Tail Asymptotics for a Retrial Queue with Bernoulli Schedule

Bin Liu $^{1,a}$ and Yiqiang Q. Zhao $^b$

a. School of Mathematics & Physics, Anhui Jianzhu University, Hefei 230601, P.R. China
b. School of Mathematics and Statistics, Carleton University, Ottawa, ON, Canada K1S 5B6

December 2017

Abstract

In this paper, we study the asymptotic behavior of the tail probability of the number of customers in the steady-state $M/G/1$ retrial queue with Bernoulli schedule, under the assumption that the service time distribution has a regularly varying tail. Detailed tail asymptotic properties are obtained for the (conditional and unconditional) probability of the number of customers in the (priority) queue, orbit and system, respectively.

Keywords: $M/G/1$ retrial queue, Bernoulli schedule, Number of customers, Asymptotic tail probability, Regularly varying distribution.

Mathematics Subject Classification (2010): 60K25; 60G50; 90B22.

1 Introduction

As one of the very important types of queueing systems, retrial queues have been extensively studied for more than 40 years, which produced more than 100 literature publications. Research on retrial queues is still very active and new challenges are continuing to emerge. A general picture of retrial models, together with their applications in various areas, and basic results on retrial queues can be acquired from surveys or books, such as Yang and Templeton [36], Falin [12], Kulkarni and Liang [25], Falin and Templeton [14], Artalejo [1, 2], Choi and Chang [9], Artajelo and Gómez-Corral [4], Artalejo [3], Kim and Kim [21], among possible others.

Tail asymptotic analysis of retrial queueing systems, especially asymptotic properties in tail stationary probabilities for a stable retrial queue have been a focus of the investigation in the past 10 years or so, due to two main reasons: first, for most of retrial queues, it is not expected to have explicit non-Laplace transform solutions for stationary distributions, and second, tail asymptotic properties often lead to approximations, bounds on performance, and numerical algorithms. Both light-tailed and heavy-tailed properties have been obtained for a number of retrial queues, including the following incomplete list: Kim, Kim and Ko [22], Liu and Zhao [32], Kim, Kim and Kim [20], Liu, Wang and Zhao [31], Kim, Kim and Kim [21], Kim and Kim [19], Artalejo and Phung-Duc [5], Kim [18], while for heavy-tailed behaviour, readers may refer to Shang, Liu and Li [34], Kim, Kim and Kim [23], Yamamuro [35], Liu, Wang and Zhao [30], Masuyama [33], Liu, Min and Zhao [28], and Liu and Zhao [29].

In this paper, we consider a $M/G/1$ retrial queue with Bernoulli schedule. This model was first proposed and studied by Choi and Park [10]. Specifically, this $M/G/1$ retrial queueing system has a
(priority) queue of infinite waiting capacity and an orbit. External customers arrive to this system according to a Poisson process with rate $\lambda$. There is a single server in this system. If the server is idle upon the arrival of a customer, the customer receives the service immediately and leaves the system after the completion of the service. Otherwise, or if the server is busy, the arriving customer would join the queue with probability $q$, becoming a priority customer waiting for the service according to first-in-first-out discipline; or the orbit with probability $p = 1 - q$, becoming a repeated customer who will retry later for receiving its service. A priority customer has priority over a repeated customer for receiving the service. It implies that upon the completion of a service, if there is a customer in the queue, the server will serve the customer at the head of the queue; otherwise the server becomes idle. Each of the repeated customers in the orbit independently repeatedly tries for receiving service according to a Poisson process with the retrial rate $\mu$ until he finds the idle server, and then immediately receive the service and leaves the system after the completion of the service. All customers receive the same service time whose distribution $F_\beta(x)$ with $F_\beta(0) = 0$ is assumed to have a finite mean $\beta_1$. The Laplace-Stieltjes transform (LST) of the distribution function $F_\beta(x)$ is denoted by $\beta(s)$.

Closely related to the model studied in [10] and also in this paper, are Falin, Artalejo and Martin [13], in which a model with two independent (primary) Poisson arrival streams was considered. The priority customers when blocked upon arrival are queued and waiting for service, while the non-priority customers when blocked join the orbit and retry for service later; Li and Yang [27], in which the discrete-time counterpart to the model studied in this paper, or a Geo/G/1 retrial queue with Bernoulli schedule, was considered; and Atencia and Moreno [7], in which the M/G/1 retrial queueing system with Bernoulli schedule has a general retrial time, but only the customer at the head of the orbit is allowed to retry for service, or retrials with a constant rate.

Let $\rho = \lambda \beta_1$. It follows from [10] that the system considered in this paper is stable if and only if (iff) $\rho < 1$, which is assumed to be true throughout the paper. Our focus is on asymptotic properties for various tail stationary probabilities of the M/G/1 retrial queue with Bernoulli schedule, which have not been studied before. The start point of our study is based on two expressions for probability transformations obtained in [10]. Stochastic decompositions are tools in our analysis. By assuming a regular varying tail in the service time distribution, we obtain in this paper asymptotic properties for the tail probabilities of customers: (1) in the orbit given that the server is idle (Section 4.1); (2) in the queue given that the server is busy (Section 4.2); (3) in the orbit given that the server is busy (Section 4.3); (4) in the queue and in the orbit, respectively (Section 4.4); (5) in the whole system (total number of customers in the queue and the orbit, plus the possible one in the service) (Section 4.5); (6) in the orbit given that the server is busy and the queue is empty (Section 4.6).

The rest of the paper is organized as follows: preliminary results are provided in Section 2; stochastic decompositions are obtained in Section 3 and the main results, asymptotic properties for tail probabilities, are derived in Section 4.

## 2 Preliminary

Assume that the system is in steady state. Let $R_{\text{que}}$ be the number of priority customers in the queue, let $R_{\text{orb}}$ be the number of repeated customers in the orbit, and let $I_{\text{sev}} = 1$ or 0, whenever the server is busy or idle, respectively. Let $R_0$ be a random variable (r.v.) whose distribution
is easily obtained as follows:

\[ \lambda_p \]

Poisson process with arrival rate

\[ b \]

immediately from equations (12) and (13) in \[10\]:

Following the discussions in \[10\], let

\[ G \]

where

\[ I \]

\[ R \]

coincides with the conditional distribution of \( R_{orb} \) given that \( I_{sev} = 0 \), and let \((R_{11}, R_{12})\) be a two-dimensional r.v. whose distribution coincides with the conditional distribution of \((R_{que}, R_{orb})\) given that \( I_{sev} = 1 \). Precisely, \( R_0, R_{11} \) and \( R_{12} \) are all nonnegative integer-valued random variables; \( R_0 \) has the probability generating function (PGF): \( R_0(z) = E(z^{R_0}) = E(z_{2, orb}^{R_{orb}}|I_{sev} = 0), \) and \((R_{11}, R_{12})\) has the PGF:

\[ R_1(z, z_2) = E(z^{R_{11}}_{z_{1}, orb = 0} z^{R_{12}}_{z_{2}, orb = 0}) = E(z_{1, que}^{R_{que}} z_{2, orb}^{R_{orb}}|I_{sev} = 1). \]

Our start point for tail asymptotic analysis is based on the expressions for \( R_0(z_2) \) and \( R_1(z_1, z_2) \). Following the discussions in \[10\], let

\[
M_a(z_1, z_2) = \frac{1}{\rho} \cdot \frac{1 - \beta(\lambda - \lambda p z_2 - \lambda q z_1)}{1 - p z_2 - q z_1}, \tag{2.1}
\]

\[
M_b(z_1, z_2) = (1 - \rho q) \cdot \frac{G(z_2) - z_1}{\beta(\lambda - \lambda p z_2 - \lambda q z_1) - z_1}, \tag{2.2}
\]

\[
M_c(z_2) = \frac{1 - \rho}{1 - \rho q} \cdot \frac{1 - z_2}{G(z_2) - z_2}, \tag{2.3}
\]

where \( G(\cdot) \) is determined uniquely by the following equation

\[ G(z) = \beta(\lambda - \lambda p z - \lambda q G(z)). \tag{2.4} \]

Since \( P\{I_{sev} = 0\} = 1 - \rho \) and \( P\{I_{sev} = 1\} = \rho \), obtained in \[10\], we have the following expressions immediately from equations (12) and (13) in \[10\]:

\[
R_0(z_2) = \exp \left\{ -\frac{\lambda}{\mu} \int_{z_2}^{\infty} \frac{1 - G(u)}{G(u) - u} du \right\}, \tag{2.5}
\]

\[
R_1(z_1, z_2) = M_a(z_1, z_2) \cdot M_b(z_1, z_2) \cdot M_c(z_2) \cdot R_0(z_2). \tag{2.6}
\]

In the next section, we will prove that \( R_0(z_2) \), \( M_c(z_2) \), \( M_a(z_1, z_2) \) and \( M_b(z_1, z_2) \) can be viewed as the PGFs of four one or two dimensional) r.v.s.

