CrossMark

Customer Sojourn Time in $GI/GI/1$ Feedback Queue in the Presence of Heavy Tails

Sergey Foss[1,2] · Masakiyo Miyazawa[3]

Abstract We consider a single-server $GI/GI/1$ queueing system with feedback. We assume the service time distribution to be (intermediate) regularly varying. We find the tail asymptotics for a customer's sojourn time in two cases: the customer arrives in an empty system, and the customer arrives in the system in the stationary regime. In particular, in the case of Poisson input we obtain more explicit formulae than those in the general case. As auxiliary results, we find the tail asymptotics for the busy period distribution in a single-server queue with an intermediate varying service times distribution and establish the principle-of-a-single-big-jump equivalences that characterise the asymptotics.

1 Introduction

In queueing theory, the sojourn time U of a customer in a queueing system is one of important characteristics, this is the time from its arrival instant to departure instant. In general, the distribution of U is hard to find analytically, and research interest is directed to the asymptotics of the tail probability, $\mathbb{P}(U > x)$, as $x \to \infty$ under various stochastic assumptions. Among them, the following assumption is typically used.

✉ Sergey Foss
  s.foss@hw.ac.uk

[1] Heriot-Watt University, Edinburgh, Scotland, UK

[2] Novosibirsk State University, Novosibirsk, Russia

[3] Tokyo University of Science, Noda, Chiba, Japan

(1) Delaying arrivals leads to increase of the characteristic.

The assumption (1) is known as a *monotonicity* property, it plays an important role in the asymptotic analysis of various characteristics. Another important factor for the tail asymptotic problem is the heaviness of service time distributions. A nonnegative random variable X is said to have a heavy tail distribution if $\mathbb{E} \exp(sX) = \infty$ for all $s > 0$, and a light tail distribution, otherwise. If service time distributions have heavy tails (or light tails), we talk about a heavy (or light) tail regime of the system.

Under the heavy tail regime, the sojourn time problem is relatively well studied for the system satisfying the monotonicity assumption (1) (see e.g. [2,4,5,11] and references therein). However, if a customer separately takes multiple services, then the monotonicity is violated, in general. This is a common phenomenon in queueing networks. Clearly, *non-monotone* characteristics are also important, but we are unaware of any asymptotic results for them in the presence of heavy tails. So we challenge the tail asymptotics problem of a non-monotone characteristic under the heavy tail regime, using a relatively simple model as one of the first attempts.

We consider the following single server system. Exogenous customers arrive at the system subject to *i.i.d.* inter-arrival times $\{t_n\}$ with finite mean a, join at the end of a queue, and are served in the first-in first-out order by a single server with *i.i.d.* service times $\{\sigma_n\}$. When a customer completes service, it returns to the end of the queue for another service, with probability $p \in (0, 1)$, or leaves the system, with probability $q = 1 - p$. Both transitions are independent of everything else. Note that if a customer requires more than one service, its sojourn time is the sum of several periods of waiting in queue with following service. One can see that the sojourn time U does not have natural monotonicity properties with respect to inter-arrival times. We refer this system as a $GI/GI/1$ feedback queue. If $p = 0$ (no customer returns), then it is reduced to a standard $GI/GI/1$ queue (e.g., see [1]).

Because of independent feedback of customers completing service, the ith arriving customer, customer i for short, requires a geometric number of services, say K_i, and their total duration is $\sum_{j=1}^{K_i} \sigma_i^{(j)}$ where $\{\sigma_i^{(j)}\}$ are *i.i.d.* with finite mean b and do not depend on K_i. Throughout the paper, we assume that the system is stable, that is, $\rho \equiv \lambda b/q < 1$, where $\lambda = 1/a$, and the service-time distribution is heavy-tailed. More precisely, we assume the distribution to be *intermediate regularly varying*—see the list of heavy-tailed distributions at the end of this section and in Appendix A.

For this $GI/GI/1$ feedback queue, we derive the asymptotics of the probability $\mathbb{P}(U > x)$ for the sojourn time U of customer 1 as $x \to \infty$ under the heavy tail regime. This U depends on the state of the system just before the arrival instant of customer 1. We consider two scenarios for the system state found by customer 1. The first one is that customer 1 enters an empty queue. It is the most difficult because coefficients appearing in the asymptotics are sensitive not only to the first moments, but to the whole distribution of inter-arrival and service times, in general. The second one is that customer 1 enters a stationary queue. In both cases, we also obtain simpler formulae in the case when the input is Poisson. In particular, the first moment of the sojourn time can be obtained in a closed form in this special case. A part of this information will be used for the general renewal arrival case.

Our model may be considered as a particular case of a two-server *generalised Jackson network* where server $i = 1, 2$ has service times $\{\sigma_n^{(i)}\}$. Each family of service times is *i.i.d.* and they are mutually independent. Customers arrive in a renewal input to server 1 and join the queue there. After service completion at server 1, a customer either leaves the network, with probability p_{10}, or joins the queue to the second server, with probability $p_{12} = 1 - p_{10}$. Similarly, after service completion at server 2, a customer either leaves the network, with

probability p_{20}, or joins the queue to server 1, with probability $p_{21} = 1 - p_{20}$. Customers are server in the order of their (external and internal) arrival to the servers.

If we let $\sigma_n^{(2)} \equiv 0$ and $p_{12}p_{21} = p$, we obtain our model as a particular case indeed. So the study of our model is not only of interest itself, but also opens a window to analysing a broad class of more general models.

In the $GI/GI/1$ feedback queue, one may change the service order in such a way that each customer continuously gets service without interruption when it completes service and returns to the queue. Then such a system is nothing else than the standard $GI/GI/1$ queue with "new" $i.i.d.$ service times $\sum_{j=1}^{K_i} \sigma_i^{(j)}$, and the sojourn time is again the sum of the waiting time and of the (new) service time. Although the busy period, which is the time from the moment when the system becomes non-empty to the moment when it is again empty, is unchanged by this modification, the sojourn time does change. We will use this modified system to study the tail asymptotic of the busy period.

Thus, our analysis is connected to the standard $GI/GI/1$ queue. In this case, there is no feedback, and the monotonicity is satisfied. Hence, the waiting time of a tagged customer is a key characteristic because the sojourn time U is the sum of the waiting and service times, which are independent. In particular, the stationary waiting time is a major target for the tail asymptotic analysis. Let u_0 be the unfinished work found by the "initial" customer 1 that arrives at the system at time 0, and let W_n be the waiting times of the nth arriving customer. Then, $W_1 = u_0$ and we have the *Lindley recursion*:

$$W_{n+1} = \max(0, W_n + \sigma_n - t_n), \qquad n \geq 1. \tag{1.1}$$

We assume both inter-arrival and service times to have finite means, $a = \mathbb{E}t_n$ and $b = \mathbb{E}\sigma_n$. Here W_n forms a Markov chain which is *stable* (i.e. converges in distribution to the limiting/stationary random variable $W = W_\infty$) if the *traffic intensity* $\rho := b/a$ is less than 1. It is well-known (see e.g. [1]) that if $u_0 = 0$, then W coincides in distribution with the supremum $M = \sup_{n \geq 0} \sum_{i=1}^{n}(\sigma_n - t_n)$ of a random walk with increments $\sigma_n - t_n$.

The tail asymptotics for $\mathbb{P}(M > x)$ as $x \to \infty$ is known under the *light-tail* and *heavy-tail* regimes. In the case of light tails, there are three types of the tail asymptotics, depending on properties of the moment generating function $\varphi(s) = \mathbb{E}\exp(s\sigma)$ – see e.g. [6] and references therein. In the case of heavy tails, the tail asymptotics are known in the class of so-called *subexponential distributions* and are based on the *principle of a single big jump (PSBJ)*: M takes a large value if one of the service times is large. This PSBJ has been used for the asymptotic analysis in several other (relatively simple) stable queueing models, for a number of characteristics (waiting time, sojourn time, queue length, busy cycle/period, maximal data, etc.) that possess the monotonicity property (1) (see e.g. [2]).

Our proofs rely on the tail asymptotics for the first and stationary busy periods of the system. We establish the PSBJ for the busy period first. This allows us to establish the principle for the sojourn time since the tail distribution asymptotics of the busy period is of the same order with that of the sojourn time. Then insensitivity properties of the *intermediate varying distributions* (see Appendix A again) allow us to compute the exact tail asymptotics for the sojourn time. The main result from [7] is a key tool in our analysis.

The paper is organised as follows. Section 2 formally introduces the model and presents main results. Section 3 states the tail asymptotic of the busy period and the PSBJ. All theorems from Sects. 2 and 3 are proved in Sect. 4. The Appendix consists of three parts. Part A contains an overview on basic properties of heavy-tailed distributions and part B the proof of Corollary 3.1. In part C, we propose an alternative approach to the proof of Corollary 3.2.

Throughout the paper, we use the following notation: $1(\cdot)$ is the indicator function of the event "$\cdot$". For two positive functions f and g, we write $f(x) \sim g(x)$ if $f(x)/g(x) \to 1$ as $x \to \infty$, $f(x) \gtrsim g(x)$ if $\liminf_{x\to\infty} f(x)/g(x) \ge 1$ and $f(x) \lesssim g(x)$ if $\limsup_{x\to\infty} f(x)/g(x) \le 1$. For a distribution function F, its tail $\overline{F}$ is defined as

$$\overline{F}(x) = 1 - F(x).$$

For random variables X, Y with distributions F, G, respectively, $X =_{st} Y$ if $F = G$, and $X \le_{st} Y$ if $\overline{F}(x) \le \overline{G}(x)$ for all real x. Two families of events A_x and B_x of non-zero probabilities are equivalent, $A_x \simeq B_x$, if

$$\mathbb{P}(A_x \Delta B_x) = o(\mathbb{P}(A_x)), \quad \text{as} \quad x \to \infty, \tag{1.2}$$

where $A_x \Delta B_x = (A_x \setminus B_x) \cup (B_x \setminus A_x)$ is the symmetric difference of A_x and B_x. Note that equivalence $A_x \simeq B_x$ is symmetric, since

$$|\mathbb{P}(B_x) - \mathbb{P}(A_x)| \le \mathbb{P}(A_x \Delta B_x).$$

Note also that $A_x \simeq B_x$ is stronger than equivalence $\mathbb{P}(A_x) \sim \mathbb{P}(B_x)$.

We complete the Introduction by a short

Summary of main classes of heavy tail distributions

In this paper, we are concerned with several classes of heavy tail distributions. We list their definitions below. Their basic properties are discussed in Appendix A.

In all definitions below, we assume that $\overline{F}(x) > 0$ for all x.

