RAIRO-Oper. Res. 54 (2020) 231–249
https://doi.org/10.1051/ro/2018117

RAIRO Operations Research
www.rairo-ro.org

STOCHASTIC ANALYSIS OF A PREEMPTIVE RETRIAL QUEUE WITH ORBITAL SEARCH AND MULTIPLE VACATIONS

Shan Gao[1] and Jinting Wang[2,*]

Abstract. This paper deals with a preemptive priority *M/G/1* retrial queue with orbital search and exhaustive multiple vacations. By using embedded Markov chain technique and the supplementary variable method, we discuss the necessary and sufficient condition for the system to be stable and the joint queue length distribution in steady state as well as some important performance measures and the Laplace–Stieltjes transform of the busy period. Also, we establish a special case and the stochastic decomposition laws for this preemptive retrial queueing system. Finally, some numerical examples and cost optimization analysis are presented.

Mathematics Subject Classification. 60K25, 90B22.

Received January 31, 2015. Accepted December 9, 2018.

1. Introduction

During last two decades, the study of retrial queueing systems has been reinvigorated due to the progresses in computer and telecommunication networking technologies. The feature of retrial queueing systems is that customers who find all servers unavailable will join the retrial orbit and retry for service after a random length (called retrial time). Detailed overviews of the related literatures on retrial queues can be referred the books of Falin and Templeton [1], Artalejo and Gómez-Corral [2], the survey papers of Artalejo [3, 4], Kim and Kim [5] and recent papers, *e.g.* Dimitriou [6–8], Boxma and Resing [9] and Abidini *et al.* [10] and references therein.

Recently, preemptive retrial queueing systems have gained wide attention due to their applications in many real world situations. For example, in the computer and communication networks with priority messages, an arriving message of higher priority may push-out the job of lower priority whose service is ongoing, to the orbit to commence its own service. Krishna Kumar *et al.* [11] considered an *M/G/1* retrial queueing system with two-phase service and possible preemptive resume at the first phase of service, Krishna Kumar *et al.* [12] studied the busy period of a retrial queue, which generalized the retrial queue in Krishna Kumar *et al.* [11] by considering that the service time of the second phase is generally distributed. Wu *et al.* [13] presented a discrete-time *Geo/G/1* retrial queue with preferred customers and impatient customers, where the arriving customer may push out the customer in service to commence his own service with some probability. Dimitriou [14] considered a retrial queueing model accepting two types of positive customers and negative arrivals, where

Keywords. Nonlinear system, preemptive retrial queue, orbital search, embedded Markov chain, supplementary variable.

[1] Department of Mathematics and Statistics, Fuyang Normal College, 236037 Fuyang, PR China.
[2] Department of Mathematics, Beijing Jiaotong University, 100044 Beijing, PR China.
*Corresponding author: jtwang@bjtu.edu.cn

Article published by EDP Sciences

an arriving P_1 customer can preempt the service of a P_2 customer and force the server to start his service. Moreover, Artalejo *et al.* [15] considered a retrial queueing system where repeated customers have preemptive priority over customers at the waiting line. To the authors' best knowledge, there have been a little attention on research in preemptive retrial queues with orbital search.

Most of the literature on retrial queues assumes that after completion of each service the server will remain idle in the system until the arrival of the next primary or retrial customer. But in real life, we always want to minimize the idle time of the server and minimize the holding costs. So it is necessary to study the retrial queue with orbital search. That is, after completion of each service, the server may initiatively search the customer from the orbit. Then a service is followed by another service if a search is made, otherwise a service is followed by an idle period. So far seldom research work on retrial queue with orbital search has been done, readers are referred to Artalejo *et al.* [16], Dudin *et al.* [17], Krishnamoorthy *et al.* [18], Chakravarthy *et al.* [19], Zhang and Wang [20], Deepak *et al.* [21] and references therein. Besides introducing the orbital search policy into the preemptive retrial queue, we will also consider exhaustive multiple vacation policy, which is different from vacation policy in Arivudainambi and Dodhandaraman [22], in which the authors presented an *M/G/1* retrial queue with with balking, second optional service and single vacation. Under exhaustive multiple vacation and orbital search protocol, the server takes a vacation with arbitrary random length whenever the system becomes empty. After returning from that vacation, he keeps on taking vacations till he finds at least one customer in the orbit (multiple vacations). If there are customers in the orbit at the completion epoch of a service or the end of a vacation, the server searches for the customers in the orbit or remains idle (orbital search policy).

As an application of the retrial queueing system under investigation, let's consider a call center with an interactive voice response units (IVRUs) (see [23]) and without a waiting loop. Typically, callers arrive at the call center following the Poisson stream. If the agent (server) is busy, an arriving caller may interrupt the customer in service with probability α if he has urgent accident or has to hang up and retry to access to the server later with probability $1 - \alpha$. Such a retrying caller is said to be in orbit. If no callers in the system, the agent can pursuit secondary jobs (*e.g.*, outgoing calls) as long as callers are stored to be served (multiple vacation). Suppose if the agent is idle and there are callers in the orbit, the agent is able to log the phone numbers of unserved customer in the orbit according to the FCFS service discipline. In addition, each customer in the orbit either chooses to leave voice messages to explain the business that he needs to be handled with probability p, or chooses to retry for his service later with probability $1 - p$. When the agent becomes idle, he can immediately rend the service to the customer in the orbit with probability p according to his voice messages, or may need some time to contact the customer with probability $1 - p$, which corresponds to the general retrial policy.

The rest of this paper is organized as follows. In Section 2, we give the description of the queueing system. Section 3 presents the stable condition of the system, steady-state analysis including the joint distribution of the orbit size and the server state at arbitrary epoch, some system characteristics and analysis of busy period. In Section 4, we present a special case of our model and develop the stochastic decomposition property of the system size distribution for our queueing model. In Section 5, some numerical examples and cost optimization analysis are given.

2. Model descriptions

In this section, we consider a single server retrial queue with orbital search and preemptive priority under multiple vacations policy. Assume that customers arrive at the system according to a Poisson process with rate λ. If an arriving customer finds the server idle, the customer obtains service immediately and departs the system after completion of his service. Otherwise, if the server is busy at the arrival epoch, the arriving customer either interrupts the customer in service to commence his own service with probability α (the customer is called preemptive customer) or enters into the orbit with probability $\overline{\alpha} = 1 - \alpha$ (according to FCFS discipline). The interrupted customer enters into the orbit and his service resumes from the beginning. Because we only focus on the distribution of the number of customers in the system and won't consider the waiting and sojourn time distributions, we have no need to specify what position the interrupted customer will occupy.

We assume that the server always takes a vacation when the system becomes empty after a service completion. If the server finds the system is still empty upon returning from his vacation, he will again take another vacation, *i.e.*, the exhaustive multiple vacation policy is adopted. Otherwise, if there are customers in the orbit at a service or a vacation completion epoch, the server will search for the orbital customer with probability p or remains idle with complementary probability $\bar{p}$ $(= 1 - p)$. The search time is assumed to be negligible.

The service times of customers follow an arbitrary distribution with distribution function (d.f.) $B(x)$, probability density function (p.d.f.) $b(x)$, finite first two moments μ_1, μ_2. The vacation times of the server follow an arbitrary distribution with d.f. $V(x)$, p.d.f. $v(x)$, finite first two moments ν_1, ν_2.

We assume that only the customer at the head of the orbit is allowed to access to the server, *i.e.*, if there are customers in the orbit at a service or a vacation completion epoch and no orbital search occurs, the customer at the head of the orbit begins retrying to connect with the server, and the inter-retrial times have an arbitrary distribution with d.f. $R(x)$, p.d.f. $r(x)$. This retrial policy is called general retrial and it was extensively studied by Gómez-Corral [24].

All the random variables defined above are independent.

Throughout the rest of the paper, for a d.f. $F(x)$, define $\overline{F}(x) := 1 - F(x)$, $\widetilde{F}(s) := \int_0^\infty e^{-sx}\mathrm{d}F(x)$.

Define the functions $\beta(x), \eta(x), \gamma(x)$ to be the conditional completion rates for service, for vacation and for retrial attempt respectively, *i.e.*,

$$\beta(x) = \frac{b(x)}{\overline{B}(x)}, \eta(x) = \frac{v(x)}{\overline{V}(x)}, \gamma(x) = \frac{r(x)}{\overline{R}(x)}\cdot$$

The state of the system at time t can be described by the process $Y(t) = \{J(t), N(t), \xi_0(t), \xi_1(t), \xi_2(t)\}$, $(t \geq 0)$, where $J(t)$ is the indicator function of the server state: $J(t)$ is equal to $0, 1$ or 2 depending on whether the server is idle, busy or on vacation at time t; $N(t)$ is the number of customers in the orbit. The random variable $\xi_0(t)$ represents the elapse retrial time of the customer at the head of the orbit at time t if $J(t) = 0$ and $N(t) \geq 1$; $\xi_1(t)$ represents the elapse service time of the customer being served at time t if $J(t) = 1$; and $\xi_2(t)$ represents the elapse vacation time at time t if $J(t) = 2$.