Next, we provide a probabilistic interpretation for \( G(z) \). Let \( T_a \) be the busy period of the standard \( M/G/1 \) queue with arrival rate \( \lambda q \) and the service time \( T_s \). By \( F_\alpha(x) \) we denote the probability distribution function of \( T_\alpha \), and by \( \alpha(s) \) the LST of \( F_\alpha(x) \). The following are classic results on the busy period of this standard \( M/G/1 \) queue:

\[
\alpha(s) = \beta(s + \lambda q - \lambda q \alpha(s)), \tag{2.7}
\]

\[
\alpha_1 \defeq E(T_\alpha) = \beta_1/(1 - \rho q). \tag{2.8}
\]

Throughout paper we will use the notation \( N_b(t) \) to represent the number of Poisson arrivals with rate \( b \) within the time interval \((0, t)\), and \( N_{\lambda p}(T_\alpha) \) to represent the number of arrivals of a Poisson process with arrival rate \( \lambda p \) within the independent random time \( T_\alpha \). The PGF of \( N_{\lambda p}(T_\alpha) \) is easily obtained as follows:

\[
E(z^{N_{\lambda p}(T_\alpha)}) = \int_0^{\infty} \sum_{n=0}^{\infty} \frac{(\lambda p x)^n}{n!} e^{-\lambda p x} dF_\alpha(x) = \alpha(\lambda p - \lambda p). \tag{2.9}
\]
It follows from (2.7) that
\[ \alpha(\lambda p - \lambda pz) = \beta(\lambda - \lambda pz - \lambda q \cdot \alpha(\lambda p - \lambda pz)). \tag{2.10} \]

By comparing (2.4) and (2.10) and noticing the uniqueness of \( G(z) \), we immediately have
\[ G(z) = \alpha(\lambda p - \lambda pz) = E(z^{N_{\lambda p}(T_\alpha)})). \tag{2.11} \]

**Remark 2.1** \( G(z) \) is the PGF of the number of arrivals of a Poisson process with arrival rate \( \lambda p \) within an independent random time \( T_\alpha \), where \( T_\alpha \) has the same probability distribution as that for the busy period of the standard \( M/G/1 \) queue with arrival rate \( \lambda q \) and service time \( T_\beta \).

Our tail asymptotic analysis is based on the assumption that the service time \( T_\beta \) has a so-called regularly varying tail at \( \infty \).

**Definition 2.1** (e.g., Bingham et al. [8]) A measurable function \( U : (0, \infty) \to (0, \infty) \) is regularly varying at \( \infty \) with index \( \sigma \in (-\infty, \infty) \) (written \( U \in \mathcal{R}_\sigma \)) iff \( \lim_{t \to \infty} U(xt)/U(t) = x^\sigma \) for all \( x > 0 \). If \( \sigma = 0 \) we call \( U \) slowly varying, i.e., \( \lim_{t \to \infty} U(xt)/U(t) = 1 \) for all \( x > 0 \).

**Definition 2.2** (e.g., Foss et al. [16]) A distribution \( F \) on \((0, \infty)\) belongs to the class of subexponential distribution (written \( F \in \mathcal{S} \)) if \( \lim_{t \to \infty} F^*2(t)/F(t) = 2 \), where \( F = 1 - F \) and \( F^*2 \) denotes the second convolution of \( F \).

It is well known that for a distribution \( F \) on \((0, \infty)\), if \( F \in \mathcal{R}_{-\alpha} \), \( \alpha \geq 0 \), then \( F \in \mathcal{S} \) (see, e.g., Embrechts et al. [11]). Throughout the paper we will make the following basic assumption.

**Assumption A.** The service time \( T_\beta \) has tail probability \( P\{T_\beta > t\} \sim t^{-\alpha}L(t) \) as \( t \to \infty \), where \( \alpha > 1 \) and \( L(t) \) is a slowly varying function at \( \infty \).

**Lemma 2.1** (de Meyer and Teugels [26]) Under Assumption A,
\[ P\{T_\alpha > t\} \sim \frac{1}{(1 - \rho q)^{\alpha - 1}} \cdot t^{-\alpha}L(t) \quad \text{as} \quad t \to \infty. \tag{2.12} \]

The result in (2.12) is straightforward due to Remark 2.1 and the main theorem in [26].

## 3 Stochastic decompositions

In this section, we will verify that \( R_0(z_2), M_a(z_1, z_2), M_b(z_1, z_2) \) and \( M_c(z_2) \) can be viewed as the PGFs of four r.v.s, and at the same time we will find the stochastic decompositions of these r.v.s, which will be used for the asymptotic analysis in later sections.
3.1 Probabilistic interpretation for PGF $R_0(z_2)$

Based on the definition of $R_0$ given earlier, this is a r.v. whose distribution coincides with the conditional distribution of $R_{orb}$ given that $I_{sev} = 0$. In this section, we provide a new probabilistic interpretation for $R_0$, which is useful for our tail asymptotic analysis.

Substituting (2.11) into (2.5), we obtain

$$R_0(z_2) = \exp \left\{ -\frac{\lambda}{\mu} \int_{z_2}^{1} \frac{1 - \alpha(\lambda p - \lambda pu)}{\alpha(\lambda p - \lambda pu) - u} \, du \right\}. \quad (3.1)$$

We now rewrite (3.1). Let

$$\psi = \frac{\rho}{\mu(1-\rho)}, \quad (3.2)$$
$$\kappa(s) = \frac{1 - \rho}{\beta_1} \cdot \frac{1 - \alpha(s)}{s - \lambda p + \lambda p\alpha(s)}, \quad (3.3)$$
$$\tau(s) = \exp \left\{ -\psi \int_0^s \kappa(u) \, du \right\}, \quad (3.4)$$

It is easy to see, from (3.1)–(3.4), that

$$R_0(z_2) = \exp \left\{ -\frac{\lambda}{\mu} \int_0^{\lambda p - \lambda p z_2} \frac{1 - \alpha(s)}{s - \lambda p + \lambda p\alpha(s)} \, ds \right\} = \tau(\lambda p - \lambda p z_2). \quad (3.5)$$

In the following, we show that both $\kappa(s)$ and $\tau(s)$ are the LSTs of two probability distribution functions, respectively. For the first assertion, let $F^{(e)}_{\alpha}(x)$ be the equilibrium distribution of $F^{(e)}_{\alpha}(x)$, which is defined as $F^{(e)}_{\alpha}(x) = \alpha^{-1} \int_0^x (1 - F_{\alpha}(t)) \, dt$ where $\alpha_1 = E(T_{\alpha})$ given in (2.8). The LST of $F^{(e)}_{\alpha}(x)$ can be written as $\alpha^{(e)}(s) = (1 - \alpha(s))/\alpha_1$. From (3.3), we have

$$\kappa(s) = \frac{(1 - \vartheta)\alpha^{(e)}(s)}{1 - \vartheta\alpha^{(e)}(s)} = \sum_{k=1}^{\infty} (1 - \vartheta) \vartheta^{k-1} (\alpha^{(e)}(s))^k, \quad (3.6)$$

where

$$\vartheta = \lambda p\alpha_1 = \rho p/(1 - \rho q) < 1. \quad (3.7)$$

The following remark confirms the assertion.