1. Distribution F on the real line belongs to the class $\mathcal{L}$ of *long-tailed* distributions if, for some $y > 0$ and as $x \to \infty$,

$$\frac{\overline{F}(x + y)}{\overline{F}(x)} \to 1 \tag{1.3}$$

   (we may write equivalently $\overline{F}(x + y) \sim \overline{F}(x)$).
2. Distribution F on the positive half-line belongs to the class $\mathcal{S}$ of *subexponential* distributions if

$$\int_0^x F(dt)\overline{F}(x - t) \sim \overline{F}(x) \quad \text{as} \quad x \to \infty,$$

   which is equivalent to $\int_x^\infty F(dt)F(x - t) \sim 2\overline{F}(x)$. Distribution F of a real-valued random variable ξ is subexponential if distribution $F^+(x) = F(x)1(x \ge 0)$ of random variable $\xi^+ = \max(0, \xi)$ is subexponential.
3. Distribution F on the real line belongs to the class $\mathcal{S}^*$ of *strong subexponential* distributions if $m^+(F) \equiv \int_0^\infty \overline{F}(x)dx$ is finite and

$$\int_0^x \overline{F}(y)\overline{F}(x - y)dy \sim 2m^+(F)\overline{F}(x) \quad \text{as} \quad x \to \infty.$$

4. Distribution F on the real line belongs to the class $\mathcal{D}$ of *dominantly varying* distributions if there exists $\alpha > 1$ (or, equivalently, for all $\alpha > 1$) such that

$$\liminf_{x\to\infty} \frac{\overline{F}(\alpha x)}{\overline{F}(x)} > 0 \tag{1.4}$$

5. Distribution F on the real line belongs to the class $\mathcal{IRV}$ of *intermediate regularly varying* distributions if

$$\lim_{\alpha\downarrow 1} \liminf_{x\to\infty} \frac{\overline{F}(\alpha x)}{\overline{F}(x)} = 1. \tag{1.5}$$

6. Distribution F on the real line belongs to the class $\mathcal{RV}$ of *regularly varying* distributions if, for some $\beta > 0$,

$$\overline{F}(x) = x^{-\beta} L(x), \tag{1.6}$$

where $L(x)$ is a *slowly varying* function, i.e. $L(cx) \sim L(x)$ as $x \to \infty$, for any $c > 0$.

The following relations between the classes introduced above may be found, say, in the books [3] or [9]: in the class of distributions F with finite $m^+(F)$,

$$\mathcal{RV} \subset \mathcal{IRV} \subset \mathcal{L} \cap \mathcal{D} \subset \mathcal{S}^* \subset \mathcal{S} \subset \mathcal{L}. \tag{1.7}$$

2 The Modelling Assumptions and Main Results

In this section, we describe the dynamics of the sojourn time of a tagged customer (customer 1), and present main results on the tail asymptotics of its sojourn time. In Sect. 2.1, we formally introduce the $GI/GI/1$ feedback queue and then, in Sect. 2.2, a particular $M/GI/1$ feedback queue, which has Poisson arrivals. Then the main results are presented in Sect. 2.3.

2.1 $GI/GI/1$ Feedback Queue

Let K be the number of services of the tagged customer until its departure. By the feedback assumption, K is geometrically distributed with parameter p, that is,

$$\mathbb{P}(K = k) = qp^{k-1}, \qquad k = 1, 2, \ldots, \tag{2.1}$$

and independent of everything else. Throughout the paper, we make the following assumptions:

 (i) The exogenous arrival process is a renewal process with a finite mean interarrival time $a > 0$.
 (ii) All the service times that start after time 0 are *i.i.d.* with finite mean $b > 0$, they are jointly independent of the arrival process.
(iii) The system is stable, that is,

$$\rho \equiv \lambda b/q < 1, \tag{2.2}$$

where $\lambda = 1/a$.

We denote the counting process of the exogenous arrivals by $N^e(\cdot) \equiv \{N^e(t); t \geq 0\}$. We use the notation G for the service time distribution, and use σ for a random variable subject to G.

Let (X_0, R_0^s) be the pair of the number of earlier customers and the remaining service time of a customer being served at time 0, where $R_0^s = 0$ if there is no customer in the system. Let u_0 be the waiting time of the tagged customer before the start of its first service, Then

$$u_0 = R_0^s + \sum_{i=1}^{X_0-1} \sigma_{0,i+1}, \tag{2.3}$$

where $\sigma_{0,i}$'s for $i \geq 2$ are *i.i.d.* random variables each of which has the same distribution as σ. There are two typical scenarios for the initial distribution, that is, the distribution of (X_0, R_0^s).

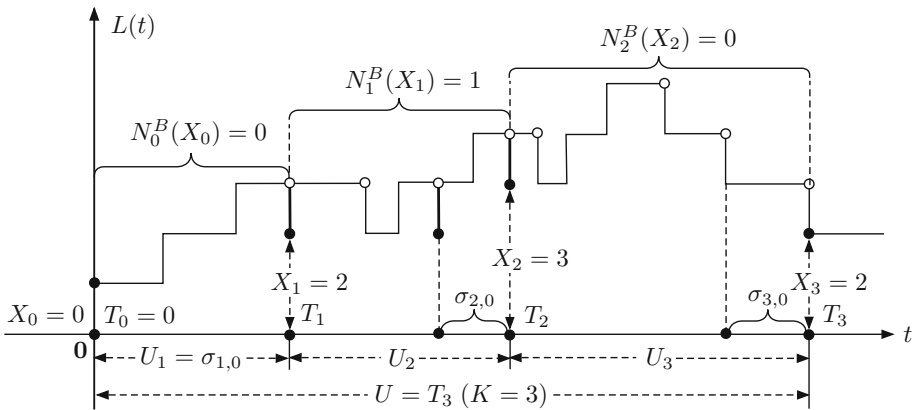(2a) A tagged arriving customer finds the system empty. That is, $(X_0, R_0^s) = (0, 0)$.

Fig. 1 Sample path of the queue length process $L(t)$ and (X_k, U_k)

(2b) A tagged arriving customer finds X_0 customers and the remaining service time R_0^s of the customer being served. Thus, the initial state $(X_0, R_0^s) \neq (0, 0)$.

In this paper, we assume that the service time distribution is heavy tailed, and mainly consider the tail asymptotic of the sojourn time distribution of the $GI/GI/1$ feedback queue under the scenario (2a). The case (2b) when X_0 and R_0^s are bounded by a constant may be studied very similarly to the case (2a), therefore we do not analyse it. We consider the case (2b) when (X_0, R_0^s) is subject to the stationary distribution embedded at the arrival instants.

For given (X_0, R_0^s), we have defined u_0. Let X_k be the queue length behind the tagged customer when it finished its kth service for $k \geq 1$ when the tagged customer gets service at least k times. Similarly, let U_k be the sojourn time of the tagged customer measured from its $(k-1)$th service completion to its kth service completion, and let T_k be the sojourn time of the tagged customer just after its kth service completion.

We now formally define random variables X_k, U_k and T_k by induction. Let $T_0 = 0$. Denote the kth service time of the tagged customer by $\sigma_{k,0}$, while $\sigma_{k,0}$, $i = 1, \ldots, X_{k-1}$ are the service times of the customers waiting before the tagged one on its kth return. Note that $\sigma_{k,i}$'s for $k \geq 1$, $i \geq 0$ are $i.i.d.$ random variables subject to the same distribution as σ. Then, X_k, U_k and T_k for $k \geq 1$ are defined as

$$U_k = \begin{cases} \sigma_{1,0} + u_0 & k = 1, \\ \sum_{i=0}^{X_{k-1}} \sigma_{k,i} & k \geq 2, \end{cases} \tag{2.4}$$

$$T_k = T_{k-1} + U_k, \tag{2.5}$$

$$X_k = N^e(T_k) - N^e(T_{k-1}) + N_k^B(X_{k-1}), \tag{2.6}$$

where u_0 is given by (2.3), and $N_k^B(n)$'s are $i.i.d.$ random variables each of which is subject to the Binomial distribution with parameters n, p. The dynamics of the sojourn time is depicted above when $X_0 = 0$, that is, a tagged customer finds the system empty (Fig. 1).

To make clear the dependence of X_k, U_k, T_k, we introduce a filtration $\{\mathcal{F}_t; t \geq 0\}$ as

$$\mathcal{F}_t = \sigma\text{-field generated by } \{X_0, R_0^s, (N^e(u), N^s(u), N^r(u)), u \leq t\},$$

where $N^s(t)$ and $N^r(t)$ are the numbers of customers who completed service and who return to the queue, respectively, up to time t. Clearly, T_k is a $\mathcal{F}_t$-stopping time, and X_k and U_k are

$\mathcal{F}_{T_k}$-measurable. Furthermore, $\sigma_{k,0}$ and $\sigma_{k,i}$ for $i \geq 1$ are independent of $\mathcal{F}_{T_{k-1}}$. Then U, the sojourn time of the tagged customer, may be represented as

$$U = u_0 + \sum_{k=1}^{K} U_k = u_0 + \sum_{k=1}^{K} \sum_{i=0}^{X_{k-1}} \sigma_{k,i}. \tag{2.7}$$

For $k \geq 0$, let $Y_k = \sum_{\ell=0}^{k} X_\ell$ for $k \geq 0$, which is the total number of external and internal arrivals to the queue up to time T_k plus the number of customers in system at time 0. Then

$$Y_k = X_0 + N^e(T_k) + \sum_{\ell=1}^{k} N_\ell^B\left(X_{\ell-1}\right) =_{st} X_0 + N^e(T_k) + N_k^B\left(Y_{k-1}\right). \tag{2.8}$$

Hence, under scenario (2a), we have $u_0 = X_0 = 0$, so

$$U = \sum_{k=1}^{K} \sum_{i=0}^{X_{k-1}} \sigma_{ki} =_{st} \sum_{i=1}^{K+Y_{K-1}} \sigma_i, \tag{2.9}$$

while, under scenario (2b),

$$U =_{st} u_0 + \sum_{i=1}^{K+Y_{K-1}} \sigma_i, \tag{2.10}$$

where σ_i's are $i.i.d.$ random variables each of which has the same distribution as σ. Note that $K + Y_{K-1}$ is $\mathcal{F}_{T_{K-1}}$-measurable that depends, in general, on all σ_i's of customers who arrive before T_{K-1}. This causes considerable difficulty in the asymptotic analysis of U.

Thus, we need to consider dependence structure in the representation of U. Furthermore, $\{(U_k, X_k); k \geq 0\}$ is generally not a Markov chain for a general renewal process.

On the other hand, if the arrival process $N^e(\cdot)$ is Poisson, then not only $\{(U_k, X_k); k \geq 0\}$ but also $\{X_k; k \geq 0\}$ is a Markov chain with respect to the filtration $\{\mathcal{F}_{T_k}; k \geq 0\}$. In this case, we may obtain exact expressions for $\mathbb{E}X_k$ and then an explicit form for the tail asymptotics.

