiating the probability generating functions (17)–(23), with respect to z_2, at the point $z_1=z_2=1$. Then

$$E(N_2;\xi=0)=\frac{\lambda_{10}\rho_d+\lambda_{20}\rho_w}{\alpha(1-\rho)},$$

$$E(N_2;\xi=2)=\frac{(1-\beta_2^*(\mu+\nu))(M+\lambda_{20}E(N_2;\xi=0))}{\mu+\nu}$$
$$+\frac{P_2(1)(\lambda_{12}m_1+\lambda_{22})(\dot{\beta}_2^*(\mu+\nu)+\frac{1-\beta_2^*(\mu+\nu)}{\mu+\nu})}{1-\beta_2^*(\mu+\nu)},$$

$$E(N_2;\xi=1,u=1)=\frac{\lambda_{12}P_2(1)(1-\beta_1^*(\mu+\nu))}{2L_1^2(\mu+\nu)}S$$
$$+\frac{\lambda_{12}E(N_2;\xi=2)(1-\beta_1^*(\mu+\nu))}{(1-\rho_1)(\mu+\nu)}$$
$$+\frac{\lambda_{12}P_2(1)\lambda_{21}(\dot{\beta}_1^*(\mu+\nu)+\frac{1-\beta_1^*(\mu+\nu)}{\mu+\nu})}{(1-\rho_1)(\mu+\nu)},$$

$$E(N_2;\xi=3,u=1)=\mu\bar{b}_3E(N_2;\xi=1,u=1)+\mu P_{11}(1,1)\left(\bar{b}_3+\frac{\lambda_{23}\bar{b}_3^{(2)}}{2}\right),$$

$$E(N_2;\xi=4)=\bar{b}_4G+\lambda_{24}\bar{b}_4^{(2)}\frac{P_4(1,1)}{2\bar{b}_4},$$

$$E(N_2;\xi=1,u=0)$$
$$=\frac{1-\beta_1^*(\mu+\nu)}{\mu+\nu}\Bigg\{\frac{\lambda_{10}E(N_2;\xi=0)}{1-\rho_1}+\frac{\lambda_{10}Q(1)S}{2L_1^2}+\frac{\lambda_{13}\mu\bar{b}_3E(N_2;\xi=2)}{1-\rho_1}$$

$$+ P_2(1)\left\{\frac{\lambda_{13}\mu\bar{b}_3}{1-\rho_1} + \frac{\mu[\lambda_{13}m_1(\bar{b}_3 + (\lambda_{13}m_1 + 2\lambda_{23})\bar{b}_3^{(2)}) + \frac{\lambda_{13}\bar{b}_3 S}{L_1}]}{2L_1}\right\}$$

$$+ \frac{\lambda_{14}\bar{b}_4}{1-\rho_1}G + \frac{P_4(1)[\lambda_{14}m_1(\lambda_{14}m_1 + 2\lambda_{24})\bar{b}_4^{(2)}) + \frac{\lambda_{14}\bar{b}_4 S}{L_1}]}{2\bar{b}_4 L_1}\Bigg\}$$

$$+ \frac{\lambda_{21}P_1(1,1)(\dot{\beta}_1^*(\mu+\nu) + \frac{1-\beta_1^*(\mu+\nu)}{\mu+\nu})}{1-\beta_1^*(\mu+\nu)},$$

$$E(N_2;\xi=3,u=0) = \mu\bar{b}_3\big[E(N_2;\xi=1,u=0) + E(N_2;\xi=2)\big]$$

$$+ \mu\big[P_1(1,1) + P_1(1)\big]\left(b_3 + \frac{\lambda_{23}\bar{b}_3^{(2)}}{2}\right),$$

where

$$S = \frac{2m_1(\dot{\beta}_1^*(\mu+\nu) + \frac{1-\beta_1^*(\mu+\nu)}{\mu+\nu})(\lambda_{21} + \mu(1+\lambda_{23}\bar{b}_3))(\lambda_{11}m_1 + \lambda_{21}\rho_1)}{(\mu+\nu)(1-\rho_1)}$$

$$+ \frac{\lambda_{13}m_1^2(1-\beta_1^*(\mu+\nu))[2\mu\bar{b}_3 + \mu\bar{b}_3^{(2)}(\lambda_{13}m_1 + 2\lambda_{23})]}{\mu+\nu},$$

$$G = \frac{M + E(N_2;\xi=0)(\lambda_0 + \alpha\tilde{\rho}_2) + Q(1)(\lambda_{10}m_1 + \lambda_{20}\tilde{\rho}_2)}{\beta_4^*(\lambda_{14})}$$

$$+ \frac{\frac{P_4(1,1)}{\bar{b}_4}[(\lambda_{14}m_1 + \lambda_{24})\bar{b}_4 + \lambda_{24}\dot{\beta}(\lambda_{14})]}{\beta_4^*(\lambda_{14})},$$