**Remark 3.1** Let $T_\kappa$ be the geometric sum of i.i.d. random variables $T^{(e)}_{\alpha,j}$, $j \geq 1$, each with the distribution $F^{(e)}_{\alpha}(x)$; or more precisely, $T_\kappa \overset{d}{=} T^{(e)}_{\alpha,1} + T^{(e)}_{\alpha,2} + \cdots + T^{(e)}_{\alpha,J}$, where $P(J = j) = (1 - \vartheta)\vartheta^{j-1}$, $j \geq 1$ and $J$ is independent of $T^{(e)}_{\alpha,j}$ for $j \geq 1$. Then, immediately from (3.6), $\kappa(s)$ can be viewed as the LST of distribution function of the r.v. $T_\kappa$, whose distribution function is denoted by $F_\kappa(x)$. In the above, $\overset{d}{=} \; \text{"\(\text{d}\) as an} \; \text{equalit y in probability distribution.} \; \text{used (which will be used later again) to mean the equality in probability distribution.} \; \text{\(\text{d}\) as an} \;$$
For the second assertion that $\tau(s)$ is the LST of a probability distribution function, denoted by $F_\tau(x)$, on $[0, \infty)$, by Theorem 1 in Feller [15] (pp.439) this is true as long as $\tau(s)$ is completely monotone, i.e., $\tau(s)$ possesses derivatives $\tau^{(n)}(s)$ of all orders such that $(-1)^n \tau^{(n)}(s) \geq 0$ for $s > 0$, and $\tau(0) = 1$. A detailed proof can be found in [29]. The following remark provides the detailed interpretation.

**Remark 3.2** Let $T_\tau$ be a r.v. having the distribution $F_\tau(x)$. Since $\tau(s)$ is the LST of the probability distribution $F_\tau(x)$, the expression in (3.3) implies that $R_0$ can be regarded as the number of Poisson arrivals with rate $\lambda p$ within an independent random time $T_\tau$, i.e., $R_0 \overset{d}{=} N_{\lambda p}(T_\tau)$.

### 3.2 Probabilistic interpretation for PGF $M_c(z_2)$

Recall that $M_c(z_2)$ is defined in (2.3). In this section, we show that $M_c(z_2)$ is the PGF of a r.v., denoted by $M_c$, or $M_c(z_2) = E(z_2^{M_c})$, and provide a probabilistic interpretation for $M_c$.

It follows from (2.3) and (2.11) that

$$M_c(z_2) = \frac{(1 - \theta) \cdot (1 - z_2)}{\alpha(\lambda p - \lambda p z_2) - z_2} = \frac{1 - \theta}{1 - \theta \alpha(e) (\lambda p - \lambda p z_2) - z_2}$$

$$= \sum_{k=0}^{\infty} (1 - \theta) \theta^k (\alpha(e) (\lambda p - \lambda p z_2))^k$$

$$= 1 - \theta + \theta \cdot \kappa(\lambda p - \lambda p z_2). \quad (3.8)$$

Define

$$T_\eta = \begin{cases} 0, & \text{with probability } 1 - \theta, \\ T_\kappa, & \text{with probability } \theta. \end{cases} \quad (3.9)$$

Denote by $\eta(s)$ the LST of the probability distribution function of $T_\eta$. Immediately, $\eta(s) = 1 - \theta + \theta \cdot \kappa(s)$, which, together with (3.8), yields

$$M_c(z_2) = \eta(\lambda p - \lambda p z_2). \quad (3.10)$$

The following remark provides a probabilistic interpretation for $M_c$.

**Remark 3.3** With the same argument as that in Remark 3.2, (3.10) allows us to interpret the r.v. $M_c$ as the number of Poisson arrivals with rate $\lambda p$ within an independent random time $T_\eta$, i.e., $M_c \overset{d}{=} N_{\lambda p}(T_\eta)$.

### 3.3 Probabilistic interpretation for PGF $M_a(z_1, z_2)$

In this section, we prove that $M_a(z_1, z_2)$, defined in (2.1) is the PGF of a two-dimensional r.v., denoted by $(M_{a1}, M_{a2})$, or $M_a(z_1, z_2) = E(z_1^{M_{a1}} z_2^{M_{a2}})$. Denote by $F_\beta^{(e)}(x)$ the equilibrium distribution of $F_\beta(x)$, that is, $F_\beta^{(e)}(x) = \beta_1^{-1} \int_0^x (1 - F_\beta(t)) dt$. The LST of $F_\beta^{(e)}(x)$ can be written as $\beta^{(e)}(s) = (1 - \beta(s))/(\beta_1 s)$. It follows from (2.1) that

$$M_a(z_1, z_2) = \beta^{(e)}(\lambda - \lambda p z_2 - \lambda q z_1). \quad (3.11)$$
For simplicity, we introduce the following concept:

**Definition 3.1** Let \( N \) be a non-negative integer valued r.v., let \( \{X_k\}_{k=1}^{\infty} \) be a sequence of i.i.d. Bernoulli r.v.s having a common distribution \( P\{X_k = 1\} = c \) and \( P\{X_k = 0\} = 1 - c \) where \( 0 < c < 1 \), which is assumed to be independent of \( N \). The two-dimensional r.v. \((\sum_{k=1}^{N} X_k, N - \sum_{k=1}^{N} X_k)\), where \( \sum_{k=1}^{0} = 0 \), is called an independent \((c, 1 - c)\)-splitting of \( N \), denoted by \( \text{split}(N; c, 1 - c) \).

The following remark provides a definition for \((M_{a_1}, M_{a_2})\), together with its probabilistic interpretation.

**Remark 3.4** \((M_{a_1}, M_{a_2}) \overset{d}{=} \text{split}(N_{\lambda}(T_{\beta}^{(e)}); q, p)\), which can be easily checked because the right hand side of (3.11) can be written as \( \int_{0}^{\infty} \sum_{n=0}^{\infty} [\sum_{k=0}^{n} z_1^k z_2^{-k-n} (n!)^k q^{n-k} ((\lambda x)^{n}/n!) e^{-\lambda x} dF_{\beta}(x) ] \). Therefore, \( M_{a}(z_1, z_2) \) is the PGF of Poisson arrivals with rate \( \lambda \), split into two components according to the independent \((c, 1 - c)\)-splitting defined above, within an independent random time \( T_{\beta}^{(e)} \) whose distribution coincides with the equilibrium distribution \( F_{\beta}(x) \).