2.2 $M/GI/1$ Feedback Queue and Branching Process

In this subsection, we assume that the exogenous arrival process is Poisson with rate $\lambda > 0$. This model is analytically studied using Laplace transforms in [12], but no asymptotic results are given there. Note that we may consider $\{X_k; k \geq 0\}$ as a branching process and directly compute $\mathbb{E}(X_k)$, which then will be used for the general renewal input case.

Since the Poisson process $N^e(\cdot)$ has independent increments, (2.6) is simplified to

$$X_k = N_k^e(U_k) + N_k^B(X_{k-1}), \qquad 1 \leq k \leq K, \tag{2.11}$$

using independent Poisson processes N_k^e and independent Binomial random variables $N_k^B(n)$. Furthermore, (2.11) can be written as

$$X_k = \begin{cases} N_{1,1}^e(\sigma_{1,0} + u_0) + N_1^B(X_0), & k = 1, \\ N_{k,0}^e(\sigma_{k,0}) + \sum_{i=1}^{X_{k-1}}(N_{k,i}^e(\sigma_{k,i}) + N_{k,i}^B(1)), & k \geq 2, \end{cases} \tag{2.12}$$

where $N_{k,i}^e(\cdot)$'s are independent Poisson processes with rate λ. Hence, $\{X_k; k \geq 1\}$ is a branching process with immigration.

Due to the branching structure, we can compute the moments of X_k explicitly. We are particularly interested in their means. From (2.12), we have

$$\mathbb{E}(X_k) = \begin{cases} \lambda(b + \mathbb{E}(u_0)) + p\mathbb{E}(X_0), & k = 1, \\ \lambda b + \mathbb{E}(X_{k-1})r, & k \geq 2, \end{cases}$$

where $r = \lambda b + p$. By the stability condition (2.2), $r < 1$, and we have

$$\mathbb{E}(X_k) = \frac{1 - r^{k-1}}{1 - r}\lambda b + \mathbb{E}(X_1)r^{k-1}, \qquad k \geq 1. \tag{2.13}$$

Hence, we have a uniform bound:

$$\mathbb{E}(X_k) \leq \frac{1}{1 - r}\lambda b + \mathbb{E}(X_1), \qquad k \geq 1. \tag{2.14}$$

Furthermore, we have

$$\mathbb{E}(K + Y_{K-1}) = \mathbb{E}(K) + \sum_{k=1}^{\infty}\sum_{\ell=0}^{k-1} \mathbb{E}(X_\ell)\mathbb{P}(K = k)$$

$$= \mathbb{E}(K) + \mathbb{E}(X_0) + \sum_{\ell=1}^{\infty} \mathbb{E}(X_\ell)\mathbb{P}(K \geq \ell + 1)$$

$$= \frac{1}{q} + \mathbb{E}(X_0) + \frac{\lambda bp}{(1 - r)q} + \frac{(1 - r)\mathbb{E}(X_1) - \lambda b}{(1 - r)(1 - pr)}p. \tag{2.15}$$

Under the scenario (2a), $\mathbb{E}(X_k)$ of the $M/GI/1$ feedback queue will be used for the tail asymptotic of the sojourn time in the $GI/GI/1$ feedback queue. Thus, we introduce notations for them. Let $X_k^{(0)}(M/GI/1)$ be the X_k of the $M/GI/1$ feedback queue for $X_0 = 0$, then define $m_k^{(0)}$ as

$$m_0^{(0)} = 0, \qquad m_k^{(0)} = \mathbb{E}(X_k^{(0)}(M/GI/1)), \quad k \geq 1.$$

From (2.13), we have

$$m_k^{(0)} = \frac{1 - r^k}{1 - r}\lambda b = (1 + r + \cdots + r^{k-1})\lambda b. \tag{2.16}$$

We will use $m_{K-1}^{(0)}$ and $m_K^{(0)}$ in main results the next section. It should be noticed that they are random variables obtained by substituting $K - 1$ and K into k of $m_k^{(0)}$.

2.3 Main Results

We are ready to present the main results of this paper. They are proved in Sect. 4.

Theorem 2.1 *For the stable $GI/GI/1$ feedback queue, assume that its service time distribution is intermediate regularly varying ($\mathcal{IRV}$). If the tagged customer finds the system empty, then*

$$\mathbb{P}(U > x) \sim \frac{1}{q}\mathbb{E}\left(1 + X_{K-1}^{(0)}\right)\mathbb{P}\left((1 + m_{K-1}^{(0)})\sigma > x\right) \quad as\ x \to \infty, \tag{2.17}$$

where $X_k^{(0)}$ is X_k for $X_0 = 0$, $m_k^{(0)} = \frac{1-r^k}{1-r}\lambda b$ for $r = p + \lambda b$ by (2.16). Here the random variable K does not depend on $\{X_k^{(0)}\}$ and σ.

Remark 2.1 In the case of a Poisson input, $\mathbb{E}\big(1 + X_{K-1}^{(0)}\big) = \frac{q(1+p)}{1-pr}$. In general, it is hard to evaluate $\mathbb{E}\big(1 + X_{K-1}^{(0)}\big)$ because this requires computing $\mathbb{E}(N^e(T_k))$.

We prove this theorem in Sect. 4.3 using the PSBJ that is established in Theorem 3.2.

Corollary 2.1 *Under the assumptions of Theorem 2.1, for each $k \geq 1$,*

$$\mathbb{P}\left(T_k > x\right) \sim \sum_{\ell=0}^{k-1} \mathbb{E}\left(1 + X_{k-\ell-1}^{(0)}\right) \mathbb{P}\left(\big(1 + m_\ell^{(0)}\big)\sigma > x\right) \quad as \; x \to \infty. \tag{2.18}$$

This corollary is easily obtained from arguments used in the proof of Theorem 2.1. On the other hand, if we take the geometrically weighted sum of (2.18) and if the interchange of this sum and the asymptotic limit are allowed, then we have (2.17). This interchange of the limits is legitimated by Theorem 3.2. However, Corollary 2.1 itself can be directly proved. We provide such a proof for a slightly extended version of Corollary 2.1 in Appendix C.

Corollary 2.2 *Assume that the assumptions of Theorem 2.1 hold.*

(a) *If the distribution of service times is $\mathcal{RV}$, $\overline{G}(x) = L(x)/x^{\alpha+1}$ with $\alpha > 0$ and slowly varying function $L(x)$, then*

$$\mathbb{P}\left(\big(1 + m_{K-1}^{(0)}\big)\sigma > x\right) \sim \frac{qL(x)}{x^{\alpha+1}} \sum_{k=0}^{\infty} p^k \left(\frac{1 - (p + r^k \lambda b)}{1 - r}\right)^{\alpha+1},$$

where $r = p + \lambda b$.

(b) *Under the assumption in (a), if the input stream is Poisson with parameter λ, then*

$$\mathbb{P}(U > x) \sim CL(x)x^{-(\alpha+1)}$$

where

$$C = \frac{1+p}{(1-r)^{\alpha+1}(1-rp)} 1 + q \sum_{k=0}^{\infty} p^k (1 - (p + r^k \lambda b))^{\alpha+1}.$$

We next present the tail asymptotic for a tagged customer that arrives in the stationary system. By "stationary" we mean stationary in discrete time, i.e. at embedded arrival epochs, this is detailed in Sect. 4.4.

Theorem 2.2 *Let U^0 be the sojourn time of a typical customer in the stationary $GI/GI/1$ feedback queue with $\mathcal{IRV}$ distribution G of service times with mean b, i.i.d. inter-arrival times with mean a and probability of feedback $p = 1 - q \in (0, 1)$. Let σ_I be a random variable having the distribution function $G_I(x) \equiv 1 - \min\left(1, \int_x^\infty \overline{G}(u)du\right)$.*

(a) *Then, as $x \to \infty$,*

$$\mathbb{P}(U^0 > x) \sim \frac{\lambda}{1 - \lambda b} \left(\mathbb{P}\left(m_K^{(0)}\sigma_I > x\right) + \frac{(1-q)\rho}{1-\rho}\mathbb{P}\left(\big(\lambda b\, m_K^{(0)}\big)\sigma_I > x\right)\right), \tag{2.19}$$

where $m_k^{(0)}$ is defined by (2.16), and σ_I is independent of K.

(b) *In particular, if G is an $\mathcal{RV}$ distribution, $\overline{G}(x) = L(x)/x^{\alpha+1}$ with $\alpha > 0$, where $L(x)$ is a slowly varying function, then*

$$\mathbb{P}(U^0 > x) \sim \frac{L(x)}{x^\alpha} \cdot \frac{q(1-r)^{-\alpha}}{\alpha(a-b)} \left(1 + \frac{b(1-q)}{aq-b}\left(\frac{b}{a}\right)^\alpha\right) \sum_{k=1}^{\infty} p^{k-1}(1 - r^k)^\alpha. \tag{2.20}$$

Remark 2.2 (a) The tail function $\overline{G_I}(x) \equiv 1 - G_I(x)$ is called the "integrated tail" of G. Instead of G_I, we may use the stationary excess distribution $G_e(x) \equiv \frac{1}{b} \int_0^x \overline{G}(u) du$ since $\overline{G_I}(x) = b\,\overline{G_e}(x)$ for sufficiently large x. In this case, let σ_e be a random variable subject to G_e, then we can replace σ_I by σ_e in (2.19), multiplying its right-hand side by b.

(b) It is notable that no $X_{K-1}^{(0)}$ for the renewal arrivals is involved in (2.19). This is different from (2.17), and may come from averaging in the steady state.

(c) It may be interesting to compare the asymptotics in (2.19) with those without feedback, which is well known (e.g., see [2]). Namely, let the stationary sojourn time $\widetilde{U}^0$ in the standard $GI/GI/1$ queue with inter-arrival times $\{t_n\}$ and with service times $\{\sigma_n^H\}$. where σ_n^H has the same distribution as $\sum_{i=1}^K \sigma_i$. If σ_I has a subexponential distribution, then

$$\mathbb{P}(\widetilde{U}^0 > x) \sim \frac{\lambda}{1-\rho} \int_x^\infty \mathbb{P}\left(\sum_{i=1}^K \sigma_i > u\right) du \sim \frac{\lambda}{q(1-\rho)} \mathbb{P}(\sigma_I > x), \qquad (2.21)$$

where the second asymptotic equivalence follows from Lhópital's theorem and (A.4). Thus, one can see that (2.19) is asymptotically compatible with (2.21) as $q \uparrow 1$, because $\rho \to \lambda b$ and $m_K^{(0)} \to 1$ almost surely as $q \uparrow 1$.