$$M = \frac{Q(1)(1-\rho_1)[\lambda_{10}\tilde{\rho}_d^{(2)} + \lambda_{20}\tilde{\rho}_w^{(2)}]}{2(1-\rho)}$$

$$+ \frac{(\lambda_{10}\rho_d + \lambda_{20}\rho_w)[E(N_2;\xi=0) + \frac{Q(1)(1-\rho_1)\tilde{\rho}_w^{(2)}}{2(1-\rho)}}{1-\rho}.$$

$$m_2 = \frac{2(\lambda_{11}m_1 + \lambda_{21})(\dot{\beta}_1^*(\mu+\nu) + \frac{1-\beta_1^*(\mu+\nu)}{\mu+\nu})(\lambda_{21} + \mu(1+\lambda_{23}\bar{b}_3))}{(\mu+\nu)(1-\rho_1)^2}$$

$$+ \frac{(\lambda_{13}m_1 + \lambda_{23})(1-\beta_1^*(\mu+\nu))[2\mu\bar{b}_3 + \mu\bar{b}_3^{(2)}(\lambda_{13}m_1 + \lambda_{23})]}{(1-\rho_1)(\mu+\nu)},$$

$$\tilde{\rho}_d^{(2)} = m_2 + 2m_1\tilde{\rho}_4 + \tilde{\rho}_4^{(2)}, \qquad \tilde{\rho}_w^{(2)} = \tilde{\rho}^{(2)} + 2\tilde{\rho}_2\tilde{\rho}_4 + \tilde{\rho}_4^{(2)},$$

$$\tilde{\rho}_2^{(2)} = \frac{(\lambda_{12} + \lambda_{13}\mu\bar{b}_3)(1-\beta_2^*(\mu+\nu))m_2}{\mu+\nu}$$

$$+ \frac{(\lambda_{13}m_1 + \lambda_{23})(1-\beta_2^*(\mu+\nu))[2\mu\bar{b}_3 + \mu\bar{b}_3^{(2)}(\lambda_{13}m_1 + \lambda_{23})]}{\mu+\nu}$$

$$+\left\{2(\lambda_{12}m_1+\lambda_{22})\left(\dot{\beta}_2^*(\mu+\nu)+\frac{1-\beta_2^*(\mu+\nu)}{\mu+\nu}\right)\right.$$
$$\left.\times\left((\lambda_{12}+\lambda_{13}\mu\bar{b}_3)m_1+\lambda_{21}+\mu(1+\lambda_{23}\bar{b}_3)\right)\right\}\left\{(\mu+\nu)\right\}^{-1},$$

$$\tilde{\rho}_4^{(2)}=\frac{\lambda_{14}\bar{b}_4}{\beta_4^*(\lambda_{14})}m_2+\frac{(\lambda_{14}m_1+\lambda_{24})^2\bar{b}_4^{(2)}}{\beta_4^*(\lambda_{14})}$$
$$+\frac{2(\lambda_{14}m_1+\lambda_{24})\bar{b}_4[(\lambda_{14}m_1+\lambda_{24})\bar{b}_4+\lambda_{24}\dot{\beta}_4^*(\lambda_{14})]}{(\beta_4^*(\lambda_{14}))^2},$$

where $\dot{\beta}^*(.)$ is the first order derivative of the function $\beta^*(.)$.

8 Reliability analysis

This section discusses some reliability indices of the model under consideration. Specifically, we analyze the availability of the server, the failure frequency and the time to first failure of the server. Let $AV(t)$ be the availability of the server at time t, that is, the probability that the server is either serving a customer or is in the idle period or during a single vacation. Define the steady state ($\rho<1$) availability by $AV=\lim_{t\to\infty}AV(t)$. Then,

Theorem 7 *The steady state availability of the server is* $AV=1-\mu\bar{b}_3[P_1(1,1)+P_2(1)(1+\frac{\lambda_{12}(1-\beta_1^*(\mu+\nu))}{(1-\rho_1)(\mu+\nu)})]$.

Proof: This is readily obtained by considering that

$$AV=1-P_3(1,1)-P_{31}(1,1). \tag{27}$$

Corollary 8 *The steady state failure frequency of the server is* $F=\mu[P_1(1,1)+P_2(1)(1+\frac{\lambda_{12}(1-\beta_1^*(\mu+\nu))}{(1-\rho_1)(\mu+\nu)})]$.