3. Analysis of the model

In this section, we will focus on the discussion of the stability condition of the system by using embedded Markov chain technique and steady-state distribution of the system by using supplementary variable method.

3.1. Embedded Markov chain

For future use, we first introduce some notations.

$$a_k = \int_0^\infty \frac{(\lambda x)^k}{k!} e^{-\lambda x}\overline{\alpha}^k \mathrm{d}B(x), k \geq 0,$$
$$b_k = \int_0^\infty \frac{(\lambda x)^{k-1}}{(k-1)!} \lambda e^{-\lambda x}\overline{\alpha}^{k-1}\alpha\overline{B}(x)\mathrm{d}x, k \geq 1,$$
$$c_k = \int_0^\infty \frac{(\lambda x)^k}{k!} e^{-\lambda x}\mathrm{d}V(x), k \geq 0,$$

where a_k is the probability that there are k new arrivals during the service of a customer and all arrivals don't interrupt the customer being served; b_k is the probability that there are at least k new arrivals during a customer's service and the kth arrival preempts the customer; c_k is the probability that there are k new arrivals during the vacation time.

Evidently, $\sum_{k=0}^\infty a_k = \widetilde{B}(\lambda\alpha), \sum_{k=1}^\infty b_k = 1 - \widetilde{B}(\lambda\alpha), \sum_{k=0}^\infty c_k = 1$, which indicate that $\{a_k, k \geq 0\}$ and $\{b_k, k \geq 0\}$ are two non-complete probability distributions, $\{c_k, k \geq 0\}$ is a complete probability distribution.

The generating functions of $\{a_k, k \geq 0\}$, $\{b_k, k \geq 0\}$ and $\{c_k, k \geq 0\}$ can be obtained as follows

$$A(z) = \sum_{k=0}^{\infty} z^k a_k = \widetilde{B}(\lambda(1-\overline{\alpha}z)),$$
$$B(z) = \sum_{k=1}^{\infty} z^k b_k = \frac{\alpha z}{1-\overline{\alpha}z}(1-A(z)),$$
$$V(z) = \sum_{k=0}^{\infty} z^k c_k = \widetilde{V}(\lambda(1-z)).$$

Note that we introduce the preemptive resume and multiple vacation policy into the retrial queue, to obtain the stability condition of the system, we follow the argument of embedded Markov chain. Let $\{T_n, n \geq 0\}$ be the sequence of epochs at which either a service is totally completed or a preemption occurs or a proper vacation period ends. Then the sequence of random vectors $\{(N(T_n^+), I(T_n)), n \geq 0\}$ forms a Markov chain which is the embedded Markov chain for our queuing system, where $N(T_n^+)$ denotes the number of customers in the orbit immediately after the embedded epoch T_n; $I(T_n)$ denotes the factor which make the system's state change at time epoch T_n: $I(T_n) = 0$ represents that at epoch T_n a customer's service is preempted by a new arrival and then he enters into the orbit, $I(T_n) = 1$ denotes that at epoch T_n a customer's service completely ends and he departs the system, $I(T_n) = 2$ denotes that a proper vacation ends at epoch T_n. The state space of the embedded Markov chain $\{(N(T_n^+), I(T_n)), n \geq 0\}$ is $\{(0,1),(0,2)\} \cup \{(n,j), n \geq 1; j = 0,1,2\}$.

Using the above notations and lexicographical sequence for the states, we can have the transition probability matrix of $\{(N(T_n^+), I(T_n)), n \geq 0\}$ as follows:

$$\mathscr{P} = \begin{bmatrix} \mathbf{B}_0 & \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{B}_3 & \cdots \\ \mathbf{C}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & \cdots \\ & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \cdots \\ & & \mathbf{A}_0 & \mathbf{A}_1 & \cdots \\ & & & \vdots & \vdots \end{bmatrix},$$

where

$$\mathbf{B}_0 = \begin{bmatrix} 0 & c_0 \\ 0 & c_0 \end{bmatrix}, \ \mathbf{B}_n = \begin{bmatrix} 0 & 0 & c_n \\ 0 & 0 & c_n \end{bmatrix}, n \geq 1,$$
$$\mathbf{C}_0 = \begin{bmatrix} 0 & 0 \\ d_0 & 0 \\ d_0 & 0 \end{bmatrix}, \ \mathbf{A}_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & d_0 & 0 \\ 0 & d_0 & 0 \end{bmatrix},$$
$$\mathbf{A}_n = \begin{bmatrix} b_{n-1} & a_{n-1} & 0 \\ e_n & d_n & 0 \\ e_n & d_n & 0 \end{bmatrix}, n \geq 1,$$

and

$$d_n = (p + \overline{p}\widetilde{R}(\lambda))a_n + (1-\delta_{0,n})\overline{p}(1-\widetilde{R}(\lambda))a_{n-1}, n \geq 0,$$
$$e_n = (p + \overline{p}\widetilde{R}(\lambda))b_n + \overline{p}(1-\widetilde{R}(\lambda))b_{n-1}, n \geq 1,$$

$\delta_{0,n}$ is the Kronecker's symbol, and $b_0 = 0$. Obviously, $\sum_{n=0}^{\infty} d_n + \sum_{n=1}^{\infty} e_n = 1$.

Theorem 3.1. *The Markov chain $\{(N(T_n^+), I(T_n)), n \geq 0\}$ is ergodic if and only if*

$$\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)} < p + \overline{p}\widetilde{R}(\lambda).$$

Proof. From the expression of the transition matrix $\mathscr{P}$, we know that

$$\mathbf{A} = \sum_{n=0}^{\infty} \mathbf{A}_n = \begin{bmatrix} 1-\widetilde{B}(\lambda\alpha) & \widetilde{B}(\lambda\alpha) & 0 \\ 1-\widetilde{B}(\lambda\alpha) & \widetilde{B}(\lambda\alpha) & 0 \\ 1-\widetilde{B}(\lambda\alpha) & \widetilde{B}(\lambda\alpha) & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{A}(1) & \mathbf{O} \\ \mathbf{T}(1) & \mathbf{T}(0) \end{bmatrix},$$

where

$$\mathbf{O} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \mathbf{A}(1) = \begin{bmatrix} 1-\widetilde{B}(\lambda\alpha) & \widetilde{B}(\lambda\alpha) \\ 1-\widetilde{B}(\lambda\alpha) & \widetilde{B}(\lambda\alpha) \end{bmatrix}$$

and

$$\mathbf{T}(0) = 0, \ \mathbf{T}(1) = \begin{bmatrix} 1-\widetilde{B}(\lambda\alpha) & \widetilde{B}(\lambda\alpha) \end{bmatrix}.$$

Thus the matrix $\mathbf{A}$ is a reducible stochastic matrix. Evidently, the matrix $\mathbf{A}(1)$ is an irreducible stochastic matrix and has invariant probability vector $\pi(1) = \begin{pmatrix} 1-\widetilde{B}(\lambda\alpha) & \widetilde{B}(\lambda\alpha) \end{pmatrix}$. In addition, define the vector $\beta(1) = \sum_{n=0}^{\infty} n\mathbf{A}_n(1)\mathbf{e}$, where

$$\mathbf{A}_0(1) = \begin{bmatrix} 0 & 0 \\ 0 & d_0 \end{bmatrix}, \ \mathbf{A}_n(1) = \begin{bmatrix} b_{n-1} & a_{n-1} \\ e_n & d_n \end{bmatrix}, n \geq 1, \ \mathbf{e} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Then we can have that $\beta(1) = \begin{bmatrix} 1 + \frac{1-\widetilde{B}(\lambda\alpha)}{\alpha} \\ \frac{1-\widetilde{B}(\lambda\alpha)}{\alpha} + \overline{p}(1-\widetilde{R}(\lambda)) \end{bmatrix}$.

Hence by Theorem 2.3.3 in Neuts [25], the Markov chain $\{(N(T_n^+), I(T_n)), n \geq 0\}$ is positive recurrent if and only if

$$\pi(1)\beta(1) = 1-\widetilde{B}(\lambda\alpha) + \frac{1-\widetilde{B}(\lambda\alpha)}{\alpha} + \overline{p}\widetilde{B}(\lambda\alpha)(1-\widetilde{R}(\lambda)) < 1,$$

which leads to the inequality

$$\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)} < p + \overline{p}\widetilde{R}(\lambda).$$

□

3.2. Steady-state distribution of the system

In this section, we assume that the steady state condition $\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)} < p+\overline{p}\widetilde{R}(\lambda)$ holds. We will use the method of supplementary variable technique to derive the stationary distribution of the system.

Let $Y = \{N, J, \xi_0, \xi_1, \xi_2\}$ be the stationary limit of the Markov process $Y(t) = \{N(t), J(t), \xi_0(t), \xi_1(t), \xi_2(t)\}$.