3 Busy Period and the Principle of a Single Big Jump

In this section, we present the Principle of a Single Big Jump (PSJB) in Theorem 3.2 below, which will be used for a proof Theorem 2.1. For that, we first provide an auxiliary result on the tail asymptotics of the busy period in the $GI/GI/1$ queue without feedback. Denote its service time distribution by H and let σ_i^H be the ith service time. It is assumed that the arrivals are subject to the renewal process N^e with interarrival times t_i with mean a, and H has a finite and positive mean $b_H > 0$. Denote the traffic intensity by $\rho \equiv b_H/a < 1$. Let B be the (duration of the) first busy period in this $GI/GI/1$ queue, which is the time from the instant when the system becomes non-empty to the instant when it again becomes empty. We here omit the subscript $_H$ for ρ, B, because they will be unchanged for the $GI/GI/1$ feedback queue. We finally let τ^H be the number of customers served in the first busy period. We let $\xi_i^H = \sigma_i^H - t_i$, and let $S_n^H = \sum_1^n \xi_i^H$. Then $\tau^H = \min\{n \geq 1 : S_n^H \leq 0\}$. Recall the definitions of classes of heavy-tailed distributions $\mathcal{L}, \mathcal{S}^*, \mathcal{IRV}$ and $\mathcal{RV}$ at the end of Sect. 1. The following theorem is proved in Sect. 4.1.

Theorem 3.1 *Consider a stable $GI/GI/1$ queue, $\rho < 1$, with the service time distribution H.*

If $H \in \mathcal{L}$, then

$$\liminf_{x\to\infty} \frac{\mathbb{P}(B > x)}{\mathbb{E}\tau^H \overline{H}(x(1-\rho))} \geq 1 \quad and \quad \liminf_{x\to\infty} \frac{\mathbb{P}(\tau^H > x)}{\mathbb{E}\tau^H \overline{H}(x(a-b_H))} \geq 1. \qquad (3.1)$$

If, in addition, $H \in \mathcal{S}^*$, then, for any $0 < c < 1$,

$$\limsup_{x\to\infty} \frac{\mathbb{P}(B > x)}{\mathbb{E}\tau^H \overline{H}(cx(1-\rho))} \leq 1 \quad and \quad \limsup_{x\to\infty} \frac{\mathbb{P}(\tau^H > x)}{\mathbb{E}\tau^H \overline{H}(cx(a-b_H))} \leq 1. \qquad (3.2)$$

Finally, if $H \in \mathcal{IRV}$, then, as $x \to \infty$,

$$\mathbb{P}(B > x) \sim \mathbb{E}(\tau^H)\overline{H}(x(1-\rho)) \quad and \quad \mathbb{P}(\tau^H > x) \sim \mathbb{E}(\tau^H)\overline{H}(x(a-b_H)). \qquad (3.3)$$

Remark 3.1 For the class of regularly varying tails, the equivalence (3.3) was proved by Zwart in [13]. We provide a different proof which is shorter and works for a broader class of distributions. Our proof is based on probabilistic intuition related to the principle of a single big jump. A similar result holds for another class of distributions that overlaps with the $\mathcal{IRV}$ class but does not contain it, see e.g. [11].

Recall the equivalence $A_x \simeq B_x$ for two families of events A_x and B_x with variable x. We have the following corollary, which is proved in Appendix B.

Corollary 3.1 *Consider the same $GI/GI/1$ queue as in Theorem 3.1. Let σ_n^H be the service time of the nth arriving customer. If $H \in \mathcal{IRV}$, then, for the busy period B, as $x \to \infty$,*

$$\mathbb{P}(B > x) \sim \sum_{n \geq 1} \mathbb{P}\left(\tau^H \geq n\right) \mathbb{P}\left(\sigma_n^H > x(1-\rho)\right) = \mathbb{E}\tau^H \mathbb{P}\left(\sigma_1^H > x(1-\rho)\right), \quad (3.4)$$

and, for any $\varepsilon > 0$, one can choose $N = N^e(\varepsilon) \geq 1$ such that, as $x \to \infty$,

$$\mathbb{P}(B > x) \gtrsim \sum_{n=1}^{N} \mathbb{P}\left(\tau^H \geq n, \sigma_n^H \geq x(1-\rho)\right) \gtrsim (1-\varepsilon)\mathbb{P}(B > x). \quad (3.5)$$

Furthermore, the following PSBJ holds:

$$\{B > x\} \simeq \cup_{n \geq 1}\left\{\tau^H \geq n, \sigma_n^H > x(1-\rho)\right\}, \quad x \to \infty. \quad (3.6)$$

We now return to the $GI/GI/1$ feedback queue with the service time distribution G. Assume that the first customer arrives at the system at time instant $T_0 = 0$ and finds it empty. Recall that K_i is the number of services ith customer has in the system, K_i's are independent of everything else and *i.i.d* with the same geometric distribution as K [see (2.1)]. For convenience, we let $K = K_1$. Denote by $\sigma_i^{(j)}$ the jth service time of the ith customer in the $GI/GI/1$ feedback queue. Recall that $\sigma_i^{(j)}$ has the same distribution G.

Consider the $GI/GI/1$ queue without feedback and with service times σ_i^H where

$$\sigma_i^H = \sum_{j=1}^{K_i} \sigma_i^{(j)},$$

and denote its distribution by H. Since the length of the busy period, B, does not depend on the order of services, we may allow the server to proceed with services of lengths σ_i^j, like in the queue with feedback, and conclude that the (the lengths of) the busy periods are the same in both queues. Similarly, the traffic intensity ρ in the new queue without feedback coincides with that in the $GI/GI/1$ queue with feedback. Furthermore, let τ be the number of service times in the first busy period of this feedback queue. Then, $\tau = \sum_{i=1}^{\tau^H} K_i$, and therefore we have

$$\mathbb{E}(\tau) = \mathbb{E}(K)\mathbb{E}(\tau^H), \quad b_H = \mathbb{E}(K)b.$$

We now consider the $GI/GI/1$ feedback queue introduced in Sect. 2.1. We establish the PSBJ, i.e. show that, for large x, the rare event $\{U > x\}$ occurs mostly due to a big value of one of the service times. Our proof of Theorem 3.2 is based on Theorem 3.1 and is given in Sect. 4.2.

Theorem 3.2 *Consider a stable single-server queue $GI/GI/1$ with feedback. Assume that the service times distribution is intermediate regularly varying. Denote by U be the sojourn time of the first customer, and let*

$$P_{k,\ell,i,j}(x) = \mathbb{P}(U > x, K = k + \ell, X_{k-1} = j, \sigma_{k,i} > x(1 - \rho)). \tag{3.7}$$

If there exists a collection of positive functions $\{g_{k,\ell,i,j}(x)\}$ such that, as $x \to \infty$,

$$P_{k,\ell,i,j}(x) \sim g_{k,\ell,i,j}(x), \qquad \forall k \geq 1, \ell \geq 0, 0 \leq i \leq j, \tag{3.8}$$

and constants $C_{k,\ell,i,j}$ such that, for any $k \geq 1, \ell \geq 0, j \geq 0, 0 \leq i \leq j, x \geq 0$,

$$g_{k,\ell,i,j}(x) \leq C_{k,\ell,i,j} \cdot \mathbb{P}(\sigma > x), \tag{3.9}$$

$$C := \sum_{k=1}^{\infty} \sum_{\ell=0}^{\infty} \sum_{j=0}^{\infty} \sum_{i=0}^{j} C_{k,\ell,i,j} < \infty, \tag{3.10}$$

then

$$\mathbb{P}(U > x) \sim \sum_{k=1}^{\infty} \sum_{\ell=0}^{\infty} \sum_{j=0}^{\infty} \sum_{i=0}^{j} g_{k,\ell,i,j}(x). \tag{3.11}$$

4 Proofs of the Theorems

4.1 Proof of Theorem 3.1

We will prove Theorem 3.1 for the tail asymptotics of the busy period B only. The proof for τ^H, the number of arriving customers in the busy period, is similar.

It is enough to prove the lower and upper bounds in (3.1) and (3.2). Then the equivalences in (3.3) follow by letting c tend to 1 and using the property of $\mathcal{IRV}$ distributions.
Lower bound Since $\xi_1^H = \sigma_1^H - t_1$, $d_0 \equiv b^H/(a - b^H)$ is the solution to the equation

$$\mathbb{E}\sigma_1^H + d_0 \mathbb{E}\xi_1^H = 0.$$

We put $\widehat{\psi}_n^H = \sigma_n^H + d_1\xi_n^H$ for any positive number $d_1 < d_0$. Then $\{\widehat{\psi}_n^H\}$ are i.i.d. random variables with common mean $\mathbb{E}\widehat{\psi}_1^H > 0$. Recall that $\tau^H = \min\{n \geq 1 : S_n^H \leq 0\}$. For any fixed real $C > 0, R > 0$, integer $N \geq 1$, and for $x \geq 0$, define events D_i and A_i for $i \geq 1$ as

$$D_i = \left\{ \sum_{j=1}^{i-1} |\widehat{\psi}_j^H| \leq C, \tau^H \geq i, \widehat{\psi}_i^H > x + C + R \right\}, \qquad A_i = \bigcap_{\ell \geq 1} \left\{ \sum_{j=1}^{\ell} \widehat{\psi}_{i+j}^H \geq -R \right\}.$$

Then, we have

$$\mathbb{P}(B > x) \geq \mathbb{P}\left(\sum_{i=1}^{\tau^H} \widehat{\psi}_i^H > x \right) \geq \sum_{i=1}^{N} \mathbb{P}(D_i \cap A_i). \tag{4.1}$$

Here, the first inequality in (4.1) holds since $S_{\tau^H}^H$ is non-positive, and the second inequality comes from the following facts. Events D_i are disjoint and, given the event D_i, we have $\sum_{j=1}^{i} \widehat{\psi}_j^H > x + R$. Then, given the event $D_i \cap A_i$, we have $\sum_{j=1}^{k} \widehat{\psi}_j^H \geq x$ for all $k \geq i$ and, in particular, $\sum_{j=1}^{\tau^H} \widehat{\psi}_j^H > x$. Thus, (4.1) holds.

The events $\{A_i\}$ form a stationary sequence. Due to the SLLN, for any $\varepsilon > 0$, one can choose R so large that $\mathbb{P}(A_i) \geq 1 - \varepsilon$. For this ε and any $N \geq 1$, we can choose sufficiently large C such that

$$\mathbb{P}\left(\sum_{j=1}^{i-1} |\widehat{\psi}_j^H| \leq C, \tau^H \geq i\right) \geq (1-\varepsilon)\mathbb{P}(\tau^H \geq i), \quad i \leq N.$$

Hence, (4.1) implies that, as $x \to \infty$,

$$\mathbb{P}(B > x) \geq \sum_{i=1}^{N} \mathbb{P}\left(\sum_{j=1}^{i-1} |\widehat{\psi}_j^H| \leq C, \tau^H \geq i\right) \mathbb{P}(\widehat{\psi}_i^H > x + C + R)\mathbb{P}(A_i)$$

$$\geq (1-\varepsilon)^2 \mathbb{P}(\widehat{\psi}_1^H > x + C + R) \sum_{i=1}^{N} \mathbb{P}(\tau^H \geq i),$$

and therefore the long-tailedness of distribution H and (iii) of Remark A.1 yield

$$\liminf_{x \to \infty} \frac{\mathbb{P}(B > x)}{\mathbb{P}(\widehat{\psi}_1^H > x)} = \liminf_{x \to \infty} \frac{\mathbb{P}(B > x)}{\overline{H}(x/(1+d_1))} \geq (1-\varepsilon)^2 \sum_{i=1}^{N} \mathbb{P}(\tau^H \geq i).$$

Letting first N to infinity and then ε to zero completes the proof of the first inequality of (3.1).