### 3.4 Probabilistic interpretation for PGF \( M_{b}(z_1, z_2) \)

In this section, we prove that \( M_{b}(z_1, z_2) \), defined in (2.2) is the PGF of a two-dimensional r.v., denoted by \((M_{b_1}, M_{b_2})\), or \( M_{b}(z_1, z_2) = E(z_1^{M_{b_1}} z_2^{M_{b_2}}) \). Let

\[
H(z_1, z_2) = \frac{1}{\rho q} \cdot \frac{\beta(\lambda - \lambda p z_2 - \lambda q z_1) - G(z_2)}{z_1 - G(z_2)}. \tag{3.12}
\]

It follows from (2.2) that

\[
M_{b}(z_1, z_2) = \frac{1 - \rho q}{1 - \rho q H(z_1, z_2)} = \sum_{n=0}^{\infty} (1 - \rho q)(\rho q)^n (H(z_1, z_2))^n, \tag{3.13}
\]

Clearly, \((M_{b_1}, M_{b_2})\) will be a random sum of two-dimensional r.v.s. provided that \( H(z_1, z_2) \) can be verified to be the PGF of a two-dimensional r.v. To this end, let us write (3.12) in a power series. Let

\[
b_k = \int_{0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} dF_{\beta}(t), \quad k \geq 0, \quad \text{and} \quad \overline{b}_m = \sum_{k=m}^{\infty} b_k, \quad m \geq 1. \tag{3.14}
\]

Hence

\[
\beta(\lambda - \lambda p z_2 - \lambda q z_1) = \int_{0}^{\infty} \sum_{k=0}^{\infty} \frac{(\lambda(pz_2 + qz_1)t)^k}{k!} e^{-\lambda t} dF_{\beta}(t) = \beta(\lambda) + \sum_{k=1}^{\infty} b_k(pz_2 + qz_1)^k. \tag{3.15}
\]

By (2.4) and (3.15),

\[
G(z_2) = \beta(\lambda - \lambda p z_2 - \lambda q G(z_2)) = \beta(\lambda) + \sum_{k=1}^{\infty} b_k(pz_2 + q G(z_2))^k. \tag{3.16}
\]
Substituting (3.15) and (3.16) into the numerator of the right-hand side of (3.12), we obtain

\[
H(z_1, z_2) = \frac{1}{pq} \cdot \sum_{k=1}^{\infty} b_k \left( (pz_2 + qz_1)^k - (pz_2 + qG(z_2))^k \right) \left( z_1 - G(z_2) \right)
\]

\[
= \frac{1}{\rho} \cdot \sum_{k=1}^{\infty} b_k \sum_{m=1}^{\infty} (pz_2 + qz_1)^{m-1} (pz_2 + qG(z_2))^{k-m}
\]

\[
= \frac{1}{\rho} \cdot \sum_{m=1}^{\infty} (pz_2 + qz_1)^{m-1} \sum_{k=m}^{\infty} b_k (pz_2 + qG(z_2))^{k-m}.
\]

Further, let

\[
D_m(z_2) = \sum_{k=m}^{\infty} b_k (pz_2 + qG(z_2))^{k-m} = \sum_{k=0}^{\infty} \frac{b_{m+k}}{\rho} (pz_2 + qG(z_2))^k.
\]

It is obvious that \(\sum_{k=0}^{\infty} b_{m+k}/\rho = 1\) and \(pz_2 + qG(z_2)\) is the PGF of a r.v. So \(D_m(z_2)\) is the PGF of a r.v., denoted by \(D_m\). Substituting (3.18) into (3.17), we have

\[
H(z_1, z_2) = \sum_{m=1}^{\infty} \frac{b_m}{\rho} (pz_2 + qz_1)^{m-1} D_m(z_2).
\]

Note that \(pz_2 + qz_1\) is the PGF of a two-dimensional r.v., and

\[
\sum_{m=1}^{\infty} b_m = \sum_{m=1}^{\infty} mb_m = \sum_{m=1}^{\infty} \int_0^{\infty} \lambda^m (m-1)! e^{-\lambda t} dB(t) = \lambda \int_0^{\infty} t dB(t) = \rho.
\]

Namely, \(\sum_{m=1}^{\infty} b_m/\rho = 1\), which together with (3.19) imply that \(H(z_1, z_2)\) is the PGF of a mixed two-dimensional PGFs. So it is the PGF of a two-dimensional r.v.

Precisely, the implication of the above argument is summarized in the following two remarks.

**Remark 3.5** Suppose that, for each \(m \geq 1\), \(D_m\) is a r.v. with the PGF given in (3.15). Let \(\{(Y_1^{(i)}, Y_2^{(i)})\}_{i=1}^{\infty}\) be i.i.d. two-dimensional r.v.s having a common PGF \(E(z_1^{Y_1^{(i)}}, z_2^{Y_2^{(i)}}) = pz_2 + qz_1\). Assume that \(D_m, m \geq 1\), are independent of \(\{(Y_1^{(i)}, Y_2^{(i)})\}_{i=1}^{\infty}\). It follows from (3.19) that

\[
(H_1, H_2) \overset{d}{=} \sum_{i=1}^{m-1} (Y_1^{(i)}, Y_2^{(i)}) + (0, D_m) \quad \text{with probability } b_m/\rho \text{ for each } m \geq 1.
\]

**Remark 3.6** It follows from (3.13) that \((M_{b_1}, M_{b_2})\) is a random sum of i.i.d. two-dimensional r.v.s \((H_1^{(i)}, H_2^{(i)})\), \(i \geq 1\), each with the same PGF \(H(z_1, z_2)\); or more precisely,

\[
(M_{b_1}, M_{b_2}) \overset{d}{=} \begin{cases} 
0, & \text{with probability } 1 - \rho q, \\
\sum_{i=1}^{J} (H_1^{(i)}, H_2^{(i)}) & \text{with probability } \rho q,
\end{cases}
\]

where \(P(J = n) = (1 - \rho q)(\rho q)^{n-1}, n \geq 1\) and \(J\) is independent of \((H_1^{(i)}, H_2^{(i)})\) for \(i \geq 1\).
4 Tail Asymptotics

In this section, we will study the asymptotic behavior for the tail probabilities \( P\{R_{orb} > j | I_{sev} = 0 \} \), \( P\{R_{que} > j | I_{sev} = 1 \} \), \( P\{R_{orb} > j | I_{sev} = 1 \} \), \( P\{R_{que} > j \} \), \( P\{R_{orb} > j \} \), \( P\{L_{\mu} > j \} \) and \( P\{R_{orb} > j | R_{que} = 0, I_{sev} = 1 \} \) as \( j \to \infty \), respectively, where \( L_{\mu} \) represents the total number of customers in the system. The following literature asymptotic properties will be needed in our analysis. They are for the tail probability of the geometric sum of i.i.d. subexponential r.v.s, the Poisson number within a heavy-tailed random time, and the sum of independent subexponential r.v.s, respectively.

Lemma 4.1 (pp.580–581 in [11]) Let \( N \) be a r.v. with \( P\{N = k\} = (1 - \sigma)^{k-1}, 0 < \sigma < 1, k \geq 1, \) and \( \{Y_k\}_{k=1}^{\infty} \) be a sequence of non-negative, i.i.d. r.v.s having a common subexponential distribution \( F \). Define \( S_n = \sum_{k=1}^{n} Y_k \). Then
\[
P\{S_n > t\} \sim \frac{1}{1 - \sigma}(1 - F(t)), \quad t \to \infty.
\] (4.1)

Lemma 4.2 (Proposition 3.1 in [6]) Let \( N_\lambda(t) \) be a Poison process with rate \( \lambda \) and let \( T \) be a positive r.v. with distribution \( F \), which is independent of \( N_\lambda(t) \). If \( F(t) = P\{T > t\} \) is heavier than \( e^{-\sqrt{t}} \) as \( t \to \infty \), then \( P\{N_\lambda(T) > j\} \sim P\{T > j/\lambda\} \) as \( j \to \infty \).

Lemma 4.3 (p.48 in [16]) Let \( F \), \( F_1 \) and \( F_2 \) be distribution functions. Suppose that \( F \in S \). If \( F_i(t)/F(t) \to c_i \) as \( t \to \infty \) for some \( c_i \geq 0 \), \( i = 1, 2 \), then \( F_1 * F_2(t)/F(t) \to c_1 + c_2 \) as \( t \to \infty \), where the symbol \( F \overset{\text{def}}{=} 1 - F \) and “\( F_1 * F_2 \)” stands for the convolution of \( F_1 \) and \( F_2 \).