Define the following joint state probability densities at time t $(t \geq 0)$:

$$p_{k,j}(t,x) = \lim_{h\to 0} \frac{P(N(t)=k, J(t)=0, x<\xi_0(t)\leq x+h)}{h}, j=0, k\geq 1; j=1,2; k\geq 0; x\geq 0.$$

And denote

$$p_{k,j}(x) = \lim_{t\to\infty} p_{k,j}(t,x), j=0, k\geq 1; j=1,2; k\geq 0; x\geq 0.$$

By the method of the supplementary variable, we easily obtain the set of equilibrium equations:

$$\frac{\mathrm{d}}{\mathrm{d}x}p_{k,0}(x) = -(\lambda+\gamma(x))p_{k,0}(x), \ k\geq 1, \tag{3.1}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}p_{k,1}(x) = -(\lambda+\beta(x))p_{k,1}(x) + (1-\delta_{k,0})\lambda\overline{\alpha}p_{k-1,1}(x), \ k\geq 0, \tag{3.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}p_{k,2}(x) = -(\lambda+\eta(x))p_{k,2}(x) + (1-\delta_{k,0})\lambda p_{k-1,2}(x), \ k\geq 0, \tag{3.3}$$

where $\delta_{k,0}$ is the Kronecker's symbol, and the set of equations (3.1)–(3.3) is a version of Kolmogorov's equations. The boundary conditions are

$$p_{k,0}(0) = \overline{p}\bigg(\int_0^\infty p_{k,1}(x)\beta(x)\mathrm{d}x + \int_0^\infty p_{k,2}(x)\eta(x)\mathrm{d}x \bigg),\ k \geq 1, \tag{3.4}$$

$$\begin{aligned} p_{k,1}(0) = {} & (1-\delta_{k,0})\lambda \int_0^\infty p_{k,0}(x)\mathrm{d}x + \int_0^\infty p_{k+1,0}(x)\gamma(x)\mathrm{d}x + (1-\delta_{k,0})\lambda\alpha \int_0^\infty p_{k-1,1}(x)\mathrm{d}x \\ & + p\left(\int_0^\infty p_{k+1,1}(x)\beta(x)\mathrm{d}x + \int_0^\infty p_{k+1,2}(x)\eta(x)\mathrm{d}x \right),\ k \geq 0, \end{aligned} \tag{3.5}$$

$$p_{0,2}(0) = \int_0^\infty p_{0,1}(x)\beta(x)\mathrm{d}x + \int_0^\infty p_{0,2}(x)\eta(x)\mathrm{d}x, \tag{3.6}$$

$$p_{k,2}(0) = 0, k \geq 1, \tag{3.7}$$

and the normalization condition is

$$\sum_{k=1}^\infty \int_0^\infty p_{k,0}(x)\mathrm{d}x + \sum_{k=0}^\infty \int_0^\infty \Big(p_{k,1}(x) + p_{k,2}(x)\Big)\mathrm{d}x = 1.$$

Define the following generating functions

$$P_0(x,z) = \sum_{k=1}^\infty z^k p_{k,0}(x),$$

and

$$P_j(x,z) = \sum_{k=0}^\infty z^k p_{k,j}(x),\ j = 1, 2.$$

From equations (3.1)–(3.3), we have

$$P_0(x,z) = P_0(0,z)\exp\{-\lambda x\}\overline{R}(x), \tag{3.8}$$

$$P_1(x,z) = P_1(0,z)\exp\{-\lambda(1-\overline{\alpha}z)x\}\overline{B}(x), \tag{3.9}$$

$$P_2(x,z) = P_2(0,z)\exp\{-\lambda(1-z)x\}\overline{V}(x). \tag{3.10}$$

From (3.6) and (3.7), we can obtain

$$P_2(0,z) = P_{2,0}(0) := K. \tag{3.11}$$

Then from (3.4)–(3.6) and (3.9)–(3.11), we can obtain

$$P_0(0,z) = \overline{p}\Big(P_1(0,z)A(z) + K\big(V(z)-1\big)\Big), \tag{3.12}$$

$$P_1(0,z) = P_1(0,z)\Big(\frac{p}{z}A(z) + B(z)\Big) + \frac{p}{z}K(V(z)-1) + \frac{1}{z}P_0(0,z)\Big(\widetilde{R}(\lambda) + z(1-\widetilde{R}(\lambda))\Big). \tag{3.13}$$

Combining (3.12) and (3.13), we arrive at

$$P_0(0,z) = \frac{\bar{p}z(1-V(z))(1-B(z))}{\mathcal{D}(z)}K, \tag{3.14}$$

$$P_1(0,z) = \frac{(1-V(z))\Big(p + \overline{p}(\widetilde{R}(\lambda) + z(1-\widetilde{R}(\lambda)))\Big)}{\mathcal{D}(z)}K, \tag{3.15}$$

where

$$\mathcal{D}(z) = A(z)\Big(p + \overline{p}(\widetilde{R}(\lambda) + z(1 - \widetilde{R}(\lambda)))\Big) - z(1 - B(z)).$$

Here we should investigate the zeros of the equation $\mathcal{D}(z) = 0$ in the range $0 \le z \le 1$. For this point, we have the following lemma.

Lemma 3.2. *Under the stationary condition* $\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)} < p + \overline{p}\widetilde{R}(\lambda)$, *the equation*

$$A(z)\Big(p + \overline{p}(\widetilde{R}(\lambda) + z(1 - \widetilde{R}(\lambda)))\Big) - z(1 - B(z)) = 0,$$

has no roots in the range $0 \le z \le 1$ *and has the minimal nonnegative root* $z = 1$.

Proof. We only need to verify that $u(z) \triangleq A(z)\Big(p+\overline{p}(\widetilde{R}(\lambda)+z(1-\widetilde{R}(\lambda)))\Big)+zB(z)$ is a probability generating function of a nonnegative integer valued random variable.

Let U be the time period from the epoch a service begins to the next service begins given that the orbit is non-empty, and let N_U be the the change of the orbit size during U ($N_u = -1, 0, 1, 2, \dots$) and define

$$u_k(t)\mathrm{d}t = P(t < U \le t + \mathrm{d}t, N_U = k).$$

Then,

$$u_k(t) = I(k \ge 1)b_k(t) + pa_{k+1}(t) + \overline{p}\Big(a_{k+1}(t) \star (e^{-\lambda t}r(t)) + I(k \ge 0)a_k(t) \star (\lambda e^{-\lambda t}\overline{R}(t))\Big), k = 0, 1, 2, \dots,$$

where $I(\bullet)$ is an indicate function, $\star$ means convolution,

$$a_k(t) = \frac{(\lambda t)^k}{k!}e^{-\lambda t}\overline{\alpha}^k b(t), k \ge 0,$$
$$b_k(t) = \frac{(\lambda t)^{k-1}}{(k-1)!}\lambda e^{-\lambda t}\overline{\alpha}^{k-1}\alpha\overline{B}(t), k \ge 0.$$

Denote by $N_U(z)$ the generating function of N_U, after careful calculation, we have that

$$N_U(z) = E[z^{N_U}] = \sum_{k=-1}^{\infty} z^k \int_0^{\infty} u_k(t)\mathrm{d}t$$
$$= B(z) + \frac{A(z)}{z}\Big(p + \overline{p}(\widetilde{R}(\lambda) + z(1 - \widetilde{R}(\lambda)))\Big),$$

which shows that $u(z) \triangleq A(z)\Big(p+\overline{p}(\widetilde{R}(\lambda)+z(1-\widetilde{R}(\lambda)))\Big)+zB(z) = E[z^{N_U+1}]$ is exactly a probability generating function of random variable $N_U + 1$. From the condition the stationary condition $\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)} < p + \overline{p}\widetilde{R}(\lambda)$, $u(z)$ has the following properties:

(1) $u(0) = a_0\Big(p + \overline{p}\widetilde{R}(\lambda)\Big) > 0$;
(2) $u(1) = 1$,
(3) $u'(1) = \frac{d}{dz}u(z)|_{z=1} = 1 - \Big((p+\overline{p}\widetilde{R}(\lambda))\widetilde{B}(\lambda\alpha) - \frac{1-\widetilde{B}(\lambda\alpha)}{\alpha}\Big) < 1$. Since the probability generating function $u(z)$ is a convex and monotonically increasing function of z for $0 \le z \le 1$, we can easily prove the expected result of Lemma 3.2.

□

Applying the normalization condition, we obtain

$$\int_0^\infty P_0(x,1)\mathrm{d}x + \int_0^\infty \Big(P_1(x,1)+P_2(x,1)\Big)\mathrm{d}x = 1,$$

and by using (3.8)–(3.10) and (3.14), (3.15), we can arrive at

$$K = \frac{1}{\nu_1}\left(p+\overline{p}\widetilde{R}(\lambda) - \frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}\right). \tag{3.16}$$

Now we summarize the above results in the following theorem.

Theorem 3.3. *Under the stationary condition* $\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)} < p+\overline{p}\widetilde{R}(\lambda)$, *the generating functions of the stationary joint distribution of the orbit size and the server state are given by:*

$$\begin{aligned}
P_0(x,z) &= K\frac{\overline{p}z(1-V(z))(1-B(z))}{\mathcal{D}(z)}\exp\{-\lambda x\}\overline{R}(x),\\
P_1(x,z) &= K\frac{(1-V(z))\Big(p+\overline{p}(\widetilde{R}(\lambda)+z(1-\widetilde{R}(\lambda)))\Big)}{\mathcal{D}(z)}\exp\{-\lambda(1-\overline{\alpha}z)x\}\overline{B}(x),\\
P_2(x,z) &= K\exp\{-\lambda(1-z)x\}\overline{V}(x).
\end{aligned}$$

Next we focus on the marginal generating functions of the orbit size and the server states given in the following corollary.