The result follows directly by

$$F=\mu\left\{\int_0^\infty\left[P_1(1,1,x)+P_2(1,x)\right]dx+\int_0^\infty\int_0^\infty P_{11}(1,1,x,y)\,dx\,dy\right\}.$$

Let us assume that the system is empty at time $t=0$. That is $q(0,0)=1$. Denote by τ to be the time to first failure of the server. Then the reliability function of the server is $U(t)=P(\tau>t)$. In order to find $U(t)$ we regard a new system where the failure states of the server are assumed to be absorbent states. In the new system we use the same notations as in Sect. 2. Then we can get the following set of equations:

$$\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial t}+\lambda_{1i}+\lambda_{2i}+\eta_i(x)+\delta_{\{i=1\}}(\mu+\nu)\right)p_i(k_1,k_2,x,t)$$

$$
\begin{aligned}
&= \lambda_{1i} p_i(k_1-1,k_2,x,t) + \lambda_{2i} p_i(k_1,k_2-1,x,t), \quad i=1,4,\\
&\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial t} + \lambda_{12} + \lambda_{22} + \eta_2(x) + \mu + \nu\right) p_2(k_2,x,t)\\
&= \lambda_{22} p_2(k_2-1,x,t)\\
&\quad + \int_0^\infty p_{11}(0,k_2,y,x,t)\eta_1(y)\,dy + \nu \int_0^\infty p_{11}(0,k_2,y,x,t)\,dy, \qquad (28)\\
&\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial t} + \lambda_{11} + \lambda_{21} + \eta_1(x) + \mu + \nu\right) p_{11}(k_1,k_2,x,t)\\
&= \lambda_{11} p_{11}(k_1-1,k_2,x,t) + \lambda_{21} p_{11}(k_1,k_2-1,x,t),\\
&\left(\frac{\partial}{\partial t} + \lambda_0 + k_2\alpha\right) q(k_2,t) = \int_0^\infty p_4(0,k_2,x,t)\eta_4(x)\,dx.
\end{aligned}
$$

$$
\begin{aligned}
p_{11}(k_1,k_2,0,x,t) &= \lambda_{12} p_2(k_2,x,t)\delta_{\{k_1=0\}} + \int_0^\infty p_{11}(k_1{+}1,k_2,y,x,t)\eta_1(y)\,dy\\
&\quad + \nu \int_0^\infty p_{11}(k_1+1,k_2,y,x,t)\,dy,\\
p_2(k_2,0,t) &= \lambda_{20} q(k_2,t) + \alpha(k_2+1)q(k_2+1,t), \qquad (29)\\
p_4(0,k_2,0,t) &= \int_0^\infty p_1(0,k_2,x,t)\eta_1(x)\,dx + \nu \int_0^\infty p_1(0,k_2,x,t)\,dx\\
&\quad + \int_0^\infty p_2(k_2,x,t)\eta_2(x)\,dx, +\nu \int_0^\infty p_2(k_2,x,t)\,dx,\\
p_1(k_1,k_2,0,t) &= \lambda_{10} q(k_2,t)\delta_{\{k_1=0\}} + \int_0^\infty p_1(k_1+1,k_2,x,t)\eta_1(x)\,dx\\
&\quad + \nu \int_0^\infty p_1(k_1+1,k_2,x,t)\,dx\\
&\quad + \int_0^\infty p_4(k_1+1,k_2,x,t)\eta_4(x)\,dx. \qquad (30)
\end{aligned}
$$

Consider now the LST of the generating functions

$$
P_i^*(z_1,z_2,x,s) = \sum_{k_1\ge 0}\sum_{k_2\ge 0} \int_0^\infty e^{-st} p_i(k_1,k_2,x,t) z_1^{k_1} z_2^{k_2}\,dt,
$$

$$
P_2^*(z_2,x,s) = \sum_{k_2\ge 0} \int_0^\infty e^{-st} p_2(k_2,x,t) z_2^{k_2}\,dt,
$$

$$
P_{11}^*(z_1,z_2,x,y,s) = \sum_{k_1\ge 0}\sum_{k_2\ge 0} \int_0^\infty e^{-st} p_{11}(k_1,k_2,x,y,t) z_1^{k_1} z_2^{k_2}\,dt,
$$

$$Q^*(z_2,s)=\sum_{k_2\geq 0}\int_0^\infty e^{-st}q(k_2,t)z_2^{k_2}\,dt.$$

After manipulations

$$\begin{aligned}
P_i^*(z_1,z_2,s)&=P_i^*(z_1,z_2,0,s)\frac{1-\beta_i^*(\sigma_i(z_1,z_2,s))}{\sigma_i(z_1,z_2,s)},\quad i=1,4,\\
P_{11}^*(z_1,z_2,s)&=P_{11}^*(z_1,z_2,0,s)\frac{1-\beta_1^*(\sigma_1(z_1,z_2,s))}{\sigma_1(z_1,z_2,s)},\\
(s+\lambda_0)Q^*(z_2,s)&+\alpha z_2\frac{\partial}{\partial z_2}Q^*(z_2,s)\\
&=1+P_4^*(0,z_2,0,s)\beta_4^*\big(\sigma_4(0,z_2,s)\big).
\end{aligned}\tag{31}$$