4.1 Asymptotic tail probability for \( P\{R_{orb} > j | I_{sev} = 0 \} \)

In this subsection, we will study the asymptotic behavior for the tail probability \( P\{R_{orb} > j | I_{sev} = 0 \} \equiv P\{R_0 > j \} \) as \( j \to \infty \). By Lemma 2.4, we know \( P\{T_\alpha > t\} \sim (1 - \rho q)^{-a-1} t^{-a} L(t) \) as \( t \to \infty \), where the r.v. \( T_\alpha \) is the busy period defined in Section 2. Applying Karamata’s theorem (e.g., p.28 in [5]), we have \( \int_{t}^{\infty} (1 - F_\alpha(x)) dx = (a - 1)^{-1} (1 - \rho q)^{-a-1} t^{-a+1} L(t), \) which implies \( 1 - F_\alpha^{(e)}(t) \sim ((a - 1)\alpha_1)^{-1} (1 - \rho q)^{-a-1} t^{-a+1} L(t), t \to \infty \). By Remark 3.1 and applying Lemma 4.1, we have
\[
P\{T_\kappa > t\} \sim \frac{1}{\alpha_1(1 - \vartheta)(a - 1)(1 - \rho q)^{a+1}} \cdot t^{-a+1} L(t)
= \frac{1}{\beta_1(1 - \rho)(a - 1)(1 - \rho q)^{a-1}} \cdot t^{-a+1} L(t), \quad t \to \infty.
\] (4.2)

Note that \( T_\kappa \) has the distribution function \( F_\tau(x) \) defined in terms of its LST \( \tau(s) \) in [34]. Therefore, the tail probability of \( T_\tau \) is determined by the tail probability of \( T_\kappa \). The following
asymptotic result is proved in [29] (Theorem 2):

\[ P\{T_r > t\} \sim \frac{\lambda}{a\mu(1 - \rho)^2(1 - \rho q)^{a - 1}} \cdot t^{-a}L(t), \quad t \to \infty. \quad (4.3) \]

Recall Remark 3.2, we know \( R_0 = N_{\lambda p}(T_r) \). Lemma 4.1 leads to

\[ P\{R_0 > j\} \sim P\{T_r > j/(\lambda p)\} \sim \frac{\lambda(\lambda p)^a}{a\mu(1 - \rho)^2(1 - \rho q)^{a - 1}} \cdot j^{-a}L(j), \quad j \to \infty. \quad (4.4) \]

### 4.2 Asymptotic tail probability for \( P\{R_{que} > j | I_{sev} = 1\} \)

In this subsection, we will study the asymptotic behavior of the tail probability \( P\{R_{que} > j | I_{sev} = 1\} \equiv P\{R_{11} > j\} \) as \( j \to \infty \). Note that \( E(z_1^{R_{11}}) = R_1(z_1, 1) \). Taking \( z_2 \to 1 \) in (2.3), (2.5) and (2.6) and using the fact that \( \frac{d}{dz} G(z)|_{z=1} = \rho p/(1 - \rho q) \), we immediately have \( M_e(1) = 1, R_0(1) = 1 \) and

\[ E(z_1^{R_{11}}) = M_a(z_1, 1)M_b(z_1, 1). \quad (4.5) \]

It follows from (2.1), (3.13) and (3.12) that

\[
\begin{align*}
M_a(z_1, 1) & = \frac{1}{\rho} \cdot \frac{1 - \beta(\lambda q - \lambda q z_1)}{q - q z_1} = \beta(e)(\lambda q - \lambda q z_1), \\
M_b(z_1, 1) & = \sum_{n=0}^{\infty} (1 - \rho q)(\rho q)^n (H(z_1, 1))^n, \\
H(z_1, 1) & = \frac{1}{\rho q} \cdot \frac{\beta(\lambda q - \lambda q z_1) - 1}{z_1 - 1} = \beta(e)(\lambda q - \lambda q z_1).
\end{align*}
\]

Substituting (4.8) into (4.7) and applying (4.6) and (4.5), we have

\[ E(z_1^{R_{11}}) = \xi(\lambda q - \lambda q z_1), \quad (4.9) \]

where

\[ \xi(s) = \sum_{n=1}^{\infty} (1 - \rho q)(\rho q)^{n-1}(\beta(e)(s))^n. \quad (4.10) \]

**Remark 4.1** Immediately from [41.10], \( \xi(s) \) can be viewed as the LST of the distribution function of \( r.v. T_{\xi} \overset{d}{=} T_{\beta,1}^{(e)} + T_{\beta,2}^{(e)} + \cdots + T_{\beta,j}^{(e)}, \) where \( T_{\beta,j}^{(e)}, j \geq 1, \) are i.i.d. \( r.v.s. \) with a common distribution \( F_{\beta}^{(e)}(x) \), \( P(J = j) = (1 - \rho q)(\rho q)^{j-1}, j \geq 1, \) and \( J \) is independent of \( T_{\beta,j}^{(e)}, j \geq 1. \)

Under Assumption A, by Karamata’s theorem (e.g., p.28 in [8]), we have \( \int_t^{\infty} (1 - F_{\beta}(x))dx \sim (a - 1)^{-1}t^{-a+1}L(t) \), which implies \( 1 - F_{\beta}^{(e)}(t) \sim ((a - 1)\beta_1)^{-1}t^{-a+1}L(t), t \to \infty. \) By Remark 4.1 and applying Lemma 4.1, we have

\[ P\{T_{\xi} > t\} \sim \frac{1}{(1 - \rho q)(a - 1)\beta_1} \cdot t^{-a+1}L(t), \quad t \to \infty. \quad (4.11) \]
Remark 4.2 With (4.9), one can interpret $R_{11}$ as the number of Poisson arrivals with rate $\lambda q$ within an independent random time $T_\xi$, i.e., $R_{11} \overset{d}{=} N_q(T_\xi)$.

By Remark 4.2 and applying Lemma 4.2, we have

$$P\{R_{11} > j\} \sim P\{T_\xi > j/(\lambda q)\} \sim \frac{(\lambda q)^{a-1}}{(1-\rho)(a-1)\beta_1} \cdot j^{-a+1}L(j), \quad j \to \infty. \quad (4.12)$$

4.3 Asymptotic tail probability for $P\{R_{orb} > j|I_{sev} = 1\}$

In this subsection, we will study the asymptotic behavior of tail probability $P\{R_{orb} > j|I_{sev} = 1\} \equiv P\{R_{12} > j\}$ as $j \to \infty$. Note that $E(z_{R_{12}}^2) = R_{1}(1,z)$. Taking $z_1 \to 1$ in (2.1)–(2.3), we have

$$M_a(1,z_2) \cdot M_b(1,z_2) \cdot M_c(z_2) = \frac{1-\rho}{\rho} \cdot \frac{1-\beta(\lambda p - \lambda p z_2)}{\rho - p z_2} \cdot \frac{G(z_2) - 1}{\beta(\lambda p - \lambda p z_2) - 1} \cdot \frac{1 - z_2}{G(z_2) - z_2}.$$  

Substituting (2.11) into (4.13) gives

$$M_a(1,z_2) \cdot M_b(1,z_2) \cdot M_c(z_2) = \frac{1-\rho}{\rho} \cdot \frac{1-\alpha(\lambda p - \lambda p z_2)}{\beta(\lambda p - \lambda p z_2) - 1} \cdot \frac{1 - z_2}{G(z_2) - z_2}.$$  

where the last equality follows from (3.3).

By (2.6), (4.14) and (3.5),

$$E(z_{R_{12}}^2) = M_a(1,z_2) \cdot M_b(1,z_2) \cdot M_c(z_2) \cdot R_0(z_2) = \kappa(\lambda p - \lambda p z_2) \cdot \tau(\lambda p - \lambda p z_2). \quad (4.15)$$

Remark 4.3 With (4.15), one can interpret $R_{12}$ as the number of Poisson arrivals with rate $\lambda p$ within an independent random time $T_\kappa + T_\tau$, i.e., $R_{12} \overset{d}{=} N_\lambda(T_\kappa + T_\tau)$, where $T_\kappa$ and $T_\tau$ are assumed to be independent.