Corollary 3.4. (1) *The marginal generating function of the orbit size when the server is idle is*

$$P_0(z) = \int_0^\infty P_0(x,z)\mathrm{d}x = K\frac{\overline{p}z(1-V(z))(1-B(z))}{\mathcal{D}(z)}\frac{1-\widetilde{R}(\lambda)}{\lambda}.$$

(2) *The marginal generating function of the orbit size when the server is busy is*

$$P_1(z) = \int_0^\infty P_1(x,z)\mathrm{d}x = K\frac{(1-V(z))\Big(p+\overline{p}(\widetilde{R}(\lambda)+z(1-\widetilde{R}(\lambda)))\Big)}{\mathcal{D}(z)}\frac{1-A(z)}{\lambda(1-\overline{\alpha}z)}.$$

(3) *The marginal generating function of the orbit size when the server is on vacation is*

$$P_2(z) = \int_0^\infty P_2(x,z)\mathrm{d}x = K\frac{1-V(z)}{\lambda(1-z)}.$$

Corollary 3.5. (1) *The generating function of the orbit size,* $P(z)$, *is given by*

$$\begin{aligned}
P(z) &= P_0(z)+P_1(z)+P_2(z)\\
&= \frac{\mathcal{N}(z)}{\mathcal{D}(z)}\frac{1-V(z)}{\lambda(1-\overline{\alpha}z)}K,
\end{aligned}$$

where $\mathcal{N}(z) = \Big(p+\overline{p}\big(\widetilde{R}(\lambda)+z(1-\widetilde{R}(\lambda))\big)\Big)\Big(\alpha z A(z)+1-z\Big)$.

(2) *The generating function of the system size,* $\Phi(z)$, *is given by*

$$\begin{aligned}
\Phi(z) &= P_0(z)+zP_1(z)+P_2(z) = P(z)+(z-1)P_1(z)\\
&= K\frac{1-V(z)}{\lambda}\frac{\Big(p+\overline{p}\big(\widetilde{R}(\lambda)+z(1-\widetilde{R}(\lambda))\big)\Big)A(z)}{\mathcal{D}(z)}.
\end{aligned} \tag{3.17}$$

Using the above results, we can obtain some performance measures of the system in steady state.

Corollary 3.6. (1) *The probability that the server is idle, denoted by P_0, is given by*

$$P_0 = P_0(1) = \overline{p}(1 - \widetilde{R}(\lambda)).$$

(2) *The probability that the server is busy, denoted by P_1, is given by*

$$P_1 = P_1(1) = \frac{1 - \widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}\cdot$$

(3) *The probability that the server is on vacation, denoted by P_2, is given by*

$$P_2 = P_2(1) = p + \overline{p}\widetilde{R}(\lambda) - \frac{1 - \widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}\cdot$$

(4) *The mean orbit size, $E(L_q)$, is given by*

$$E(L_q) = P'(1) = \frac{V''(1)}{2V'(1)} - \frac{\mathcal{D}''(1)}{2\mathcal{D}'(1)} + \frac{\mathcal{N}'(1) + \overline{\alpha}A(1)}{\mathcal{N}(1)},$$

where

$$\begin{aligned}
V'(1) &= \lambda\nu_1,\\
V''(1) &= \lambda^2\nu_2,\\
\mathcal{D}'(1) &= \frac{1 - \widetilde{B}(\lambda\alpha)}{\alpha} - (p + \overline{p}\widetilde{R}(\lambda))\widetilde{B}(\lambda\alpha),\\
\mathcal{D}''(1) &= \frac{2(1 - \widetilde{B}(\lambda\alpha))}{\alpha^2} + 2\lambda\overline{\alpha}\widetilde{B}'(\lambda\alpha)\Big(p + \overline{p}\widetilde{R}(\lambda) + \frac{1}{\alpha}\Big),\\
\mathcal{N}'(1) + \overline{\alpha}A(1) &= \widetilde{B}(\lambda\alpha)\Big(1 + \overline{p}\alpha(1 - \widetilde{R}(\lambda))\Big) - \lambda\alpha\overline{\alpha}\widetilde{B}'(\lambda\alpha) - 1,\\
\mathcal{N}(1) &= \alpha\widetilde{B}(\lambda\alpha).
\end{aligned}$$

(5) *The mean system size, $E(L_s)$, is given by*

$$E(L_s) = \Phi'(1) = E(L_q) + P_1(1).$$

(6) *The mean sojourn time of an arbitrary customer in the system, $E(W_s)$, is given by Little's law:*

$$E(W_s) = \frac{E(L_s)}{\lambda}\cdot$$

Next, we consider a cycle of the system for the model under consideration. Let E_V be the expected length of the total vacation period that starts at the epoch when the server begins a vacation and ends at a vacation completion epoch with customers being in the orbit, let E_I be the expected length of the period that the server is idle, and E_B be the expected length of the period that the server is busy. Then the expected length of a cycle Θ is given by $E(\Theta) = E_V + E_I + E_B$. Then we have the following theorem.

Theorem 3.7. *Under the steady-state, we have*

$$\begin{aligned}
E(\Theta) &= \frac{\nu_1}{1-\widetilde{V}(\lambda)}\left(p+\overline{p}\widetilde{R}(\lambda)-\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}\right)^{-1}, \\
E_V &= \frac{\nu_1}{1-\widetilde{V}(\lambda)}, \\
E_I &= \frac{\nu_1}{1-\widetilde{V}(\lambda)}\left(p+\overline{p}\widetilde{R}(\lambda)-\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}\right)^{-1}\overline{p}(1-\widetilde{R}(\lambda)), \\
E_B &= \frac{\nu_1}{1-\widetilde{V}(\lambda)}\left(p+\overline{p}\widetilde{R}(\lambda)-\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}\right)^{-1}\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}.
\end{aligned}$$

Obviously, E_V consists of some consecutive vacation times. Assume that N_V is the number of vacations in E_V, then

$$P(N_V=k)=(\widetilde{V}(\lambda))^{k-1}(1-\widetilde{V}(\lambda)).$$

Then we have that

$$E_V=E[N_V]\nu_1=\frac{\nu_1}{1-\widetilde{V}(\lambda)}.$$

By applying the argument of an alternating renewal process, we have

$$P_2=\frac{E_V}{E(\Theta)},$$

which yields

$$E(\Theta)=\frac{E_V}{P_2}=\frac{\nu_1}{1-\widetilde{V}(\lambda)}\left(p+\overline{p}\widetilde{R}(\lambda)-\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}\right)^{-1}.$$

Using the argument of an alternating renewal process once again, we have

$$P_0=\frac{E_I}{E(\Theta)},\ P_1=\frac{E_B}{E(\Theta)}.$$

The we can obtain the expected results. □

3.3. Laplace–Stieltjes transform of the length of busy period

We define the server busy period of our retrial queue as the period that starts at the epoch when the server returns from the vacation and finds an nonempty orbit and ends at the first departure epoch at which the system is empty. Our interest is focused on the length of the busy period, denoted by L with distribution function $\pi(t)$ and the Laplace–Stieltjes transform $\widetilde{\Pi}(s)=E[e^{-sL}]$.

Define the transient taboo probabilities of process $Y(t)$ as follows:

$$\begin{aligned}
q_{k,0}(t,x) &= \lim_{h\to 0}\frac{P(L>t,N(t)=k,J(t)=0,\xi_0(t)\in(x,x+h))}{h}, t\geq 0, k\geq 1, x\geq 0, \\
q_{k,1}(t,x) &= \lim_{h\to 0}\frac{P(L>t,N(t)=k,J(t)=1,\xi_1(t)\in(x,x+h))}{h}, t\geq 0, k\geq 0, x\geq 0.
\end{aligned}$$

Without loss of generality, we assume that at time 0, the server just ends his vacation and there are k $(k\geq 1)$ customers in the orbit, and denote the corresponding distribution function of the length of busy period as $\pi_k(t)$ with Laplace–Stieltjes transform $\widetilde{\Pi}_k(s)$. Considering the orbital search schedule, at time 0, the system's state may be $(k,0,0)$, or $(k-1,1,0)$.