Moreover by the first of (29)

$$\begin{aligned}
&\big(z_1-y_1(z_1,z_2,s)\big)P_{11}^*(z_1,z_2,0,x,s)\\
&\quad=\lambda_{12}z_1P_2^*(z_2,x,s)\\
&\qquad-\left[\int_0^\infty P_{11}^*(0,z_2,y,x,s)\eta_1(y)\,dy+\nu\int_0^\infty P_{11}^*(0,z_2,y,x,s)\,dy\right]
\end{aligned}$$

where $y_1(z_1,z_2,s)=\beta_1^*(\sigma_1(z_1,z_2,s))+\nu\frac{1-\beta_1^*(\sigma_1(z_1,z_2,s))}{\sigma_1(z_1,z_2,s)}$. Similar to the derivation in Falin and Templeton (1997) (p. 191) we can show that $z_1-y_1(z_1,z_2,s)=0$ has exactly one root, say $z_1=\omega_1(s,z_2)$, in the disk $|z_1|\leq 1$. Then after some algebra we obtain

$$\begin{aligned}
P_{11}^*(z_1,z_2,0,x,s)&=\frac{\lambda_{12}(z_1-\omega_1(s,z_2))}{z_1-y_1(z_1,z_2,s)}P_2^*(z_2,x,s),\\
P_2^*(z_2,s)&=P_2^*(z_2,0,s)\frac{1-\beta_2^*(\hat{\sigma}_2(z_2,s))}{\hat{\sigma}_2(z_2,s)},
\end{aligned}$$

where $\hat{\sigma}_2(z_2,s)=s+\mu+\nu+\lambda_{12}(1-\omega_1(s,z_2))+\lambda_{22}(1-z_2)$. Moreover

$$\begin{aligned}
P_1^*(z_1,z_2,0,s)=&\big\{\lambda_{10}z_1Q^*(z_2,s)+P_2^*(z_2,0,s)y_2(z_2,s)-P_4^*(0,z_2,0,s)\\
&\times\big[1-\beta_4^*\big(\sigma_4(z_1,z_2,s)\big)+\beta_4^*\big(\sigma_4(0,z_2,s)\big)\big]\big\}\\
&\times\big\{z_1-y_1(z_1,z_2,s)\big\}^{-1},
\end{aligned}$$

where $y_2(z_2,s)=\beta_2^*(\hat{\sigma}_2(z_2,s))+\nu\frac{1-\beta_2^*(\hat{\sigma}_2(z_2,s))}{\hat{\sigma}_2(z_2,s)}$. Then,

$$P_4^*(0,z_2,0,s)=\frac{(\lambda_{10}\omega_1(s,z_2)+\lambda_{20}y_2(z_2,s))Q^*(z_2,s)+\alpha\frac{\partial}{\partial z_2}Q^*(z_2,s)y_2(z_2,s)}{1-\beta_4^*(\sigma_4(\omega_1(s,z_2),z_2,s))+\beta_4^*(\sigma_4(0,z_2,s))}.\tag{32}$$

Replacing (32) in the third of (31) yields

$$\alpha\big(z_2-\hat{f}_2(z_2,s)\big)\frac{\partial}{\partial z_2}Q^*(z_2,s)+\big[s+\lambda_0-\lambda_{10}\hat{f}_1(z_2,s)-\lambda_{20}\hat{f}_2(z_2,s)\big]Q^*(z_2,s)=1, \tag{33}$$

where

$$\hat{f}_1(z_2,s)=\omega_1(s,z_2)\hat{v}(s,z_2), \qquad \hat{f}_2(z_2,s)=y_2(z_2,s)\hat{v}(s,z_2),$$

where

$$\hat{v}(s,z_2)=\frac{\beta_4^*(\sigma_4(0,z_2,s))}{1-\beta_4^*(\sigma_4(\omega_1(s,z_2),z_2,s))+\beta_4^*(\sigma_4(0,z_2,s))}.$$

It is easy to realize that the equation $z_2-\hat{f}_2(z_2,s)=0$ has exactly one root, say $z_2=\omega_2(s)$, in the disk $|z_2|\le 1$. Then differential equation (33) can be solved and

$$\begin{aligned}
Q^*(z_2,s)&=\int_{z_2}^{\omega_2(s)}\frac{1}{\alpha(\hat{f}_2(y,s)-y)}\\
&\qquad\times\exp\left\{\int_y^{z_2}\frac{s+\lambda_0-\lambda_{10}\hat{f}_1(x,s)-\lambda_{20}\hat{f}_2(x,s)}{\alpha(\hat{f}_2(x,s)-x)}\,dy\right\}dx,\\
&\quad z_2\neq\omega_2(s),\\
Q^*(z_2,s)&=\frac{1}{s+\lambda_0-\lambda_{10}\hat{f}_1(z_2,s)-\lambda_{20}\hat{f}_2(z_2,s)},\qquad z_2=\omega_2(s).
\end{aligned}$$