By (4.2) and (4.3), and applying Lemma 4.3, we have

$$P\{T_\kappa + T_\tau > t\} \sim P\{T_\kappa > t\} \sim \frac{1}{\beta_1(1-\rho)(a-1)(1-\rho q)^{a-1}} \cdot t^{-a+1}L(t), \quad t \to \infty. \quad (4.16)$$

Applying Remark 4.3 and Lemma 4.2, we have

$$P\{R_{12} > j\} \sim P\{T_\kappa + T_\tau > j/(\lambda p)\} \sim \frac{(\lambda p)^{a-1}}{\beta_1(1-\rho)(a-1)(1-\rho q)^{a-1}} \cdot j^{-a+1}L(j), \quad j \to \infty. \quad (4.17)$$
4.4 Asymptotic tail probabilities for $P\{\text{R}_{\text{que}} > j\}$ and $P\{\text{R}_{\text{orb}} > j\}$

In this subsection, we will study the asymptotic behavior for the tail probabilities $P\{\text{R}_{\text{que}} > j\}$ and $P\{\text{R}_{\text{orb}} > j\}$ as $j \to \infty$. Note that

$$P\{\text{R}_{\text{que}} > j\} = \rho P\{\text{R}_{\text{orb}} > j|I_{\text{sev}} = 0\} + \rho P\{\text{R}_{\text{que}} > j|I_{\text{sev}} = 1\} = \rho P\{R_{11} > j\}. \quad (4.18)$$

It follows from (4.18) and (4.12) that

$$P\{\text{R}_{\text{que}} > j\} \sim \frac{\lambda(\lambda q)^{a-1}}{(1 - \rho q)(a - 1)} \cdot j^{-a+1}L(j), \quad j \to \infty. \quad (4.19)$$

Note that

$$P\{\text{R}_{\text{orb}} > j\} = (1 - \rho)P\{\text{R}_{\text{orb}} > j|I_{\text{sev}} = 0\} + \rho P\{\text{R}_{\text{orb}} > j|I_{\text{sev}} = 1\} = (1 - \rho)P\{R_{0} > j\} + \rho P\{R_{12} > j\}. \quad (4.20)$$

By (4.4) and (4.17), we know $P\{R_{0} > j\} = o(P\{R_{12} > j\})$ as $j \to \infty$. Applying Lemma 4.3 to (4.20), we have

$$P\{\text{R}_{\text{orb}} > j\} \sim \rho P\{R_{12} > j\} \sim \frac{\lambda(\lambda p)^{a-1}}{(1 - \rho)(a - 1)(1 - \rho q)^{a-1}} \cdot j^{-a+1}L(j), \quad j \to \infty. \quad (4.21)$$

4.5 Asymptotic tail probability for $P\{L_{\mu} > j\}$ and a tail equivalence property

We are now ready to study the asymptotic behavior for the tail probability $P\{L_{\mu} > j\}$ as $j \to \infty$, where $L_{\mu}$ represents the total number of customers in the retrial queueing system with Bernoulli schedule (including the possible one in the service), considered in this paper. The subscript $\mu$ in the notation $L_{\mu}$ is used to indicate the retrial rate. Naturally, the notation $L_{\infty}$ represents the total number of customers in the standard $M/G/1$ queueing system (without retrials). By the definitions of $\text{R}_{\text{que}}$ and $\text{R}_{\text{orb}}$, we know

$$L_{\mu} \overset{\text{def}}{=} \begin{cases} \text{R}_{\text{orb}}, & \text{with probability } 1 - \rho, \\ 1 + \text{R}_{\text{que}} + \text{R}_{\text{orb}}, & \text{with probability } \rho. \end{cases} \quad (4.22)$$

It follows that

$$E(z^{L_{\mu}}) = (1 - \rho)E(z^{R_{\text{orb}}|I_{\text{sev}} = 0}) + \rho E(z^{R_{\text{que}} + R_{\text{orb}}|I_{\text{sev}} = 1}) = (1 - \rho)R_{0}(z) + \rho z R_{1}(z, z). \quad (4.23)$$

Setting $z_{1} = z_{2} = z$ in (3.5) and (2.6), we have

$$R_{0}(z) = \tau(\lambda p - \lambda pz), \quad (4.24)$$

$$R_{1}(z, z) = M_{a}(z, z) \cdot M_{b}(z, z) \cdot M_{c}(z) \cdot R_{0}(z) = \frac{1 - \rho}{\rho} \cdot \frac{1 - \beta(\lambda - \lambda z)}{\beta(\lambda - \lambda z) - z} \cdot R_{0}(z), \quad (4.25)$$
where the last equality follows from (2.1)–(2.3). Substituting (4.25) into (4.23), we have

\[ E(z^{L_\mu}) = \frac{(1 - \rho)(1 - z)}{\beta(\lambda - \lambda z)} = \beta(\lambda - \lambda z) R_0(z) = E(z^{L_\infty}) \cdot E(z^{R_0}). \]  

\[ (4.26) \]

Therefore, \( L_\mu \) can be written as the sum of two independent r.v.s, i.e., \( L_\mu \overset{d}{=} L_\infty + R_0 \), which is a well known result for the \( M/G/1 \) retrial queue. Such a stochastic decomposition is often used to establish the asymptotic equivalence of tail probabilities under the assumption of the heavy-tailed service time (e.g., [34], [35] and [33]):

\[ P\{L_\mu > j\} \sim P\{L_\infty > j\} \sim \frac{\lambda a}{(a - 1)(1 - \rho)} \cdot j^{-a + 1} L(j) \quad \text{as} \quad j \to \infty. \]  

\[ (4.27) \]

In addition to the above asymptotic equivalence property, we now present the following refinement:

\[ P\{L_\mu > j\} - P\{L_\infty > j\} \sim P\{R_0 > j\} \sim \frac{\lambda(\lambda p)^a}{a \mu (1 - \rho)^2 (1 - \rho q)^{a - 1}} \cdot j^{-a} L(j) \quad \text{as} \quad j \to \infty. \]  

\[ (4.28) \]

In fact, based on (1.4) and the fact \( P\{R_0 > j\} = o(P\{L_\infty > j\}) \), (4.28) can be verified directly by using Lemma 6.1 in [28].

### 4.6 Asymptotic tail probability for \( P\{R_{orb} > j|I_{sev} = 1, R_{que} = 0\} \)

In previous subsections 4.1 and 4.3 it has been proved that \( P\{R_{orb} > j|I_{sev} = 0\} \) and \( P\{R_{orb} > j|I_{sev} = 1\} \) are regularly varying with index \(-a\) and \(-a + 1\), respectively (the latter is one degree heavier than the former). In this subsection, we consider an extreme case, in which the server is busy but the queue is empty, and study the asymptotic behavior of \( P\{R_{orb} > j|I_{sev} = 1, R_{que} = 0\} \equiv P\{R_{12} > j|R_{11} = 0\} \) as \( j \to \infty \).

Immediately from (2.6), we are able to decomposes \((R_{11}, R_{12})\) into the four independent components as follows:

\[ (R_{11}, R_{12}) \overset{d}{=} (M_{a1}, M_{a2}) + (M_{b1}, M_{b2}) + (0, M_c) + (0, R_0). \]  

\[ (4.29) \]

The tail asymptotic behavior of \( R_0 \) has already been obtained in Subsection 4.1 Below, we will focus on r.v.s. \((M_{a1}, M_{a2})\), \((M_{b1}, M_{b2})\) and \(M_c\).

#### 4.6.1 Tail probability \( P\{M_{a2} > j|M_{a1} = 0\} \)

We refer to any distribution function \( F(x) \) (or r.v. \( X \)) as light-tailed if \( E(e^{\varepsilon X}) < \infty \) for some \( \varepsilon > 0 \), which includes, as the classic example, the exponential distribution function, as well as all bounded r.v.s.