Upper bound Take $L > 0$ and put $\tilde{t}_n = \min(t_n, L)$, $\tilde{\xi}_n^H = \sigma_n - \tilde{t}_n$, $\tilde{S}_n^H = \sum_1^n \tilde{\xi}_i^H$, $\tilde{\tau}^H = \min\{n \geq 1 : \tilde{S}_n^H \leq 0\}$. Put also $\psi_n^H = \sigma_n^H + d_2\tilde{\xi}_n^H$ where $d_2 > d_0$ is any number. Note that $\tilde{\xi}_1^H$ converges to ξ_1^H in distribution and in mean, as $L \to \infty$. We may choose L so large that both $\mathbb{E}\tilde{\xi}_1^H$ and $\mathbb{E}\psi_1^H$ are negative. Then $\tilde{\tau}^H$ is finite and $\tilde{S}_{\tilde{\tau}^H}^H \in (-L, 0]$ a.s. Further, as L grows, $\tilde{\tau}^H$ converges to τ in distribution and in mean and, for any $\varepsilon > 0$, we may choose L so large that $\mathbb{E}\tau^H \leq \mathbb{E}\tilde{\tau}^H \leq (1+\varepsilon)\mathbb{E}\tau^H$. By Remark A.1, the distribution of ψ_1^H also belongs to the class $\mathcal{S}^*$. We have

$$\mathbb{P}(B > x) \leq \mathbb{P}\left(\sum_{i=1}^{\tilde{\tau}^H} \sigma_i^H > x\right) = \mathbb{P}\left(\sum_{i=1}^{\tilde{\tau}^H} \psi_i^H > x + d_2\tilde{S}_{\tilde{\tau}^H}^H\right)$$

$$\leq \mathbb{P}\left(\sum_{i=1}^{\tilde{\tau}^H} \psi_i^H > x - d_2 L\right) \leq \mathbb{P}\left(\max_{1 \leq j \leq \tilde{\tau}^H} \sum_{i=1}^{j} \psi_i^H > x - d_2 L\right)$$

$$\sim \mathbb{E}\tilde{\tau}^H \mathbb{P}(\psi_1^H > x - d_2 L) \leq (1+\varepsilon)\mathbb{E}\tau^H \mathbb{P}(\psi_1^H > x - d_2 L),$$

where the equivalence follows from Theorem A.1. Further,

$$\mathbb{P}\left(\psi_1^H > x - d_2 L\right) \sim \mathbb{P}\left(\psi_1^H > x\right) \sim \mathbb{P}((d_2 + 1)\sigma_1^H > x) = \overline{H}(x/(d_2 + 1))$$

where the first equivalence follows from the long-tailedness of the distribution of ψ_1^H and the second from Remark A.1. Letting ε tend to zero, we have

$$\limsup_{x \to \infty} \frac{\mathbb{P}(B > x)}{\mathbb{E}\tau^H \overline{H}(x/(d_2 + 1))} \leq 1$$

for any $d_2 > d_0$. Let $c = (d_0 + 1)/(d_2 + 1) < 1$. This proves the first inequality in (3.2).

4.2 Proof of Theorem 3.2

Recall that we consider the scenario where the initial customer 1 arrives at the empty system. Clearly, $\sigma_1^H \leq U \leq B$ a.s. where σ_1^H is the total service time of customer 1, and B is the duration of the first busy period.

Equivalence relation (A.4) and Remark A.1 from the Appendix imply that, given σ has an intermediate regularly varying distribution, random variables σ_i^H have a common intermediate varying distribution too. Since any intermediate varying distribution is dominantly varying [see Property (1.7)], we get from Theorem 3.1 that

$$\limsup_{x \to \infty} \mathbb{P}(B > x)/\mathbb{P}(\sigma_1^H > x) < \infty. \tag{4.2}$$

Relations (4.2) and (A.4) lead then to the logarithmic asymptotics:

Lemma 4.1 *Assume that the distribution of the service time $\sigma_1^{(1)}$ belongs to the class $\mathcal{IRV}$. Then, as $x \to \infty$,*

$$\log \mathbb{P}(U > x) \sim \log \mathbb{P}(\sigma_1^H > x) \sim \log \mathbb{P}(\sigma_{1,1} > x). \tag{4.3}$$

Further, by Corollary 3.1, the PSBJ for B holds:

$$\mathbb{P}(B > x) \sim \mathbb{P}\left(B > x, \cup_{i=1}^{\tau^H}\{\sigma_i^H > x(1-\rho)\}\right) \sim \mathbb{P}\left(\cup_{i=1}^{\tau^H}\{\sigma_i^H > x(1-\rho)\}\right) \tag{4.4}$$

Here τ is the number of customers served within the first busy period.

Combining (4.4) and (A.4), we arrive at the following result:

Lemma 4.2 *Consider a stable single-server queue $GI/GI/1$ with feedback. Let B be the duration of the first busy period and U the sojourn time of the first customer. Assume that the service times distribution is intermediate regularly varying. Then*

$$\mathbb{P}(U > x) = \mathbb{P}(U > x, B > x) \sim \mathbb{P}\left(\{U > x\} \bigcap \bigcup_{i=1}^{\tau^H} \bigcup_{j=1}^{K_i}\{\sigma_i^{(j)} > x(1-\rho)\}\right). \tag{4.5}$$

To derive the exact asymptotics for $\mathbb{P}(U > x)$, we recall that, for $1 \leq k < K \equiv K_1$, $X_k \geq 0$ is the total number of services of other customers between the kth and the $(k+1)$st services of customer 1, and let $\sigma_{k,i}$ be the service time of the i service there, $1 \leq i \leq X_k$. Further, under the scenario (2a), $X_0 = 0$. Then let $\nu \geq 0$ be the total number of services of other customers after the departure of the first customer within the busy period, and let σ_i^* be the ith service time there, $1 \leq i \leq \nu$. Then random variables $\sigma_{k,.i}$ and σ_i^* are i.i.d. with the same distribution as σ and U is given by (2.9). From (4.5), we get,

$$\mathbb{P}(U > x) \sim \mathbb{P}\left(\{U > x\} \bigcap \left(\bigcup_{i=1}^{K_1} \bigcup_{j=1}^{X_i}\{\sigma_j^{(i)} > x(1-\rho)\} \bigcup \bigcup_{j=1}^{\nu}\{\sigma_j^* > x(1-\rho)\}\right)\right).$$

On the other hand, we have

$$\mathbb{P}\left(\{U > x\} \bigcap \bigcup_{i=1}^{\nu}\{\sigma_i^* > x(1-\rho)\}\right) = \sum_{n \geq 1} \mathbb{P}(U > x, \nu = n) \cdot \mathbb{P}\left(\bigcup_{i=1}^{n}\{\sigma_i^* > x(1-\rho)\}\right)$$

$$\leq C \sum_{n \geq 1} \mathbb{P}(U > x, \nu = n) n \overline{G}(x)$$

$$= C\mathbb{E}\left(v \cdot 1(U > x)\right)\overline{G}(x)$$
$$\leq C\mathbb{E}\left(\tau \cdot 1(U > x)\right)\overline{G}(x) = o(\overline{G}(x)),$$

where $C = \sup_x \overline{G}(x(1-\rho))/\overline{G}(x) < \infty$ is a constant, and recall that τ is the total number of services within the busy cycle. The last line follows since $\mathbb{E}\tau = \mathbb{E}\tau^H/q$ is finite. Therefore, we have

Lemma 4.3 *In the conditions of Lemma 4.2, we have*

$$\mathbb{P}(U > x) \sim \mathbb{P}\left(\{U > x\} \cap \bigcup_{k=1}^{K_1}\bigcup_{i=0}^{X_{k-1}}\{\sigma_{k,i} > x(1-\rho)\}\right). \tag{4.6}$$

Moreover, the following result holds:

Lemma 4.4 *Assume the conditions of Lemma 4.2 hold. Then, for any $\varepsilon > 0$, one can find N such that*

$$\mathbb{P}(U > x) \gtrsim \mathbb{P}(D_N(x)) \gtrsim (1 - \varepsilon)\mathbb{P}(U > x) \tag{4.7}$$

where

$$D_N(x) = \bigcup_{k=1}^{N}\bigcup_{\ell=0}^{2N}\left\{\{U > x, K_1 = k + \ell\} \cap \bigcup_{i=0}^{X_{k-1}}\{\sigma_{k,i} > x(1-\rho)\}\right\}. \tag{4.8}$$

Proof Indeed, the term in the right-hand side of (4.6) is bigger than $\mathbb{P}(D_N(x))$ and smaller than the sum $\mathbb{P}(D_N(x)) + \mathbb{P}(U > x, K_1 > N)$, where

$$\mathbb{P}(U > x, K_1 > N) \leq \mathbb{P}(B > x, K_1 > N). \tag{4.9}$$

Consider again the auxiliary $GI/GI/1$ queue with service times $\sigma_i^H = \sum_{j=1}^{K_i}\sigma_i^{(j)}$ and the first-come-first-served service discipline. Consider the following majorant: assume that at the beginning of the first cycle, in addition to customer 1, an extra $K - 1$ new customers arrive, so there are K arrivals in total. Here K is a geometric random variable with parameter p that does not depend on service times. Then the first busy period in this queue has the same distribution as $\sum_{i=1}^{K} B_i$ where B_i are *i.i.d.* random variables that have the same distribution as B and do not depend on K. By monotonicity,

$$\mathbb{P}(B > x, K_1 > N) \leq \mathbb{P}\left(\sum_{i=1}^{K} B_i > x, K > N\right) = \mathbb{P}\left(\sum_{i=1}^{K1(K>N)} B_i > x\right).$$

Due to (A.4), the latter probability is equivalent, as $x \to \infty$, to

$$\mathbb{E}(K1(K > N))\mathbb{P}(B > x) \leq C_0\mathbb{E}(K1(K > N))\mathbb{E}K\overline{G}(x)$$

where C_0 is from (4.2). Now choose N such that $C_0\mathbb{E}(K1(K > N))\mathbb{E}K \leq \varepsilon$. Since $\mathbb{P}(U > x) \geq \overline{G}(x)$, (4.7) follows. $\square$

We can go further and obtain the following result.