Using the above results we can compute the LST

$$\begin{aligned}
U^*(s)=\int_0^\infty e^{-st}U(t)\,dt&=Q^*(1,s)+P_1^*(1,1,s)+P_2^*(1,s)\\
&\quad+P_{11}^*(1,1,s)+P_4^*(1,1,s),
\end{aligned}$$

while the mean time to first failure ($MTFF$) of the server is given by

$$MTFF=U^*(0).$$

9 Numerical results

In this section, we are going to illustrate useful numerical results that demonstrate the features of the model under consideration. To construct the tables below, we assume that service, repair times and vacation times follow exponential distributions with p.d.f.'s, given by

$$b_i(x)=\frac{1}{\bar{b}_i}e^{-(1/\bar{b}_i)x}, \qquad i=1,2,3,4.$$

Moreover, in all tables below we assume $\lambda_{20}=1.1,\ \lambda_{21}=0.6,\ \lambda_{22}=0.8,\ \lambda_{24}=0.5.$

Table 1 Values of λ_{11}^* for $\lambda_{10}=1, \lambda_{12}=0.8, \nu=4$, $\lambda_{13}=\lambda_{23}=0.1, \lambda_{14}=0.6$, $\mu=3, \bar{b}_4=0.25$

$\bar{b}_1$		$\bar{b}_2$ 0.2	0.5	1	2
0.2	$\bar{b}3=0.5$	10.036	9.323	8.804	8.4097
	$\bar{b}3=1$	9.4647	8.5319	7.8338	7.2917
	$\bar{b}3=2$	8.8948	7.6167	6.595	5.759
0.5		6.8887	6.136	5.5888	5.173
		6.523	5.6068	4.921	4.3886
		5.7355	4.4044	3.3404	2.4705
1		5.9077	5.1599	4.6164	4.2034
		5.5427	4.6318	3.9502	3.4308
		4.7559	3.4316	2.3731	1.5076
2		5.4172	4.6719	4.1302	3.7185
		5.0523	4.1443	3.4647	2.937
		4.2661	2.9452	1.8894	1.0261

Table 2 Values of $E(N_2)$ for $\lambda_{10}=1,\ \nu=4,\ \lambda_{11}=0.8=\lambda_{12},\ \lambda_{13}=0.1,\ \lambda_{14}=0.6,\ \bar{b}_1=0.5,\ \bar{b}_2=0.33,\ \bar{b}_4=0.25,\ \alpha=0.8,\ \nu=4$

λ_{23}		$\bar{b}_3$ 0.1	0.8	1.4	4	8
0	$\mu=1$	8.945	12.06	14.87	28.981	59.382
	$\mu=2$	16.619	28.191	39.527	110.91	369.61
	$\mu=3$	28.124	58.178	90.662	365.77	2886.9
0.1		9.0268	12.981	16.947	43.178	
		16.895	32.338	50.776	274.9	
		28.821	72.021	136.24	3154.9	
0.5		9.3589	17.479	29.006		
		18.051	59.021	966.2		
		31.846	213.01	4347.7		
1		9.7921	25.749			
		19.634	158.41			
		36.246	9551.8			
5		14.149				
		41.527				
		142.5				

Table 1 illustrates values of the maximum permitted arrival rate of P_1 customers given that the server is busy with a P_1 customer, λ_{11}, in order to maintain statistical equilibrium ($\lambda_{11} \leq \lambda_{11}^*$ for $\rho < 1$). Here one can observe the impact of repair times on the maximum permitted traffic of P_1 customers when the server is busy with P_1 customer.

Table 2 contains values of $E(N_2)$ for increasing values of arrival rates of P_2 customers during repair times (λ_{23}) and $\bar{b}_3$, when we vary the failure rate μ. Note that even when we do not allow P_2 arrivals during repair periods ($\lambda_{23}=0$), $E(N_2)$ in-

Table 3 Values of $E(N_2)$, $P(idle)$, $E(W_i)$, $i = 1, 2$, for $\lambda_{10} = 1$, $\mu = 3$, $\nu = 4$, $\alpha = 0.8$, $\lambda_{12} = 0.8$, $\lambda_{14} = 0.6$, $\lambda_{23} = 0.8$, $\bar{b}_1 = 0.5$, $\bar{b}_2 = \bar{b}_3 = 0.33$, $\bar{b}_4 = 0.25$