First we assume $k=1$, there are two possible cases at time 0:

(1) One is the case that orbital search occurs at the end of vacation with one customer in the orbit, then the system initial state is $(0,1,0)$, corresponding distribution function of busy period is denoted to be $\pi_{11}(t)$ with Laplace–Stieltjes transform $\widetilde{\Pi}_{11}(s)$.
Using supplementary variable method, we obtain the following Kolmogorov equations that govern the dynamic of the system:

$$\frac{\mathrm{d}}{\mathrm{d}t}\pi_{11}(t) = \int_0^\infty q_{0,1}(t,x)\beta(x)\mathrm{d}x, \tag{3.18}$$

$$\frac{\partial}{\partial t}q_{k,0}(t,x) + \frac{\partial}{\partial x}q_{k,0}(t,x) = -(\lambda+\gamma(x))q_{k,0}(t,x),\ k\geq 1, \tag{3.19}$$

$$\frac{\partial}{\partial t}q_{k,1}(t,x) + \frac{\partial}{\partial x}q_{k,1}(t,x) = -(\lambda+\beta(x))q_{k,1}(t,x) + (1-\delta_{k,0})\lambda\overline{\alpha}q_{k-1,1}(t,x),\ k\geq 0, \tag{3.20}$$

$$q_{k,0}(t,0) = \overline{p}\int_0^\infty q_{k,1}(t,x)\beta(x)\mathrm{d}x,\ k\geq 1, \tag{3.21}$$

$$\begin{aligned} q_{k,1}(t,0) = {} & (1-\delta_{k,0})\lambda\int_0^\infty q_{k,0}(t,x)\mathrm{d}x \\ & + \int_0^\infty q_{k+1,0}(t,x)\gamma(x)\mathrm{d}x + (1-\delta_{k,0})\lambda\alpha\int_0^\infty q_{k-1,1}(t,x)\mathrm{d}x \\ & + p\int_0^\infty q_{k+1,1}(t,x)\beta(x)\mathrm{d}x,\ k\geq 0, \end{aligned} \tag{3.22}$$

then the boundary conditions are

$$q_{k,0}(0,x) = 0,\ k\geq 1, q_{k,1}(0,x) = \delta_{k,0}\delta(x), k\geq 0, \tag{3.23}$$

where $\delta(x)$ is Dirac delta function.
Now, we introduce Laplace transforms and generating functions to solve the equations (3.18)–(3.22) with boundary condition (3.23):

$$\widetilde{Q}_{k,0}(s,x) = \int_0^\infty e^{-st}q_{k,0}(t,x)\mathrm{d}t, k\geq 1,\ \widetilde{Q}_0(s,x,z) = \sum_{k=1}^\infty z^k\widetilde{Q}_{k,0}(s,x),$$

$$\widetilde{Q}_{k,1}(s,x) = \int_0^\infty e^{-st}q_{k,1}(t,x)\mathrm{d}t, k\geq 0,\ \widetilde{Q}_1(s,x,z) = \sum_{k=0}^\infty z^k\widetilde{Q}_{k,1}(s,x).$$

Then equations (3.18)–(3.22) become

$$\widetilde{\Pi}_{11}(s) = \int_0^\infty \widetilde{Q}_{0,1}(s,x)\beta(x)\mathrm{d}x, \tag{3.24}$$

$$\frac{\partial}{\partial x}\widetilde{Q}_0(s,x,z) = -(s+\lambda+\gamma(x))Q_0(s,x,z), \tag{3.25}$$

$$\frac{\partial}{\partial x}\widetilde{Q}_1(s,x,z) = -(s+\lambda(1-\overline{\alpha}z)+\beta(x))\widetilde{Q}_1(s,x,z) + \delta(x), \tag{3.26}$$

$$\widetilde{Q}_0(s,0,z) = \overline{p}\int_0^\infty \widetilde{Q}_1(s,x,z)\beta(x)\mathrm{d}x - \overline{p}\widetilde{\Pi}_{11}(s), \tag{3.27}$$

$$\begin{aligned} \widetilde{Q}_1(s,0,z) = {} & \lambda\int_0^\infty \widetilde{Q}_0(s,x,z)\mathrm{d}x + \frac{1}{z}\int_0^\infty \widetilde{Q}_0(s,x,z)\gamma(x)\mathrm{d}x + \lambda\alpha z\int_0^\infty \widetilde{Q}_1(s,x,z)\mathrm{d}x \\ & + \frac{p}{z}\left(\int_0^\infty \widetilde{Q}_1(s,x,z)\beta(x)\mathrm{d}x - \widetilde{\Pi}_{11}(s)\right). \end{aligned} \tag{3.28}$$

Solving equations (3.25) and (3.26) leads to

$$\widetilde{Q}_0(s,x,z) = \widetilde{Q}_0(s,0,z)e^{-(s+\lambda)x}\overline{R}(x), \tag{3.29}$$
$$\widetilde{Q}_1(s,x,z) = (1+\widetilde{Q}_1(s,0,z))e^{-(s+\lambda(1-\overline{\alpha}z))x}\overline{B}(x). \tag{3.30}$$

Combining (3.27) and (3.30) and after algebra manipulation, we obtain

$$\widetilde{Q}_0(s,0,z) = \overline{p}[(1+\widetilde{Q}_1(s,0,z))\widetilde{B}(s+\lambda(1-\overline{\alpha}z)) - \widetilde{\Pi}_{11}(s)]. \tag{3.31}$$

Now using equations (3.28)–(3.31), after rearrangement, we have

$$\begin{aligned} &1+\widetilde{Q}_1(s,0,z)\\ &= \frac{(s+\lambda(1-\overline{\alpha}z))\Big(z(s+\lambda) - p(s+\lambda)\widetilde{\Pi}_{11}(s) + \widetilde{Q}_0(s,0,z)\big(\lambda z + (s+\lambda(1-z))\widetilde{R}(s+\lambda)\big)\Big)}{(s+\lambda)\Big(z(s+\lambda(1-z)) + \big(\lambda\alpha z^2 - p(s+\lambda(1-\overline{\alpha}z))\big)\widetilde{B}(s+\lambda(1-\overline{\alpha}z))\Big)}. \end{aligned} \tag{3.32}$$

Substituting (3.32) in (3.31) yields

$$\begin{aligned} f(s,z)\widetilde{Q}_0(s,0,z) = \overline{p}z(s+\lambda)\Big[&\widetilde{B}(s+\lambda(1-\overline{\alpha}z))(s+\lambda(1-\overline{\alpha}z)) - \widetilde{\Pi}_{11}(s)\\ &\times\Big((s+\lambda(1-z)) + \lambda\alpha z\widetilde{B}(s+\lambda(1-\overline{\alpha}z))\Big)\Big], \end{aligned} \tag{3.33}$$

where

$$\begin{aligned} f(s,z) = {}& z(s+\lambda)\Big(s+\lambda(1-z) + \lambda\alpha z\widetilde{B}(s+\lambda(1-\overline{\alpha}z))\Big)\\ &- \widetilde{B}(s+\lambda(1-\overline{\alpha}z))(s+\lambda(1-\overline{\alpha}z))\big[p(s+\lambda) + \overline{p}\big(\lambda z + (s+\lambda(1-z))\widetilde{R}(s+\lambda)\big)\big], \end{aligned} \tag{3.34}$$

for fixed s with $Re(s) > 0$, it's easy to realize that the equation $f(s,z) = 0$ has a unique solution in $0 \le z \le 1$, denoted by $z = \phi(s)(0 \le \phi(s) \le 1)$, and $\phi(s)$ satisfies: $\phi(0) = 1$, and

$$\phi'(0) = \frac{1}{\lambda}\left(1 - \frac{\alpha\widetilde{B}(\lambda\alpha)}{\widetilde{B}(\lambda\alpha)[1+\alpha(p+\overline{p}\widetilde{R}(\lambda))] - 1}\right) < 0.$$

Then, we can get $\widetilde{\Pi}_{11}(s)$ by taking $z = \phi(s)$ in (3.33) as follows

$$\widetilde{\Pi}_{11}(s) = \frac{\widetilde{B}(s+\lambda(1-\overline{\alpha}\phi(s)))(s+\lambda(1-\overline{\alpha}\phi(s)))}{(s+\lambda(1-\phi(s))) + \lambda\alpha\phi(s)\widetilde{B}(s+\lambda(1-\overline{\alpha}\phi(s)))}. \tag{3.35}$$

(2) The other one is that no orbital search occurs at the end of vacation with one customer in the orbit, then the system initial state is $(1,0,0)$, corresponding distribution function of busy period is denoted to be $\pi_{12}(t)$ with Laplace–Stieltjes transform $\widetilde{\Pi}_{12}(s)$. Then the boundary conditions are

$$q_{k,0}(0,x) = \delta_{k,1}\delta(x),\ k \ge 1, q_{k,1}(0,x) = 0, k \ge 0, \tag{3.36}$$

and (3.24) is replaced by

$$\widetilde{\Pi}_{12}(s) = \int_0^\infty \widetilde{Q}_{0,1}(s,x)\beta(x)\mathrm{d}x,$$

By taking Laplace transforms and generating functions, equations (3.19) and (3.20) with above boundary conditions become

$$\frac{\partial}{\partial x}\widetilde{Q}_0(s,x,z) = -(s+\lambda+\gamma(x))Q_0(s,x,z) + z\delta(x), \tag{3.37}$$
$$\frac{\partial}{\partial x}\widetilde{Q}_1(s,x,z) = -(s+\lambda(1-\overline{\alpha}z)+\beta(x))\widetilde{Q}_1(s,x,z), \tag{3.38}$$

and equations (3.27) and (3.28) still hold.
From (3.37) and (3.38), we get

$$\widetilde{Q}_0(s,x,z) = (Q_0(s,0,z)+z)e^{-(s+\lambda)x}\overline{R}(x), \tag{3.39}$$
$$\widetilde{Q}_1(s,x,z) = Q_1(s,0,z)e^{-(s+\lambda(1-\overline{\alpha}z))x}\overline{B}(x). \tag{3.40}$$