Lemma 4.5 *Assume the conditions of Lemma 4.2 hold, and let*

$$G_{N,R}(x) = \bigcup_{k=1}^{N}\bigcup_{\ell=0}^{2N}\left\{\{U > x, K_1 = k + \ell\} \cap \bigcup_{i=0}^{R}\{\sigma_{k,i} > x(1-\rho)\}\right\}, \tag{4.10}$$

Then, for any $\varepsilon > 0$, one can choose a positive integer R such that

$$\mathbb{P}(D_N(x)) \geq \mathbb{P}(G_{N,R}(x)) \geq (1-\varepsilon)\mathbb{P}(D_N(x)) \tag{4.11}$$

where the event $D_N(x)$ was defined in (4.8). *Further,*

$$\mathbb{P}(G_{N,R}(x)) \sim \sum_{k=1}^{N} \sum_{\ell=0}^{2N} \sum_{j=0}^{R} \sum_{i=0}^{j} \mathbb{P}(U > x, K_1 = k+\ell, X_{k-1} = j, \sigma_{k,i} > x(1-\rho)). \tag{4.12}$$

Proof Indeed,

$$\mathbb{P}\left(D_N(x) \setminus G_{N,R}(x)\right)$$

$$\leq \sum_{k=1}^{N} \sum_{\ell=0}^{2N} \sum_{j>R} \mathbb{P}\left(\{U > x, K_1 = k+\ell, X_{k-1} = j\} \bigcap \bigcup_{i=R+1}^{j} \{\sigma_{k,i} > x(1-\rho)\}\right)$$

$$\leq \sum_{k=1}^{N} \sum_{j>R} \sum_{i=0}^{j} \mathbb{P}(X_{k-1} = j, \sigma_{k,i} > x(1-\rho))$$

$$= \sum_{k=1}^{N} \sum_{j>R} (j+1)\mathbb{P}(X_{k-1} = j)\overline{G}(x(1-\rho))$$

$$= \sum_{k=1}^{N} \mathbb{E}((X_{k-1} + 1)1(X_{k-1} > R))\overline{G}(x(1-\rho))$$

$$\leq N\mathbb{E}((\tau + 1)1(\tau > R))\overline{G}(x(1-\rho))$$

where the term $\mathbb{E}((\tau + 1)1(\tau > R))$ may be made as small as possible by taking a sufficiently large R. Then (4.12) follows since the probability of a union of events is always smaller than the sum of their probabilities, and is bigger than the sum of probabilities of events minus the sum of probabilities of pairwise intersections of events. Each probability of intersection of two independent events is smaller than

$$\mathbb{P}(\sigma_1^{(1)} > x(1-\rho), \sigma_2^{(1)} > x(1-\rho)) \leq C\overline{G}^2(x) = o(\overline{G}(x)),$$

therefore their finite sum is $o(\overline{G}(x))$ and (4.12) follows.

We are now in a final step of the proof of Theorem 3.2. For $k \geq 1$, $\ell, j \geq 0$, define $D_{k,\ell,j}$ as

$$D_{k,\ell,j} = \mathbb{P}(K = k+\ell, X_{k-1} = j) = \mathbb{P}(K > k, X_{k-1} = j)\mathbb{P}(K = \ell), \tag{4.13}$$

where the second equality holds because K is geometrically distributed. Then, Lemma 4.5 implies (3.11) for $g_{k,\ell,i,j}(x) = P_{k,\ell,i,j}(x)$ since, for any $k, \ell, j \geq i$,

$$g_{k,\ell,i,j}(x) \leq \mathbb{P}(K = k+\ell, X_{k-1} = j, \sigma_{k,i} > x(1-\rho)) = D_{k,\ell,j}\mathbb{P}(\sigma > x(1-\rho)),$$

and

$$\sum_{k,\ell,i\leq j} D_{k,\ell,j} = \sum_{k,\ell,j} jD_{k,\ell,j} = \sum_{k,j} j\mathbb{P}(X_{k-1} = j, K > k) \leq \mathbb{E}\tau/q < \infty, \tag{4.14}$$

where, recall, τ is the total number of customers served in the first busy period. Clearly, (3.11) is also valid for a general $\{g_{k,\ell,i,j}(x)\}$ because of the conditions (3.9) and (3.10). This completes the proof of Theorem 3.2. $\square$

4.3 Proof of Theorem 2.1

We first recall the notation: $U_1, U_2, \ldots$ and $X_0, X_1, \ldots$ are the service cycles and the number of customers other than the tagged customer served in the cycles, respectively. Here $u_0 = X_0 = 0$. In general, the sojourn time is a randomly stopped sum of $i.i.d.$ positive random variables, and both the summands and the counting random variable have heavy-tailed distributions. It is known that it is hard to study the tail asymptotics for general heavy-tailed distributions (see, e.g., [10]) in this case. We proceed under the assumption that the service time distribution is intermediate regularly varying.

Recall that $\sigma_{k,0}$ is the kth service time of the tagged customer and, for $i = 1, \ldots, X_k, \sigma_{k,i}$ is the ith service time in the queue X_k. Further, $T_k = \sum_{\ell=1}^{k} U_\ell$ be the time instant when the kth service of the tagged customer is completed, where $U_1 = \sigma_{1,0}$. Introduce the notation

$$U_k^+ = \sum_{\ell=k+1}^{K} U_\ell, \qquad k \geq 1,$$

which is the remaining time the tagged customer spends in the system after the completion of the kth service, and let v_k be the residual inter-arrival time of the input when the kth service of the tagged customer ends.

In what follows, we will say that an event involving some constants and functions/sequences occurs "with high probability" if, for any $\varepsilon > 0$, there exists constants and functions/sequences (that depend on ε) with the desired properties such that the event occurs with probability at least $1 - \varepsilon$.

For example, let $S_n^\sigma = \sum_1^n \sigma_i$ be the sum of $i.i.d.$ random variables with finite mean b. Then the phrase "with high probability (WHP), for all $n = 1, 2, \ldots$,

$$S_n^\sigma \in (n(a - \delta_n) - C, n(a + \delta_n) + C)$$

with $C > 0$ and $\delta_n \downarrow 0$" means that "for any $\varepsilon > 0$, there exist a constant $C \equiv C_\varepsilon > 0$ and a sequence $\delta_n \equiv \delta_n(\varepsilon) \downarrow 0$ such that the probability of the event

$$\{S_n^\sigma \in (n(a - \delta_n) - C, n(a + \delta_n) + C), \text{ for all } n \geq 1\}$$

is at least $1 - \varepsilon$". We can say equivalently that "WHP, for all $n = 1, 2, \ldots, S_n^\sigma \in (na - o(n), na + o(n))$ ", or, simply, "WHP, $S_n^\sigma \sim an$", and this means that "for any $\varepsilon > 0$, there exists a positive function $h(n) = h_\varepsilon(n)$ which is an $o(n)$-function (it may tend to infinity, but slower than n) and is such that the probability of the event

$$\{S_n^\sigma \in (na - h(n), na + h(n)), \text{ for all } n\}$$

is at least $1 - \varepsilon$."

Now we show (3.8) for

$$g_{k,\ell,i,j}(x) = \mathbb{P}(K = k + \ell, X_{k-1} = j)\,\mathbb{P}((1 + m_\ell^{(0)})\sigma_{k,i} > x),$$
$$k \geq 1, \ell \geq 0, j \geq 0, 0 \leq i \leq j. \tag{4.15}$$

Namely, we show that, for all $k \geq 1, \ell \geq 0, 0 \leq i \leq j$,

$$\mathbb{P}(U > x, K = k + \ell, X_{k-1} = j, \sigma_{k,i} > x(1 - \rho)) \sim g_{k,\ell,i,j}(x). \tag{4.16}$$

We prove (4.16) by induction on $\ell \geq 0$, for each fixed $k \geq 1, 0 \leq i \leq j$.
Lower bound $\ell = 0$.

Since $\sigma_{k,i} > x$ implies that $U > x$ and $\sigma_{k,i} > (1 - \rho)x$, the lower bound for the LHS of (4.16) is

$$\mathbb{P}(K = k, X_{k-1} = j)\overline{G}(x) = g_{k,0,i,j}(x),$$

because $m_0^{(0)} = 0$.

Upper bound $\ell = 0$.
There is a constant $w > 0$ such that $T_{k-1} \leq w$ and $\sum_{0 \leq i' \leq j, i' \neq i} \sigma_{k,i'} \leq w$ WHP. Then $U \leq 2w + \sigma_{k,i}$, so the upper bound for the LHS of (4.16) is

$$\varepsilon\overline{G}(x(1 - \rho)) + (1 + o(1))\mathbb{P}(K = k, X_{k-1} = j, \sigma_{k,i} + 2w > x)$$
$$\sim \varepsilon\overline{G}(x(1 - \rho)) + \mathbb{P}(K = k, X_{k-1} = j)\overline{G}(x).$$

Letting ε tend to zero in this upper bounds yields that the lower and upper bounds are asymptotically identical. Since $m_0^{(0)} = 0$, they are further identical to $g_{k,0,i,j}(x)$ of (4.15). Thus, (4.16) is verified.

Turn to the case $\ell = 1$.

Lower bound $\ell = 1$.
Like in the case $\ell = 0$, replace all other service times $\sigma_{k,i'}, i' \neq i$ by zero. Assume that all j customers from the group X_{k-1} leave the system after their service completions. WHP, $v_{k-1} \leq w$. Given $y = \sigma_{k,i}$ is large and much bigger than w, we have that at least $N^e(y - w)$ customers arrive during time $U_k \geq \sigma_{k,i} = y$. Again WHP,

$$N^e(y - w) \in (\lambda y - o(y), \lambda y + o(y))$$

and, again WHP, their total service time is within the time interval $(\lambda by - o(y), \lambda by + o(y))$. Therefore,

$$U \geq U_k + U_{k+1} \geq y + \lambda by - o(y)$$

and the RHS is bigger than x if $y > x/(1 + \lambda b) + o(x)$. Therefore, the lower bound for the LHS of (4.16) is

$$(1 + o(1))\mathbb{P}(K = k + 1, X_{k-1} = j, \sigma_{k-1,i} > x/(1 + \lambda b) + o(x)) - \varepsilon\overline{G}(x(1 - \rho))$$
$$\sim \mathbb{P}(K = k + 1, X_{k-1} = j)\overline{G}(x/(1 + \lambda b)) - \varepsilon\overline{G}(x(1 - \rho)).$$

Upper bound $\ell = 1$.
WHP, $T_{k-1} \leq w$, $v_k \leq w$ and $\sum_{0 \leq i' \leq j, i' \neq i} \sigma_{k,i'} \leq w$. Let $\sigma_{k,i} = y \gg 1$. Then, WHP, $U_k \leq y + 2w$ and the number of external arrivals within U_k is bounded above by $1 + N^e(y + 2w) = \lambda y + o(y)$, again WHP. Assume that all $X_{k-1} = j$ customers stay in the system after their services. Then again $j + 1 + N^e(y + w) = \lambda y + o(y)$, WHP. Therefore, $U_{k+1} = b\lambda y + o(y)$. Then we arrive at the upper bound that meets the lower bound.