λ_{11}		λ_{13}				
		0	1	2	3	3.5
0	$E(N_2)$	33.194	58.637	132.41	561.74	2732.7
	$P(idle)$	0.2417	0.2011	0.1496	0.0819	0.0395
	$E(W_1)$	0.548	0.5806	0.6226	0.6783	0.7138
	$E(W_2)$	0.5436	0.5756	0.6168	0.6717	0.7066
1.5		44.61	95.343	342.69	85838.7	
		0.2193	0.1684	0.1009	0.0073	
		0.6002	0.6483	0.7138	0.8086	
		0.5548	0.5959	0.652	0.733	
3		68.505	215.77	5766.46		
		0.1887	0.1217	0.0272		
		0.6783	0.7565	0.8738		
		0.5717	0.6285	0.7138		
4.5		138.28	1611.4			
		0.1443	0.0494			
		0.8086	0.9576			
		0.5997	0.6889			
6		644.69				
		0.0743				
		1.069				
		0.6558				

creases from 16.619 to 369.61 and from 28.124 to 2886.9 when we pass from a system where $\bar{b}_3 = 0.1$ to a system where $\bar{b}_3 = 8$ for $\mu = 2$ and $\mu = 3$, respectively. Note also that for $\lambda_{23} = 1$, a small increase in the repair period from $\bar{b}_3 = 0.1$ to $\bar{b}_3 = 0.8$, $E(N_2)$ increases rapidly from 36.246 to 9551.8, when failure rate equals $\mu = 3$.

Table 3, shows the way $E(N_2)$, $P(idle)$ and $E(W_i)$, $i = 1, 2$, are affected when we vary the arrival rate of P_1 customers during repair, λ_{13}, for increasing values of the arrival rate of P_1 customers when the server, serves a P_1 customer, λ_{11}. Note that if we allow no arrivals of P_1 customers during the service time of P_1 customers ($\lambda_{11} = 0$), the mean number of P_2 customers in orbit $E(N_2)$ is increased dramatically from 33.194 to 2732.7, while $P(idle)$ is reduced for 0.2417 to 0.0395, $E(W_1)$ and $E(W_2)$ increases from 0.548 to 0.7138 and from 0.5436 to 0.7066, respectively, when we pass from a system without P_1 arrivals during repairs ($\lambda_{13} = 0$) to a system with $\lambda_{13} = 3.5$. Note also that when $\lambda_{11} = 4.5$, if we increase λ_{13} from 0 to 1, $E(N_2)$ increases from 138.28 to 1611.4, while $P(idle)$ is reduced from 0.1443 to 0.0494. The results from Table 2, indicates the crucial role of repair period in our systems. In order to manage the congestion in our model, we have to keep at low level the arrivals of P_1 customers during repair period.

Table 4 demonstrates the effect of vacation period on $E(N_2)$, $P(idle)$, $E(W_i)$, $i = 1, 2$, for increasing values of arrival rate of P_1 customers during that period (λ_{14}). Note that when $\lambda_{14} = 1$, an increase on the mean vacation period from $\bar{b}_4 = 0.25$ to

Table 4 Values of $E(N_2)$, $P(idle)$, $E(W_i)$, $i=1,2$ for $\lambda_{10}=1$, $\nu=4$, $\lambda_{11}=0.8$, $\lambda_{13}=0.1=\lambda_{23}$, $\bar{b}_1=0.5$, $\bar{b}_3=0.33$, $\mu=3$

λ_{14}		$\bar{b}_4$			
		0.25	0.4	0.6	1
0	$E(N_2)$	17.115	29.68	68.618	1482.3
	$P(idle)$	0.3043	0.2169	0.1313	0.0258
	$E(W_1)$	0.4969	0.6469	0.8457	1.2469
	$E(W_2)$	0.4722	0.6222	0.821	1.222
0.4		29.671	92.52	2180.9	
		0.2532	0.1389	0.0274	
		0.5491	0.7567	1.062	
		0.5244	0.7321	1.038	
1		85.041	9087.9		
		0.1711	0.0166		
		0.6365	0.9452		
		0.6118	0.9204		
2		7138.8			
		0.0228			
		0.8071			
		0.7824			

$\bar{b}_4 = 0.4$ increase rapidly makes $E(N_2)$ go from 85.041 to 9087.9, while $P(idle)$ is reduced from 0.1711 to 0.0166. From the above results, we have to observe that in order to reduce the congestion in retrial box and to increase the probability the customers from the orbit to connect with the server, we must keep as much as we can the rate λ_{14} at low level.

Definitely, in order to manage the congestion in our model we must control the arrival rates, at time periods where the server is dealt with other activities, such as repair and vacation period.