By using the above two equations and (3.27), (3.28), we obtain

$$\widetilde{Q}_0(s,0,z) = \overline{p}[\widetilde{Q}_1(s,0,z)\widetilde{B}(s+\lambda(1-\overline{\alpha}z)) - \widetilde{\Pi}_{12}(s)], \tag{3.41}$$

$$\widetilde{Q}_1(s,0,z) = \frac{(s+\lambda(1-\overline{\alpha}z))\Big(\big(z+\widetilde{Q}_0(s,0,z)\big)\big(\lambda z+(s+\lambda(1-z))\widetilde{R}(s+\lambda)\big) - p(s+\lambda)\widetilde{\Pi}_{12}(s)\Big)}{(s+\lambda)\Big(z(s+\lambda(1-z)) + \big(\lambda\alpha z^2 - p(s+\lambda(1-\overline{\alpha}z))\big)\widetilde{B}(s+\lambda(1-\overline{\alpha}z))\Big)}. \tag{3.42}$$

Substituting (3.42) into (3.41) leads to

$$\begin{aligned} f(s,z)\widetilde{Q}_0(s,0,z) = \overline{p}z\Big[&\widetilde{B}(s+\lambda(1-\overline{\alpha}z))(s+\lambda(1-\overline{\alpha}z))\big(\lambda z+(s+\lambda(1-z))\widetilde{R}(s+\lambda)\big) \\ &- \widetilde{\Pi}_{12}(s)(s+\lambda)\Big((s+\lambda(1-z)) + \lambda\alpha z\widetilde{B}(s+\lambda(1-\overline{\alpha}z))\Big)\Big]. \end{aligned} \tag{3.43}$$

Taking $z=\phi(s)$ in (3.43) yields

$$\widetilde{\Pi}_{12}(s) = \frac{\widetilde{B}(s+\lambda(1-\overline{\alpha}\phi(s)))(s+\lambda(1-\overline{\alpha}\phi(s)))\big(\lambda\phi(s)+(s+\lambda(1-\phi(s)))\widetilde{R}(s+\lambda)\big)}{(s+\lambda)\Big((s+\lambda(1-\phi(s))) + \lambda\alpha\phi(s)\widetilde{B}(s+\lambda(1-\overline{\alpha}\phi(s)))\Big)}. \tag{3.44}$$

Combining the two cases (1) and (2), we have that

$$\begin{aligned} \widetilde{\Pi}_1(s) &= p\widetilde{\Pi}_{11}(s) + \overline{p}\widetilde{\Pi}_{12}(s) \\ &= \frac{\widetilde{B}(s+\lambda(1-\overline{\alpha}\phi(s)))(s+\lambda(1-\overline{\alpha}\phi(s)))}{(s+\lambda(1-\phi(s))) + \lambda\alpha\phi(s)\widetilde{B}(s+\lambda(1-\overline{\alpha}\phi(s)))} \\ &\quad \times \left(p + \frac{\overline{p}\big(\lambda\phi(s)+(s+\lambda(1-\phi(s)))\widetilde{R}(s+\lambda)\big)}{s+\lambda}\right). \end{aligned}$$

Using $f(s,\phi(s))=0$, we can simplify the expression of $\widetilde{\Pi}_1(s)$ to be the following

$$\widetilde{\Pi}_1(s) = \phi(s).$$

Second, for $k \geq 2$, we have that

$$\widetilde{\Pi}_k(s) = (\widetilde{\Pi}_1(s))^k = \phi^k(s). \tag{3.45}$$

Therefore, conditioning on the numbers of customers in the orbit at the end of the vacation, we can get the Laplace–Stieltjes transform $\widetilde{\Pi}(s)$ of the busy period L as

$$\begin{aligned} \widetilde{\Pi}(s) &= \frac{1}{1-c_0}\sum_{k=1}^{\infty} c_k\widetilde{\Pi}_k(s) \\ &= \frac{1}{1-\widetilde{V}(\lambda)}\Big(\widetilde{V}\big(\lambda(1-\phi(s))\big) - \widetilde{V}(\lambda)\Big). \end{aligned} \tag{3.46}$$

Differentiating (3.46) with respective to s at $s=0$ yields

$$\begin{aligned} E[L] &= -\frac{\mathrm{d}}{\mathrm{d}s}\widetilde{\Pi}(s)\big|_{s=0} \\ &= \frac{\nu_1}{1-\widetilde{V}(\lambda)}\left(\frac{\alpha\widetilde{B}(\lambda\alpha)}{\widetilde{B}(\lambda\alpha)[1+\alpha(p+\overline{p}\widetilde{R}(\lambda))]-1}-1\right). \end{aligned} \tag{3.47}$$

Obviously, $E[L]=E(\Theta)-E_V$.

4. A special case and stochastic decomposition laws

In this section, we present a special case of our model and use it to develop the stochastic decomposition law of the system size distribution for our queueing model.

4.1. A special case and its stochastic decomposition

If $p=1$ or $\widetilde{R}(\lambda)=0$, that is, no retrial occurs, then our model is reduced to the classical preemptive *M/G/1* queue with multiple vacations, and denote the PGF of this system size by $\Phi_0(z)$, then we have that

$$\begin{aligned} \Phi_0(z) &= \left(1-\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}\right)\frac{1-V(z)}{\lambda\nu_1}\frac{A(z)}{A(z)-z(1-B(z))} \\ &= \left(1-\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}\right)\frac{(1-z)A(z)}{A(z)-z(1-B(z))}\frac{1-V(z)}{\lambda\nu_1(1-z)} \\ &= L_\infty(z)\zeta(z), \end{aligned} \tag{4.1}$$

where

$$L_\infty(z) = \left(1-\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}\right)\frac{(1-z)A(z)}{A(z)-z(1-B(z))}, \tag{4.2}$$

denotes the PGF of the stationary system size distribution of the classical preemptive *M/G/1* queue without vacation, which can be obtained by letting $\phi^*(\lambda)=0$ and $p=0$ in equation (37) of Krishna Kumar *et al.* [11] and

$$\zeta(z) = \frac{1-V(z)}{\lambda\nu_1(1-z)},$$

denotes the PGF of the number of customers that arrive during the residual life of the vacation time.

From equation (4.1), we observe that the stationary system size of the preemptive *M/G/1* queue with multiple vacations is the sum of two independent random variables: one of which is the stationary system size (L_∞) of the classical preemptive *M/G/1* queue without vacation and the other one is the number of customers (L_v) that arrive during the residual vacation time.

4.2. Stochastic decomposition laws of our model

By using the above results, we show that stochastic decomposition law still holds for our model.

By Corollaries 3.4 and 3.6, we obtain the generating functions of the conditional system size given that the server is idle or on vacation as follows

$$\begin{aligned} \chi(z) &= \frac{P_0(z)+P_2(z)}{P_0(1)+P_2(1)} \\ &= \frac{p+\overline{p}\widetilde{R}(\lambda)-\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}}{1-\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}}\frac{\left(A(z)-z(1-B(z))\right)\left(p+\overline{p}\big(\widetilde{R}(\lambda)+z(1-\widetilde{R}(\lambda))\big)\right)}{\mathcal{D}(z)}\frac{1-V(z)}{\lambda\nu_1(1-z)}. \end{aligned} \tag{4.3}$$

Comparing (3.17) and (4.2), (4.3), we have that

$$\Phi(z) = L_\infty(z)\chi(z),$$

which indicates that the the number of customers in the system under study (L_s) can be expressed as the sum of two independent random variables, the first one of which is is the stationary system size (L_∞) of the classical preemptive *M/G/1* queue without vacation, the second one is the number of customers (L_1) in the orbit given that the server is idle or on vacation. That is, $L_s = L_\infty + L_1$.

On the other hand, from (3.17) and (4.1), one can find that there exists another stochastic decomposition law of our model as follows:

$$\begin{aligned}\Phi(z) &= \left[\left(p + \overline{p}\widetilde{R}(\lambda) - \frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}\right)\frac{(1-z)A(z)}{\mathcal{D}(z)}\right]\frac{1-V(z)}{\lambda\nu_1(1-z)}\Big(p + \overline{p}\widetilde{R}(\lambda) + z\overline{p}\big(1-\widetilde{R}(\lambda)\big)\Big)\\ &= \Psi(z)\zeta(z)\xi(z),\end{aligned} \tag{4.4}$$

where

$$\Psi(z) = \left(p + \overline{p}\widetilde{R}(\lambda) - \frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)}\right)\frac{(1-z)A(z)}{\mathcal{D}(z)},$$

represents the generating function of the system size $\tilde{L}_s$ of the preemptive *M/G/1* retrial queue with orbital search and general retrial times and without vacations, which can be obtained by using the similar method in Section 3, and

$$\xi(z) = p + \overline{p}\widetilde{R}(\lambda) + z\overline{p}\big(1-\widetilde{R}(\lambda)\big),$$

which is the generation function of a discrete random variable ξ which follows Binomial distribution

$$P(\xi = 0) = p + \overline{p}\widetilde{R}(\lambda), P(\xi = 1) = \overline{p}(1-\widetilde{R}(\lambda)).$$

Equation (4.4) indicates that the the number of customers in the system under study (L_s) can be expressed as the sum of three independent random variables, the first one of which is is the stationary system size ($\tilde{L}_s$) of the preemptive *M/G/1* retrial queue with orbital search and general retrial times, the second one is the number of customers (L_v) that arrive during the residual vacation time and the third is the number of customers (ξ) in the orbit given that the server is idle or on vacation. That is, $L_s = \tilde{L}_s + L_v + \xi$.