Thus, (4.16) is verified for $g_{k,2,i,j}(x)$ because $m_1^{(0)} = \lambda b$ by (2.16).

Induction step
We can provide induction for any finite number of steps. Here is the induction base. Assume that $\sigma_{k,i} = y \gg 1$ and that, after $\ell \geq 1$ steps, $T_{k+\ell'} \sim (1 + m_{\ell'}^{(0)})y$ for $0 \leq \ell' \leq \ell$, and there are $X_{k+\ell-1}$ customers in the queue and that $X_{k+\ell-1} = wy + o(y)$, WHP, where $w > 0$. Then, combining upper and lower bounds, we may conclude that, again WHP, $U_{k+\ell} = bwy + o(y)$ and then

$$T_{k+\ell} = T_{k+\ell-1} + U_{k+\ell} \sim (1 + m_{\ell-1}^{(0)} + bw)y, \tag{4.17}$$

$$X_{k+\ell} = pX_{k+\ell-1} + \lambda U_{k+\ell} + o(y) = wyr + o(y), \tag{4.18}$$

where we recall that $r = p + \lambda b$. By the induction hypothesis, (4.17) implies that

$$1 + m_\ell^{(0)} = 1 + m_{\ell-1}^{(0)} + bw,$$

which, with (2.16), yields that

$$bw = m_\ell^{(0)} - m_{\ell-1}^{(0)} = r^{\ell-1}\lambda b.$$

Hence, by (4.18),

$$T_{k+\ell+1} = T_{k+\ell} + U_{k+\ell+1} \sim (1 + m_\ell^{(0)})y + bwry$$
$$= (1 + m_\ell^{(0)} + r^\ell \lambda b)y = (1 + m_{\ell+1}^{(0)})y.$$

This completes the induction step for $\ell + 1$.

We finally check the conditions (3.9) and (3.10). Since an intermediate regularly varying distribution is dominantly varying, it follows from (4.15) that

$$g_{k,\ell,i,j}(x) \leq \mathbb{P}(K = k + \ell, X_{k-1} = j)\,\mathbb{P}((1 + m_\infty^{(0)})\sigma > x)$$
$$\leq \mathbb{P}(K = k + \ell, X_{k-1} = j)\,c\,\mathbb{P}(\sigma > x)$$

for some $c > 0$. Hence, letting

$$C_{k,\ell,i,j} = \mathbb{P}(K = k + \ell, X_{k-1} = j)\,c,$$

(3.9) is verified, while (3.10) follows from (4.14). Thus, by Theorem 3.2,

$$\mathbb{P}(U > x) \sim \sum_{k=1}^\infty \sum_{\ell=0}^\infty \sum_{j=0}^\infty \sum_{i=0}^j \mathbb{P}(K = k + \ell, X_{k-1} = j)\,\mathbb{P}\left((1 + m_\ell^{(0)})\sigma > x\right)$$
$$= \frac{1}{q}\sum_{k=1}^\infty \sum_{\ell=0}^\infty qp^k\,\mathbb{E}(1 + X_{k-1})\,qp^{\ell-1}\mathbb{P}\left((1 + m_\ell^{(0)})\sigma > x\right),$$

which implies (2.17), and Theorem 2.1 is proved.

4.4 PSBJ for the Stationary Queue

We now consider the case where customer 1 arrives to the stationary queue and denote by U^0 its sojourn time. In this section, we frequently use the following notation: for a distribution F having a finite mean, $\overline{F_I}(x) = \min(1, \int_x^\infty \overline{F}(y)dy)$ is its integrated tail distribution (see (a) of Remark 2.2).

By "stationarity" we mean stationarity in discrete time, i.e. at embedded arrival epochs. So we assume that the system has started from time $-\infty$ and that customer 1 arrives at time $\tilde{t}_1 \equiv 0$, customers with indices $k \leq 0$ enter the system at time instants $\tilde{t}_k = -\sum_{j=k}^0 t_j$ and customers with indices $k \geq 2$ at time instants $\tilde{t}_k = \sum_{j=1}^{k-1} t_j$. For $k \leq \ell$, let $S_{k,\ell}^H = \sum_{j=k}^\ell \xi_j^H$, where $\xi_j^H = \sigma_j^H - t_j$. Then the stationary busy cycle covering 0 starts at $\tilde{t}_k, k \leq 0$ if

$$\left\{\sup_{j<k} S_{j,k}^H \leq 0, \min_{k \leq \ell \leq 0} S_{k,\ell}^H > 0\right\}.$$

So, if B^0 is the remaining duration of the busy period viewed at time 0, then

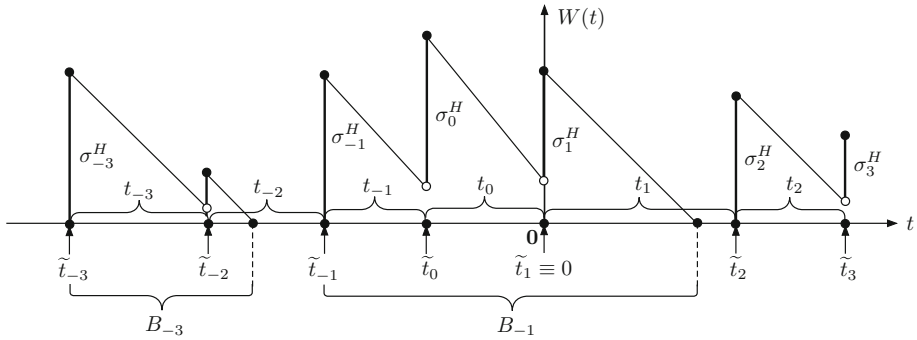$$\{B^0 > x\} = \bigcup_{k=-\infty}^0 \left\{\sup_{j<k} S_{j,k}^H \leq 0, B_k > -\tilde{t}_k + x\right\}, \tag{4.19}$$

Fig. 2 Workload and time indexes under the stationary regime: $B^0 = B_{-1} + \widetilde{t}_{-1}$ and $\arg_k\{\min_{k \le \ell \le 0} S^H_{k,\ell} > 0\} = -1$. At and after time 0, the time indexes are identical with those in the previous sections

where B_k is the duration of the period that starts at time $\widetilde{t}_k$ given that customer k arrives in the empty system (then, in particular, $B = B_0$). See Fig. 2.

Let

$$\tau^H_k = \min\left\{n \ge 1; S^H_{k,n} \le 0\right\}, \qquad k \le 0,$$

which is the number of customers arriving at or after time 0 in the busy period when it starts at time $\widetilde{t}_k$, and let

$$A^H_k = \left\{\sup_{j<k} S^H_{j,k} \le 0\right\}, \qquad k \le 0,$$

then $A^H_k \cap \{\tau^H_k \ge \ell\}$'s for $k \le 0, \ell \ge -k + 1$ are disjoint sets.

Thus, (4.19) may be written as

$$\{B^0 > x\} = \bigcup_{k=0}^{\infty} \left(A^H_{-k} \cap \left\{\tau^H_{-k} \ge k+1, B_{-k} > -\widetilde{t}_{-k} + x\right\}\right), \qquad (4.20)$$

and therefore, applying the PSBJ of Corollary 3.1 to each busy period B_{-k},

$$\{B^0 > x\} = \bigcup_{k=0}^{\infty} \left(A^H_{-k} \cap \left\{\tau^H_{-k} \ge k+1, B_{-k} > -\widetilde{t}_{-k} + x\right\}\right)$$

$$\simeq \bigcup_{k=0}^{\infty} \bigcup_{i=1}^{\infty} \left(A^H_{-k} \cap \left\{\tau^H_{-k} \ge \max(k+1, i), \sigma^H_{-k+i} > (-\widetilde{t}_{-k} + x)(1 - \rho)\right\}\right),$$

where $\sigma^H_{-k+i}, i \ge 0$ is the service time of the ith customer arriving in the busy period that starts at time $\widetilde{t}_{-k}$.

Hence, letting

$$A^0_-(x) = \bigcup_{k=1}^{\infty} \bigcup_{i=1}^{k} \left(A^H_{-k} \cap \left\{\tau^H_{-k} \ge k+1, \sigma^H_{-k+i} > (-\widetilde{t}_{-k} + x)(1 - \rho)\right\}\right)$$

$$A^0_+(x) = \bigcup_{k=0}^{\infty} \bigcup_{i=k+1}^{\infty} \left(A^H_{-k} \cap \left\{\tau^H_{-k} \ge i, \sigma^H_{-k+i} > (-\widetilde{t}_{-k} + x)(1 - \rho)\right\}\right),$$

we have

$$\{B^0 > x\} \simeq A^0_-(x) \cup A^0_+(x). \tag{4.21}$$

We first consider the event $A^0_+(x)$, which is a contribution of big jumps at or after time 0, and show that its probability is negligible with respect to $\overline{H_I}(x)$, as $x \to \infty$. Clearly, for any positive function $h(x)$ and for any $\varepsilon \in (0, a)$,

$$A^0_+(x) \subseteq \bigcup_{k \geq 0} \{-\widetilde{t}_{-k} < (a - \varepsilon)k - h(x)\}$$

$$\bigcup_{k \geq 0} \bigcup_{i \geq k+1} \left\{ \tau^H_{-k} \geq i, \sigma^H_{-k+i} > ((a - \varepsilon)k + x - h(x))(1 - \rho)) \right\}.$$

Then

$$\mathbb{P}(A^0_+(x)) \leq \sum_{k=0}^{\infty} \sum_{i=k+1}^{\infty} \mathbb{P}(\tau^H_{-k} \geq i, \sigma^H_{-k+i} > (-\widetilde{t}_{-k} + x)(1 - \rho))$$

$$\leq \sum_{k \geq 0} \mathbb{P}(-\widetilde{t}_{-k} < (a - \varepsilon)k - h(x))$$

$$+ \sum_{k \geq 0} \sum_{i \geq k+1} \mathbb{P}(\tau^H_{-k} \geq i)\mathbb{P}(\sigma^H_{-k+i} > ((a - \varepsilon)k + x - h(x))(1 - \rho))$$

$$\leq Ce^{-\alpha h(x)} + \sum_{k \geq 0} \mathbb{E}\left(\left(\tau^H - k + 1\right)^+\right)\mathbb{P}\left(\sigma^H_{-k+i} > ((a - \varepsilon)k + x - h(x))(1 - \rho)\right)$$

$$= o(\overline{H_I}(x(1 - \rho))) = o(\overline{G_I}(x(1 - \rho))),$$

if one takes, say, $h(x) = x^c$ for some $c < 1$. Here the second inequality follows since $\{\tau^H_{-k+i} \geq i\} = \{\tau^H_{-k+i} \leq i - 1\}^c$ is independent of σ^H_{-k+i}, the third inequality from Chernoff's inequality, for a small $\alpha > 0$, and the final conclusion from property (A.7) in the Appendix.