Table 5 contains values of $E(N_2)$ when we vary the mean interretrial time $E(retrial) = 1/\alpha$ for increasing values of the arrival rate of P_1 customers when the server is idle, λ_{10}, at specific values of arrival rate of P_1 customers during vacation period (λ_{14}). Note that even when λ_{10} take small values, say $\lambda_{10} = 0.5$, an increase from $\lambda_{14} = 0$ to $\lambda_{14} = 1$ causes the increase of $E(N_2)$ from 72.251 to 365.75 when $E(retrial) = 10$. Here one can realize the crucial role of keeping at low levels the arrival rate of P_1 customers during vacation, mentioned here and in Table 4, on the values of $E(N_2)$. Furthermore, if we keep at low levels the values of λ_{10}, we can control $E(N_2)$, especially for small values of $E(retrial)$. In conclusion, it is easy to observe the increase of $E(N_2)$ when $E(retrial)$ increases. This is an increase that is more apparent for large values of λ_{10}.

Table 6 contains values of the steady state availability AV and the failure frequency F of the server when we vary the mean repair time ($\bar{b}_3$) for increasing arrival rate of P_1 customers during repair period (λ_{13}) at specific values of the failure rate μ. We observe that when the failure rate increases from $\mu = 1$ to $\mu = 3$ the server availability decreases, while failure frequency increases. In addition, this change become intense as the mean repair time and λ_{13} increases. Furthermore we have to note here

Table 5 Values of $E(N_2)$, for $\mu = 3$, $\lambda_{11} = 0.8 = \lambda_{12}$, $\nu = 4$, $\bar{b}_1 = 0.5$, $\lambda_{13} = 0.1 = \lambda_{23}$, $\bar{b}_3 = 0.33 = \bar{b}_2$, $b_4 = 0.25$

λ_{10}		$E(retrial)$				
		0.02	0.2	1	3	10
0.1	$\lambda_{14} = 0$	0.716	1.3866	4.367	11.931	37.897
	$\lambda_{14} = 0.5$	1.7243	3.112	9.2778	24.927	78.648
	$\lambda_{14} = 1$	4.2	7.7192	23.358	66.048	199.29
0.5		0.9128	2.199	7.918	22.431	72.251
		2.1435	4.7516	16.343	45.76	146.74
		5.141	11.645	40.551	113.91	365.75
1		1.1623	3.4968	13.873	40.205	130.6
		2.6775	7.3467	28.099	80.765	261.56
		6.3691	17.882	69.051	198.91	644.7
2		1.7166	7.037	30.879	91.295	297.7
		3.863	14.439	61.444	180.74	590.25
		9.1637	34.954	149.57	440.47	1439.1

Table 6 Values of AV, F for $\lambda_{10} = 1$, $\lambda_{11} = 0.8 = \lambda_{12}$, $\nu = 4$, $\lambda_{14} = 0.6$, $\lambda_{23} = 0.1$, $\bar{b}_1 = 0.5$, $\bar{b}_2 = 0.33$, $\bar{b}_4 = 0.25$, $\alpha = 0.8$

λ_{13}			$\bar{b}_3$			
			0.2	1	2	6.67
0	$\mu = 1$	AV	0.9542	0.8208	0.6715	0.3667
		F	0.2289	0.1792	0.1643	0.0945
	$\mu = 3$	AV	0.8807	0.5903	0.411	0.1502
		F	0.5963	0.4097	0.2945	0.1269
0.45			0.9536	0.7938	0.6379	
			0.2319	0.2062	0.1811	
			0.8766	0.5375	0.2953	
			0.6168	0.4625	0.3524	
1			0.9528	0.778		
			0.2357	0.2219		
			0.8712	0.4509		
			0.6439	0.5491		
4			0.9482			
			0.2587			
			0.8108			
			0.9458			

that decrease of AV is smoother in case where $\mu = 1$. On the other hand, when $\mu = 3$, AV is reduced faster.

Finally Fig. 1 shows the way the traffic intensity ρ is affected by the presence of negative arrivals. Clearly, negative arrival is a useful tool to reduce the congestion of the system. For $\lambda_{10} = 1$, $\lambda_{11} = 0.8 = \lambda_{12}$, $\lambda_{13} = 0.1 = \lambda_{23}$, $\lambda_{14} = 0.6$, $\mu = 3$, $\bar{b}_1 = 0.5$, $\bar{b}_2 = 0.33 = \bar{b}_3$, $\bar{b}_4 = 0.25$, when $E(\text{negative}) = 1/\nu$ is small, ρ remains at low

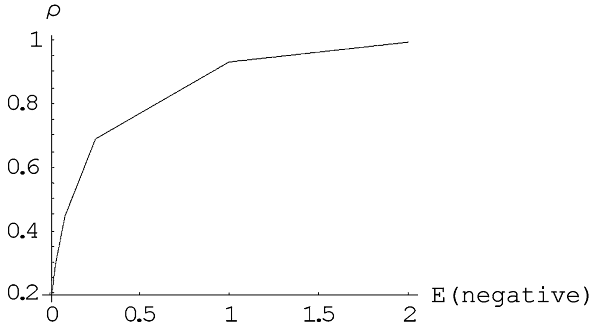

Fig. 1 The effect of negative arrivals on ρ

level. The increase of E(negative) causes the increase of the traffic intensity and as a result the increase in congestion of our model.