5. Numerical examples and cost analysis

In this section, numerical examples deal with the case of exponential distributions for vacation time and retrial time with p.d.f.s. $v(x) = \nu e^{-\nu x}$ and $r(x) = re^{-rx}$, respectively.

The first set of numerical examples is devoted to Erlangian distribution of order 2 for service time with p.d.f. $b(x) = \mu^2 x e^{-\mu x}$. We examine the effect of varying parameters on the expected length of a cycle ($E[\Theta]$) and the mean system size ($E[L_s]$), which are shown in Figure 1.

From Figure 1(a) and (b), we can see that $E[\Theta]$ and $E[L_s]$ decrease with increasing the value of orbital search probability p. The reason is that the bigger the value of orbital search probability p is, the smaller the server's idle period is, which leads to more customers get their service quickly and then the length of a cycle and the system size decrease. Figure 1(a) and (b) also show that $E[\Theta]$ and $E[L_s]$ increase monotonously with increasing the value of preemptive probability α. That is because when the preemptive probability α increases, the time of a customer spending in the system will possibly increase due to potentially being preempted by the last come customer, which leads to the expected length of a cycle $E[\Theta]$ and the mean system size $E[L_s]$ to increase.

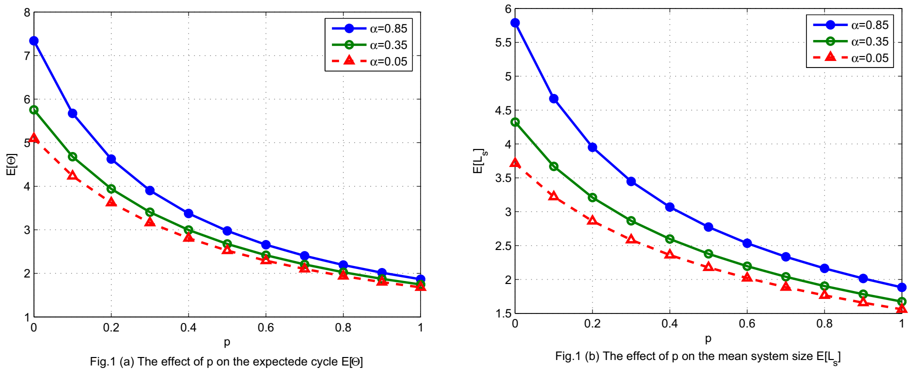

FIGURE 1. The effect of p on performance measures $E[\Theta]$ and $E[L_s]$ with $\lambda = 2, \mu = 8, \nu = 3, r = 4$.

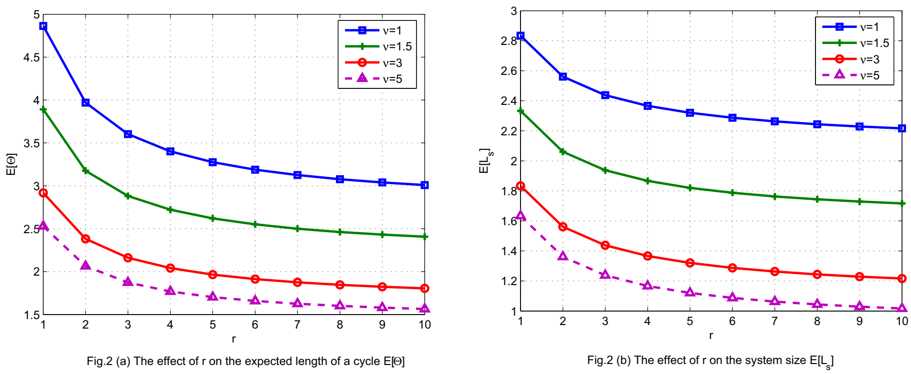

FIGURE 2. The effect of p on performance measures $E[\Theta]$ and $E[L_s]$ with $\lambda = 1.5, \mu = 8, \alpha = 0.35, p = 0.55$.

In Figure 2, we plot the expected length of a cycle $E[\Theta]$ and the mean system size $E[L_s]$ *versus* retrial rate r for different values of $\nu = 1, 1.5, 3, 5$. As seen from Figure 2, as functions of retrial rate r, $E[\Theta]$ and $E[L_s]$ decrease monotonously, which agrees with intuitive expectations. However, as functions of vacation rate ν, $E[\Theta]$ and $E[L_s]$ increase monotonously. The reason is that the bigger the value of ν is, the longer the vacation time is, and then the number of customers who enter into the orbit increases, which causes $E[\Theta]$ and $E[L_s]$ to increase.

The second set of numerical examples deals with different distributions for service times with exponential distribution with p.d.f. $b(x) = \mu e^{-\mu x}$), Erlangian distribution of order 2 with p.d.f. $b(x) = \mu^2 x e^{-\mu x}$ and hyper-exponential distribution with p.d.f. $b(x) = 0.4\mu e^{-\mu x} + 0.6\mu^2 x e^{-\mu x}$. The graphs illustrated in Figures 3 and 4 compare the behavior of $E[\Theta]$ and $E[L_s]$ against the parameters (I) p, the probability of orbital search, (II) α, preemptive probability, (III) r, retrial rate and (IV) ν, vacation rate. An interesting thing is that when the service time follows exponential distribution, the effect of the preemptive probability α on the performance measures $E[\Theta]$ and $E[L_s]$ vanishes because of the memoryless property of the exponential distribution.

Finally, we focus on a steady state total expected cost function per unit time for the preemptive *M/G/1* retrial queue with orbital search and multiple vacations, in which orbital search probability p is a decision variable. With the cost structure being considered, the objective is to determine the optimum value of p, so as to minimize the expected operating cost function per unit time.

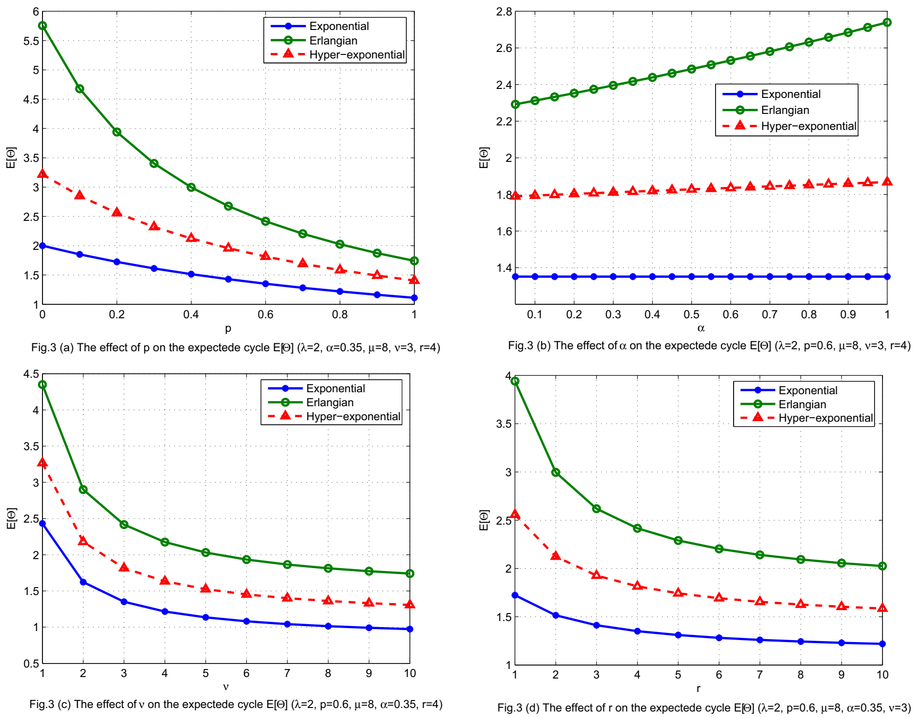

FIGURE 3. $E[\Theta]$ *versus* different parameters.