**Lemma 4.4** Let \( N \) be a discrete r.v. with an arbitrary distribution \( p_n = P\{N = n\}, n \geq 0 \). If \((N_1, N_2) = \text{split}(N; 1 - c, c), 0 < c < 1\), then the conditional tail distribution \( P\{N_2 > j|N_1 = k\} \) is a light-tailed as \( j \to \infty \) for any fixed \( k \geq 0 \), or we simply say that \( N_2|N_1 = k \) has a lighted-tail distribution.
Proof. Since

\[ P\{N_1 = k, N_2 > j\} = \sum_{n=j+1}^{\infty} p_{k,n} \binom{k+n}{k} (1-c)^k c^n \leq (1-c)^k \sum_{n=j+1}^{\infty} \binom{k+n}{k} c^n, \]  
(4.30)

there exists some \( d \in (c, 1) \) such that \( P\{N_1 = k, N_2 > j\} = O(d^j) \) for large \( j \geq 0 \). \( \square \)

Remark 4.4 By Remark 3.4, we know \( (M_1, M_2) \overset{\text{d}}{=} \text{split}(N_\lambda(T_\beta^e); q, p) \). So \( M_2|M_1 = 0 \) has a light-tailed probability distribution.

4.6.2 Tail probability \( P\{M_{b_2} > j|M_{b_1} = 0\} \)

By Remark 3.4, \( (M_{b_1}, M_{b_2}) \) can be decomposed to a random sum of r.v.s which have the same probability distribution as that for \((H_1, H_2)\).

Lemma 4.5 (Grandell [17], pp. 162–163) Let \( N \) be a discrete non-negative integer-valued r.v. with mean value \( \mu_N \), and \( \{Y_i\}_{i=1}^\infty \) be a sequence of non-negative i.i.d. r.v.s with mean value \( \mu_Y \). Define \( S_0 = 0 \) and \( S_n = \sum_{k=1}^{n} Y_i \). If \( P\{Y_i > x\} \sim cy^{-h}L(x) \) as \( x \to \infty \) and \( P\{N > m\} \sim c_N m^{-h}L(m) \) as \( m \to \infty \), where \( h > 1 \), \( c_Y \geq 0 \) and \( c_N \geq 0 \), then \( P\{S_N > x\} \sim (c_N \mu_Y^h + \mu_N c_Y)x^{-h}L(x) \) as \( x \to \infty \).

Remark 4.5 It is a convention that in Lemma 4.5 \( c_Y = 0 \) and \( c_N = 0 \) when \( \lim_{t \to \infty} P\{Y_i > x\}/(x^{-h}L(x)) = 0 \) and \( \lim_{m \to \infty} P\{N > m\}/(m^{-h}L(m)) = 0 \), respectively.

Recall 3.15. Note that \( D_m \) can be regarded as a random sum \( D_m = \sum_{i=0}^{N} X_i \), where \( X_0 = 0 \), \( N \) is an independent r.v. with probability distribution \( P\{N = k\} = b_{m+k}/b_m \), \( k \geq 0 \), and \( X_i, i \geq 1 \) have a common distribution with PGF \( E(z^{X_i}) = pz + qG(z) = pz + qE(z^{N_\lambda p(T_\alpha)}) \). By Lemma 2.1 and Lemma 4.2

\[ P\{X_i > j\} \sim qP\{N_\lambda p(T_\alpha) > j\} \sim \frac{q(\lambda p)^a}{(1 - pq)^{a+1}} \cdot j^{-a}L(j) \quad \text{as} \quad j \to \infty. \]  
(4.31)

Also,
\[ E(X) = p + q\rho p/(1 - pq) = p/(1 - pq) < \infty, \]
\[ E(N) = (1/b_m) \sum_{k=1}^{\infty} k b_{m+k} = (1/b_m) \sum_{k=1}^{\infty} b_{m+k} < \infty, \]

where the second series is convergent because \( b_{m+j} \sim \lambda^a j^{-a} L(j) \) as \( j \to \infty \) and \( a > 1 \).

In virtue of (4.31) and (4.32), we are able to apply Lemma 4.5 to obtain
\[ P\{D_m > j\} \sim \left[ \frac{q(\lambda p)^a}{b_m(1 - pq)^a} + \sum_{k=1}^{\infty} \frac{b_{m+k}}{b_m} \cdot \frac{q(\lambda p)^a}{(1 - pq)^{a+1}} \right] \cdot j^{-a}L(j). \]  
(4.33)
Next, we will study the tail probability of \((H_1, H_2)\). It follows from (3.21) that

\[
\begin{align*}
h_0 \defeq P\{H_1 = 0\} &= \sum_{m=1}^{\infty} \left( \frac{b_m}{\rho} \right) P\left\{ \sum_{i=1}^{m-1} Y_1^{(i)} = 0 \right\} = \frac{1}{\rho} \sum_{m=1}^{\infty} b_m \rho^{m-1}, \\
P\{H_2 > j|H_1 = 0\} &= \sum_{m=1}^{\infty} \left( \frac{b_m}{\rho h_0} \right) P\left\{ \sum_{i=1}^{m-1} Y_1^{(i)} = 0, \sum_{i=1}^{m-1} Y_2^{(i)} + D_m > j \right\} \\
&= \sum_{m=1}^{\infty} \left( \frac{b_m}{\rho h_0} \right) \rho^{m-1} P\left\{ \sum_{i=1}^{m-1} Y_2^{(i)} + D_m > j | \sum_{i=1}^{m-1} Y_1^{(i)} = 0 \right\} \\
&= \sum_{m=1}^{\infty} \left( \frac{b_m}{\rho h_0} \right) \rho^{m-1} P\left\{ m - 1 + D_m > j \right\}.
\end{align*}
\]

Further, by the dominated convergence theorem, we have

\[
P\{H_2 > j|H_1 = 0\} \sim \left[ \sum_{m=1}^{\infty} \left( \frac{\lambda p}{\rho h_0} \right)^m + \sum_{k=1}^{\infty} \frac{b_{m+k}}{\rho h_0} - \frac{q(\lambda p)}{(1 - \rho q) a+1} \right] \cdot \rho^{m-1} J^{-a} L(j),
\]

Now we are ready to study \(P\{M_{b2} > j|M_{b1} = 0\}\). Recall (3.22). Since \(P\{\sum_{i=1}^{n} H_1^{(i)} = 0\} = P\{H_1^{(1)} = \cdots = H_1^{(n)} = 0\} = h_0^n\), we immediately have

\[
P\{M_{b1} = 0\} = \rho q \sum_{n=1}^{\infty} \left(1 - \rho q\right) (\rho q)^{n-1} h_0^n, \tag{4.36}
\]

\[
P\{M_{b1} = 0, M_{b2} > j\} = \rho q \sum_{n=1}^{\infty} \left(1 - \rho q\right) (\rho q)^{n-1} h_0^n \cdot P\left\{ \sum_{i=1}^{n} H_2^{(i)} > j | \sum_{i=1}^{n} H_1^{(i)} = 0 \right\} \\
\sim \rho q \sum_{n=1}^{\infty} n(1 - \rho q) (\rho q)^{n-1} h_0^n \cdot nP\{H_2 > j|H_1 = 0\}, \tag{4.37}
\]

where we used the fact (property of subexponential distributions) that \(P\{\sum_{i=1}^{n} H_2^{(i)} > j | \sum_{i=1}^{n} H_1^{(i)} = 0\} \sim nP\{H_2 > j|H_1 = 0\}\) as \(j \rightarrow \infty\). Furthermore, by (4.36), (4.37) and (4.35),

\[
P\{M_{b2} > j|M_{b1} = 0\} \sim c_{M_b} \cdot \rho q J^{-a} L(j), \quad j \rightarrow \infty, \tag{4.38}
\]

where \(c_{M_b} > 0\) is a constant.