References

Aissani A (1994) A retrial queue with redundancy and unreliable server. Queueing Syst 17:431–449

Aissani A, Artalejo JR (1998) On the single server retrial queue subject to breakdowns. Queueing Syst 30:309–321

Aguir S, Karaesmen F, Aksin O, Chauvet F (2004) The impact of retrials on call centers performance. OR Spektrum 26:353–376

Anisimov V, Artalejo JR (2001) Analysis of Markov multiserver retrial queues with negative arrivals. Queueing Syst 39:157–182

Artalejo JR (1999) A classified bibliography of research on retrial queues: progress in 1990–1999. Top 7(2):187–211

Artalejo JR (2010) A classified bibliography of research on retrial queues: progress in 2000–2009. Math Comput Model 51:1071–1081

Artalejo JR, Gomez-Corral A (2008) Retrial queueing systems, a computational approach. Springer, Berlin

Artalejo JR, Gomez-Corral A (1998) Analysis of a stochastic clearing system with repeated attempts. Stoch Models 14:623–645

Artalejo JR, Gomez-Corral A (1997) Steady state solution of a single-server queue with linear requests repeated. J Appl Probab 34:223–233

Artalejo JR, Dudin AN, Klimenok VI (2001) Stationary analysis of a retrial queue with preemptive repeated attempts. Oper Res Lett 28:173–180

Artalejo JR, Li Q-L (2010) Performance analysis of a block-structured discrete-time retrial queue with state-dependent arrivals. Discrete Event Dyn Syst 20:325–347

Boon M, van Wijk A, Adan I, Boxma O (2010) A polling model with smart customers. Queueing Syst 66(3):239–274

Choi BD, Chang Y (1999) Single server retrial queues with priority customers. Math Comput Model 30:7–32

Cinlar E (1975) Introduction to stochastic processes. Prentice Hall, New York

Dimitriou I, Langaris C (2010) A repairable queueing model with two-phase service, start-up times and retrial customers. Comput Oper Res 37(7):1181–1190

Falin GI, Artalejo JR, Martin M (1993) On the single server retrial queue with priority customers. Queueing Syst 14:439–455

Falin GI, Templeton JGC (1997) Retrial queues. Chapman and Hall, London

Furhmand SW, Cooper RB (1985) Stochastic decompositions in the M/G/1 queue with generalized vacations. Oper Res 33(5):1117–1129

Gomez-Corral A, Ramalhoto MF (1999) The stationary distribution of a Markovian process arising in the theory of multiserver retrial queueing systems. Math Comput Model 30:141–158

Koole G, Mandelbaum A (2002) Queueing models of call centers, an introduction. Ann Oper Res 113:41–59

Kulkarni VG, Choi BD (1990) Retrial queue with server subject to breakdowns and repairs. Queueing Syst 7(2):191–208
Kulkarni VG, Liang HM (1997) Retrial queues revisited. In: Dshalalow JH (ed) Frontiers in Queueing. CRP Press, Burbank, pp 19–34
Langaris C, Katsaros A (1995) Time depended analysis of a queue with batch arrivals and N levels of non-preemptive priority. Queueing Syst 19:269–288
Langaris C, Moutzoukis E (1995) A retrial queue with structured batch arrivals, priorities and server vacations. Queueing Syst 20:341–368
Pakes AG (1969) Some conditions of ergodicity and recurrence of Markov chains. Oper Res 17:1058–1061
Parthasarathy PR, Shudesh R (2007) Time dependent analysis of a single server retrial queue with state dependent rates. Oper Res Lett 35:601–611
Shanthikumar JG (1988) On stochastic decomposition in M/G/1 queues with generalized server vacations. Oper Res 36(4):566–569
Shogan AW (1979) A single server queue with arrival rate dependent on server breakdowns. Nav Res Logist Q 26(3):487–497
Shin YW (2007) Multi-server retrial queue with negative customers and disasters. Queueing Syst 55:223–237
Takacs L (1962) Introduction to the theory of queues. Oxford Univ. Press, New York
Wang J, Liu B, Li J (2008) Transient analysis of an M/G/1 retrial queue subject to disasters and server failures. Eur J Oper Res 189(3):1118–1132
Wang J (2008) On the single server retrial queue with priority subscribers and server breakdowns. J Syst Sci Complex 21(2):304–315
Wang J, Cao J, Li Q-L (2001) Reliability analysis of the retrial queue with server breakdowns and repairs. Queueing Syst 38:363–380