Define the following cost elements:

$$C_h \equiv \text{unit time holding cost of every customer in the system;}$$
$$C_0 \equiv \text{unit time cost for keeping the server idle;}$$
$$C_1 \equiv \text{unit time cost for keeping the server busy;}$$
$$C_S \equiv \text{setup cost per cycle;}$$
$$R_v \equiv \text{unit time reward for the server being on vacation.}$$

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

$$TC(p) = C_h E[L_s] + C_0 P_0 + C_1 P_1 + C_S \frac{1}{E(\Theta)} - R_v P_2.$$

The cost minimization problem can be illustrated mathematically as $\min_p TC(p)$ subject to the stationary condition $\frac{1-\widetilde{B}(\lambda\alpha)}{\alpha\widetilde{B}(\lambda\alpha)} < p + \overline{p}\widetilde{R}(\lambda)$ and $0 \le p \le 1$. Because the total expected cost function per unit time is highly non-linear and complex, we can't easily get the derivatives of it. Hence we develop approximations by Matlab program to find the optimum value of p, say p^*. With assumptions that $C_h = 20, C_0 = 30, C_1 = 150, C_2 = 80, C_S = 200$, $b(x) = \mu^2 x e^{-\mu x}$ and $\lambda = 2, \alpha = 0.35, \mu = 8, \nu = 3, r = 4$. In Figure 5, we plot the effect of p on $TC(p)$.

From Figure 5, we can see that there is an optimal orbital search probability p to make the cost minimize. Implementing the computer software MATLAB by the parabolic method and the error is controlled by $\varepsilon = 10^{-6}$, we find the solution $p^* = 0.45303691$ with $TC(p^*) = 180.59771$.

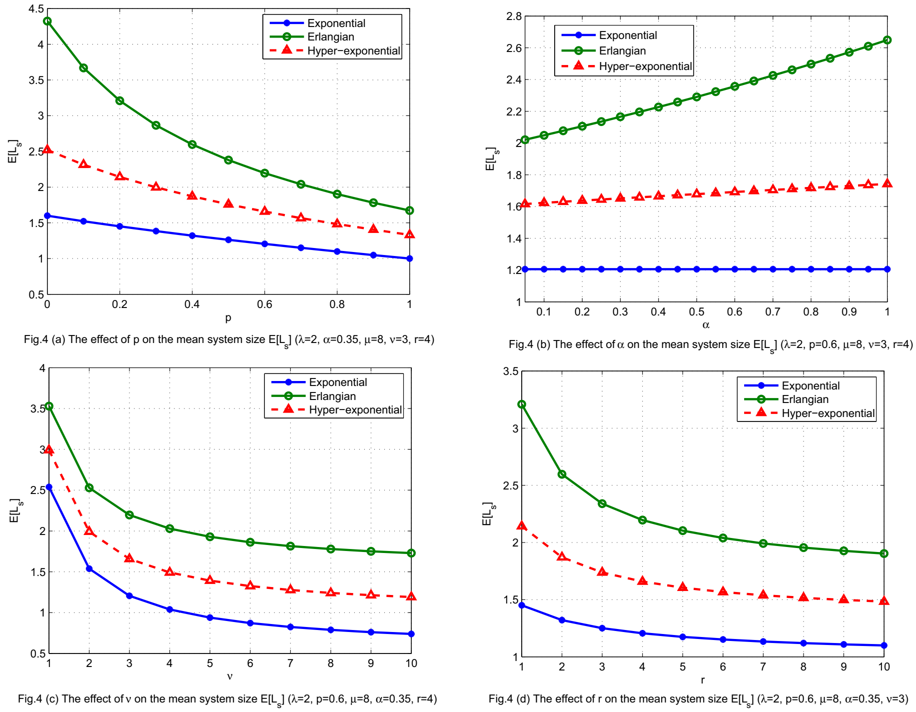

FIGURE 4. $E[L_s]$ *versus* different parameters.

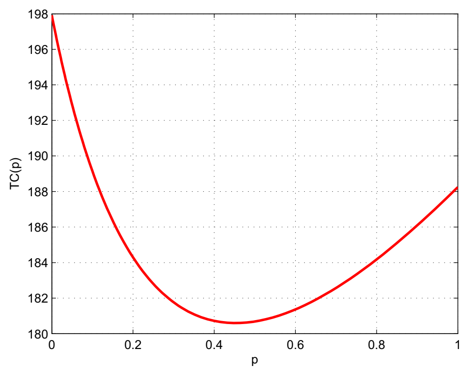

FIGURE 5. The effect of p on $TC(p)$.

Acknowledgements. The authors thank the referees for their helpful comments on an earlier version of this paper. This work is supported by the Natural Science Foundation of Anhui Higher Education Institutions of China under Grant Nos. KJ2017A340, KJ2014ZD21 and KJ2016A875; Fuyang Municipal Government-Fuyang Normal College Horizontal Cooperation projects in 2017 (NO. XDHXTD201709); the National Natural Science Foundation of China under Grant Nos. 71871008 and 71571014, the 111 Project of China (B16002).

References

[1] G.I. Falin and J.G.C. Templeton, Retrial Queues. Chapman and Hall, London (1997).

[2] J.R. Artalejo and A. Gómez-Corral, Retrial Queueing Systems. Springer-Verlag, Berlin Heidelberg (2008).

[3] J.R. Artalejo, A classified bibliography of research on retrial queues: progress in 1990–1999. *Top* **7** (1999) 187–211.

[4] J.R. Artalejo, Accessible bibliography on retrial queues: progress in 2000–2009. *Math. Comput. Model.* **51** (2010) 1071–1081.

[5] J. Kim and B. Kim, A survey of retrial queueing systems. *Ann. Oper. Res.* **247** (2016) 3–36.

[6] I. Dimitriou, A queueing model with two classes of retrial customers and paired services. *Ann. Oper. Res.* **238** (2015) 1–21.

[7] I. Dimitriou, A retrial queue to model a two-relay cooperative wireless system with simultaneous packet reception, edited by S. Wittevrongel and T. Phung-Duc. In: ASMTA 2016, LNCS. Springer, Cardiff **9845** (2016) 123–139.

[8] I. Dimitriou, A two class retrial system with coupled orbit queues. *Probab. Eng. Info. Sci.* **31** (2017) 139–179.

[9] O.J. Boxma and J.A.C. Resing, Vacation and polling models with retrials. In: 11th European Workshop on Performance Engineering (EPEW 11). Florence (September 2014).

[10] M.A. Abidini, J. Resing and O. Boxma, Analysis and optimization of vacation and polling models with retrials. *Perform. Eval.* **98** (2016) 52–69.

[11] B. Krishna Kumar, A. Vijayakumar and D. Arivudainambi, An *M/G/1* retrial queueing system with two-phase service and preemptive resume. *Ann. Oper. Res.* **113** (2002) 61–79.

[12] B. Krishna Kumar, S. Pavai Madheswari and D. Arivudainambi, On the busy period of an *M/G/1* retrial queueing system with two-phase service and preemptive resume. *Stoch. Model. App.* **8** (2005) 18–34.

[13] J. Wu, J. Wang and Z. Liu, A discrete-time *Geo/G/1* retrial queue with preferred and impatient customers. *Appl. Math. Model.* **37** (2013) 2552–2561.

[14] I. Dimitriou, A mixed priority retrial queue with negative arrivals, unreliable server and multiple vacations. *Appl. Math. Model.* **37** (2013) 1295–1309.

[15] J.R. Artalejo, A.N. Dudin and V.I. Klimenok, Stationary analysis of a retrial queue with preemptive repeated attempts. *Oper. Res. Lett.* **28** (2001) 173–180.

[16] J.R. Artalejo, V.C. Joshua and A. Krishnamurthy, An *M/G/1* retrial queue with orbital search by the server, edited by J.R. Artalejo and A. Krishnamoorthy. In: Advances in Stochastic Modelling. Notable Publications, Inc., New Jersey (2002) 41–54.

[17] A.N. Dudin, A. Krishnamoorthy, V.C. Joshua and G.V. Tsarenkov, Analysis of the *BMAP/G/1* retrial system with search of customers from the orbit. *Eur. J. Oper. Res.* **157** (2004) 169–179.

[18] A. Krishnamoorthy, T.G. Deepak and V.C. Joshua, An *M/G/1* retrial queue with nonpersistent customers and orbital search. *Stoch. Anal. App.* **23** (2005) 975–997.

[19] S.R. Chakravarthy, A. Krishnamoorthy and V.C. Joshua, Analysis of a multi-server retrial queue with search of customers from the orbit. *Perform. Eval.* **63** (2006) 776–798.

[20] F. Zhang and J. Wang, Stochastic analysis of a finite source retrial queue with spares and orbit search. MMB & DFT, LNCS 7201 (2012) 16–30.

[21] T.G. Deepak, A.N. Dudin and V.C. Joshua, A. Krishnamoorthy, On an $M^{(X)}/G/1$ Retrial System with two types of search of customers from the orbit. *Stoch. Anal. App.* **31** (2013) 92–107.

[22] D. Arivudainambi and P. Dodhandaraman, Retrial queueing system with balking, optional service and vacation. *Ann. Oper. Res.* **229** (2015) 67–84.

[23] R. Srinivasan, J. Talim and J. Wang, Performance analysis of a call center with interactive voice response units. *TOP* **12** (2004) 91–110.

[24] A. Gómez-Corral, Stochastic analysis of a single server retrial queue with general retrial time. *Nav. Res. Logist.* **46** (1999) 561–581.

[25] M.F. Neuts, Structured Stachastic Matrices of *M/G/1* Type and Their Applications. Marcel Dekker, New York, NY (1989).