### 4.6.3 Tail probability \(P\{R_{orb} > j|I_{sev} = 1, R_{que} = 0\}\)

By Remark 3.3, (3.9) and using (4.2), we have

\[
P\{M_c > j\} \sim P\{T \eta > j/(\lambda p)\} = \vartheta P\{T_\kappa > j/(\lambda p)\} \\
\sim \frac{(\lambda p)^a}{(1 - \rho)(a - 1)(1 - \rho q)^a} \cdot j^{-a+1} L(j), \quad j \rightarrow \infty. \tag{4.39}
\]
As pointed out early in (4.29), \((R_{11}, R_{12})\) can be stochastically decomposed into the sum of four independent r.v.s, which leads to

\[
P\{R_{12} > j | R_{11} = 0\} = P\{M_{a2} + M_{b2} + M_c + R_0 > j | M_{a1} = 0, M_{b1} = 0\}. \tag{4.40}
\]

It follow from (4.39), (4.4) and (4.38) and Remark 4.4 that \(P\{R_0 > j\} = o(P\{M_c > j\})\), \(P\{M_{b2} > j | M_{b1} = 0\} = o(P\{M_c > j\})\) and \(P\{M_{a2} > j | M_{a1} = 0\} = o(P\{M_c > j\})\). Applying Lemma 4.3 to (4.40), we have

\[
P\{R_{12} > j | R_{11} = 0\} \sim P\{M_c > j\}, \quad j \to \infty. \tag{4.41}
\]

Note that \(P\{R_{orb} > j | I_{sev} = 1, R_{que} = 0\} \equiv P\{R_{12} > j | R_{11} = 0\}\). By (4.39),

\[
P\{R_{orb} > j | I_{sev} = 1, R_{que} = 0\} \sim \frac{(\lambda p)^a}{(1 - \rho)(a - 1)(1 - \rho q)^a} \cdot j^{-a+1} L(j), \quad j \to \infty. \tag{4.42}
\]

**Acknowledgments**

This work was supported in part by the National Science Foundation of China (Grant No. 71571002), the Natural Science Foundation of the Anhui Higher Education Institutions of China (No. KJ2017A340), the Research Project of Anhui Jianzhu University, and a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

**References**

[1] Artalejo, J.R. (1999a). A classified bibliography of research on retrial queues: Progress in 1990–1999. *Top*, 7(2), 187–211.

[2] Artalejo, J.R. (1999b). Accessible bibliography on retrial queues. *Mathematical and Computer Modelling*, 30(3), 1-6.

[3] Artalejo, J.R. (2010). Accessible bibliography on retrial queues: Progress in 2000–2009. *Mathematical and Computer Modelling*, 51(9), 1071–1081.

[4] Artalejo, J.R., & Gómez-Corral, A. (2008). *Retrial Queueing Systems*. Springer, Berlin.

[5] Artalejo, J.R., & Phung-Duc, T. (2013). Single server retrial queues with two way communication. *Applied Mathematical Modelling*, 37(4), 1811–1822.

[6] Asmussen, S., Klüppelberg, C., & Sigman, K. (1999). Sampling at subexponential times, with queueing applications. *Stochastic Processes and their Applications*, 79(2), 265–286.

[7] Atencia, I., & Moreno, P. (2005). A single-server retrial queue with general retrial times and Bernoulli schedule. *Applied Mathematics and Computation*, 162, 855-880.

[8] Bingham, N.H., Goldie, C.M., & Teugels, J.L. (1989). *Regular Variation*. Cambridge University Press.
[9] Choi, B.D., & Chang, Y. (1999). Single server retrial queues with priority calls. *Mathematical and Computer Modelling*, 30, 7-32.

[10] Choi, B.D., & Park, K.K. (1990). The M/G/1 retrial queue with Bernoulli schedule. *Queueing Systems*, 7, 219–228.

[11] Embrechts, P., Kluppelberg, C., & Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Heidelberg.

[12] Falin, G.I. (1990). A survey of retrial queues. *Queueing Systems*, 7(2), 127–168.

[13] Falin, G.I., Artalejo, J.R., & Martin, M. (1993). On the single server retrial queue with priority customers. *Queueing Systems*, 14(3–4), 439-455.

[14] Falin, G.I., & Templeton, J.G.C. (1997). *Retrial Queues*. Chapman & Hall, London.

[15] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications, Vol. II*. John Wiley & Sons, London.

[16] Foss, S., Korshunov, D., & Zachary, S. (2011). *An Introduction to Heavy-Tailed and Subexponential Distributions*. Springer, New York.

[17] Grandell, J. (1997). *Mixed Poisson Processes*. Chapman & Hall, London.

[18] Kim, J. (2015). Tail asymptotics for the queue size distribution in an M^X/G/1 retrial queue. *Journal of Applied Mathematics & Informatics*, 33, 185–191.

[19] Kim, B., & Kim, J. (2012). Exact tail asymptotics for the M/M/m retrial queue with nonpersistent customers. *Operations Research Letters*, 40(6), 537–540.

[20] Kim, B., Kim, J., & Kim, J. (2010a). Tail asymptotics for the queue size distribution in the MAP/G/1 retrial queue. *Queueing Systems*, 66(1), 79–94.

[21] Kim, J., & Kim, B. (2016). A survey of retrial queueing systems. *Ann Oper Res*, 247(1), 3–36.

[22] Kim, J., Kim, B., & Ko, S.-S. (2007). Tail asymptotics for the queue size distribution in an M/G/1 retrial queue. *Journal of Applied Probability*, 44(4), 1111–1118.

[23] Kim, J., Kim, J., & Kim, B. (2010c). Regularly varying tail of the waiting time distribution in M/G/1 retrial queue. *Queueing Systems*, 65(4), 365–383.

[24] Kim, J., Kim, J., & Kim, B. (2012). Tail asymptotics of the queue size distribution in the M/M/m retrial queue. *Journal of Computational and Applied Mathematics*, 236(14), 3445–3460.

[25] Kulkarni, V.G., & Liang, H.M. (1997). Retrial queues revisited. In J.H. Dshalalow (Ed.), *Frontiers in Queueing: Models and Applications in Science and Engineering*, 19–34. CRC Press, Boca Raton.

[26] de Meyer, A., & Teugels, J.L. (1980). On the asymptotic behavior of the distributions of the busy period and service time in M/G/1. *Journal of Applied Probability*, 17, 802–813.
[27] Li, H., & Yang, T. (1998). Geo/G/1 discrete time retrial queue with Bernoulli schedule. *European Journal of Operational Research*, 111, 629–649.

[28] Liu, B., Min, J., & Zhao, Y.Q. (2017). Refined tail asymptotic properties for the $M^X/G/1$ retrial queue. Submitted. [arXiv:1801.02525](https://arxiv.org/abs/1801.02525).

[29] Liu, B., & Zhao, Y.Q. (2017). Second order asymptotic properties for the tail probability of the number of customers in the $M/G/1$ retrial queue. Submitted. [arXiv:1801.09607](https://arxiv.org/abs/1801.09607).

[30] Liu, B., Wang, J., & Zhao, Y.Q. (2014). Tail asymptotics of the waiting time and the busy period for the $M/G/1/K$ queues with subexponential service times. *Queueing Systems*, 76(1), 1–19.

[31] Liu, B., Wang, X., & Zhao, Y.Q. (2012). Tail asymptotics for $M/M/c$ retrial queues with nonpersistent customers. *Operational Research: An International Journal*, 12(2), 173–188.

[32] Liu, B., & Zhao, Y. Q. (2010). Analyzing retrial queues by censoring. *Queueing Systems*, 64(3), 203–225.

[33] Masuyama, H. (2014). Subexponential tail equivalence of the stationary queue length distributions of $BMAP/GI/1$ queues with and without retrials. [arXiv:1310.4608](https://arxiv.org/abs/1310.4608v3).

[34] Shang, W., Liu, L., & Li, Q.-L. (2006). Tail asymptotics for the queue length in an $M/G/1$ retrial queue. *Queueing Systems*, 52(3), 193–198.

[35] Yamamuro, K. (2012). The queue length in an $M/G/1$ batch arrival retrial queue. *Queueing Systems*, 70(2), 187–205.

[36] Yang, T., & Templeton, J.G.C. (1987). A survey on retrial queues. *Queueing Systems*, 2(3), 201–233.