Thus, we only need to evaluate the contribution of big jumps that occur before time 0. Namely, we analyse $A^0_-(x)$. Note that, for any $k_0 > 0$, the probability of the event

$$A^{(0,k_0)}_-(x) = \bigcup_{k=1}^{k_0} \bigcup_{i=1}^{k} \left(A^H_{-k} \bigcap \left\{ \tau^H_{-k} \geq k + 1, \sigma^H_{-k+i} > (-\widetilde{t}_{-k} + x)(1 - \rho) \right\} \right)$$

is of order $O(\overline{G}(x(1 - \rho)))$ which is negligible with respect to $\overline{G_I}(x(1 - \rho))$. Therefore, one can choose an integer-valued $h(x) \to \infty$ such that $\mathbb{P}\left(A^{(0,h(x))}_-(x)\right) = o(\overline{G_I}(x(1 - \rho)))$. So we may apply again the SLLN, $\widetilde{t}_{-k} \sim ak$ for sufficiently large k, to get

$$\bigcup_{k=1}^{\infty} \left(A^H_{-k} \bigcap \left(\bigcup_{i=1}^{k} \left\{ \tau^H_{-k} \geq k + 1, \sigma^H_{-k+i} > (-\widetilde{t}_{-k} + x)(1 - \rho) \right\} \right) \right)$$

$$\simeq \bigcup_{k=1}^{\infty} \bigcup_{\ell=0}^{k-1} \left(A^H_{-k} \bigcap \left\{ \tau^H_{-k} \geq k + 1, \sigma^H_{-\ell} > (ak + x)(1 - \rho) \right\} \right)$$

$$= \bigcup_{\ell=0}^{\infty} \bigcup_{k=\ell+1}^{\infty} \left(A^H_{-k} \bigcap \left\{ \tau^H_{-k} \geq k + 1, \sigma^H_{-\ell} > (ak + x)(1 - \rho) \right\} \right)$$

$$\subseteq \bigcup_{\ell=0}^{\infty} \left\{ \sigma_{-\ell}^H > (a(\ell+1)+x)(1-\rho) \right\}.$$

On the other hand, for $h(x) \uparrow \infty$ sufficiently slowly and for an appropriate sequence $\varepsilon_\ell \downarrow 0$ (that comes from the SLLN), we have

$$\bigcup_{\ell=0}^{\infty} \left\{ \sigma_{-\ell}^H > (a(\ell+1)+x)(1-\rho) \right\} \sim \bigcup_{\ell=h(x)}^{\infty} \left\{ \sigma_{-\ell}^H > (a(\ell+1)+x)(1-\rho) \right\}$$

$$\sim \bigcup_{\ell=h(x)}^{\infty} \left\{ \sigma_{-\ell}^H > ((a+2\varepsilon_\ell)(\ell+1)+x+h(x))(1-\rho)+h(x) \right\} \bigcap D_{-(\ell+1)} \equiv E(x),$$

where

$$D_{-\ell} = \bigcap_{j=1}^{\infty} \left\{ \sum_{i=-\ell+1}^{-\ell+j} t_i \leq (a+\varepsilon_j)j + h(x), \sum_{i=-\ell+1}^{-\ell+j} \sigma_i^H \geq (b^h - \varepsilon_j)j - h(x) \right\}.$$

Since $B^0 > x$ on the event $E(x)$, we arrive at the following PSBJ for the stationary busy period.

Lemma 4.6 *If the $GI/GI/1$ feedback queue is stable and its service time distribution has an $\mathcal{IRV}$ distribution with a finite mean, then*

$$\{B^0 > x\} \simeq \bigcup_{k=0}^{\infty} \left\{ \sigma_{-k}^H > (x + a(k+1))(1-\rho) \right\}. \tag{4.22}$$

The lemma implies that

$$\mathbb{P}(B^0 > x) \sim \sum_{k=0}^{\infty} \mathbb{P}(\sigma_{-k}^H > (x + a(k+1))(1-\rho)) \tag{4.23}$$

since the sum of the probabilities of pairwise intersections is of order $O(\overline{G_I}^2(x)) = o(\overline{G_I}(x))$.

Then we may conclude that the principle of a single big jump can be applied to the stationary sojourn time too:

$$\mathbb{P}(U^0 > x) \sim \sum_{n=0}^{\infty} \mathbb{P}(U^0 > x, \sigma_{-n}^H > (x + (n+1)a)(1-\rho))$$

$$\sim \sum_{n=h(x)}^{\infty} \mathbb{P}(U^0 > x, \sigma_{-n}^H > (x + (n+1)a)(1-\rho)) \tag{4.24}$$

where the second equivalence is valid for any integer-valued function $h(x) \uparrow \infty, h(x) = o(x)$ and follows from (4.19) and from the properties of $\mathcal{IRV}$ and integrated tail distributions, see Appendix A.

4.5 Proof of Theorem 2.2

First, we comment that it is easy to obtain the logarithmic asymptotics for the stationary sojourn time. Since the sojourn time of the customer entering the stationary queue at time 0 is not bigger than the stationary busy period and is not smaller than the stationary sojourn time

in the auxiliary queue without feedback, and since both bounds have tail distributions that are proportional to the integrated tail distribution of a single service time (see the Appendix for definitions), we immediately get the logarithmic tail asymptotics:

$$\log \mathbb{P}(U^0 > x) \sim \log \overline{H_I}(x) \sim \log \overline{G_I}(x). \tag{4.25}$$

Now we provide highlights for obtaining the exact tail asymptotics for the stationary sojourn time distribution and give the final answer. For this, we use the following simplifications, which are made rigorous in the "WHP" terminology and due to the $o(x)$-insensitivity of the service-time distribution.

(1) We observe that the order of services prior to time 0 is not important for the customer that enters the stationary queue at time 0: the joint distribution of the residual service time and of the queue length at time 0 stays the same for all reasonable service disciplines (that do not allow processor sharing). So we may assume that, up to time 0, all arriving customers are served in order of their external arrival: the system serves the "oldest" customer a geometric number of times and then turns to the service of the next customer.

(2) We simplify the model by assuming that all inter-arrival times are deterministic and equal to $a = \lambda^{-1}$.

(3) We further assume that all service times of all customers but one are equal to b, so every customer but one has a geometric number of services of length b. The "exceptional" customer may be any customer $-n \leq 0$, it has a geometric number of services, one of those is random and large and all others equal to b. So the total service time of the "exceptional" customer has the tail distribution equivalent to

$$\mathbb{E}K \cdot \overline{G}(x) = \frac{1}{q}\overline{G}(x).$$

(4) We assume that the "exceptional" customer arrives at an empty queue, that is, the workload found by this customer is negligible compared with his exceptional service time.

Due to the arguments explained above, we can show that the tail asymptotics of the sojourn time of customer 1 in the original and in the auxiliary system are equivalent. We start by repeating our calculations from the proof of Theorem 2.1, but in two slightly different settings.

Assume all service times but the very first one are equal to b for the exceptional customer arriving at or before time 0. Assume that, if customer 1 arriving at time 0 is not exceptional, then it finds $X_0 = N$ customers in the queue, and otherwise it finds a negligible number of customers compared with N while its first service time is Nb. Assume customer 1 leaves the system after $K = k$ services. Denote, as before, by U_i the time between its $(i-1)$st and ith services and by X_i the queue behind customer 1 after its ith service completion. How large should N be for the sojourn time of customer 1 to be bigger than x where x is large?

(A) Assume that the (residual) service time, z, of the very first customer in the queue is not bigger than b (so we may neglect it). When N is large, we get that $U_1 \sim Nb$. Then we have

$$X_i \sim X_{i-1}p + \lambda U_i, \qquad U_i \sim X_{i-1}b, \qquad i = 1, 2, \ldots, k.$$

Hence, $X_i \sim X_{i-1}p + \lambda U_i \sim Nr^i$ and $U_i \sim Nbr^{i-1}$ where $r = p + \lambda b < 1$. Then $U_1 + \cdots + U_k \sim Nb(1 - r^k)/(1 - r)$. Thus, we may conclude that

$$\mathbb{P}(U > x, K = k) \sim qp^{k-1}\mathbb{P}(Nb > x_k), \tag{4.26}$$

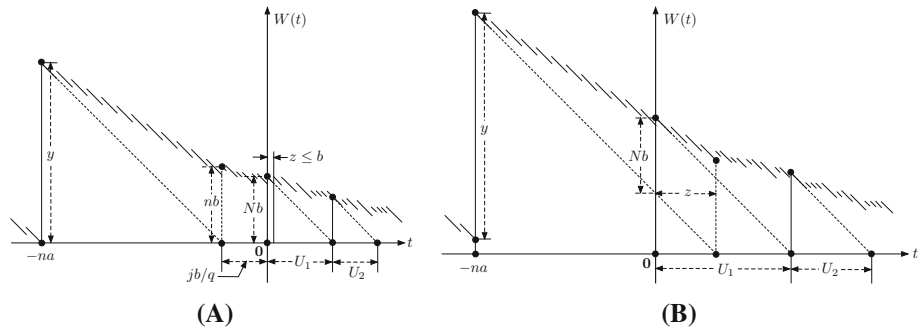where $x_k = x(1 - r)/(1 - r^k)$.

Fig. 3 Sample path of the workload $W(t)$ for the cases (A) and (B) with $K = 2$

(B) Assume now that both $X_0 = N$ and z are large. Then $U_1 \sim z + Nb$ and $X_1 \sim Np + \lambda(z + Nb)$ and, further, $X_i \sim X_{i-1}r \sim X_1 r^{i-1}$ and $U_{i+1} \sim X_i b \sim X_1 b r^{i-1} \sim Nbr^i + zr^i$. Thus, $U_1 + \cdots + U_k \sim Nb(1 - r^k)/(1 - r) + z(1 - r^k)/(1 - r)$ and

$$\mathbb{P}(U > x, K = k) \sim qp^{k-1}\mathbb{P}(Nb + z > x(1 - r)/(1 - r^k)). \quad (4.27)$$

Let $W(t)$ be the total work in the system at time t. We illustrate $W(t)$ below to see how the cases (A) and (B) occur.

We will see now that if $K = k$ and if there is a big service time of the $(-n)$th "exceptional" customer, then the case (A) occurs if $n > x_k/b$ and the case (B) if $n < x_k/b$ (Fig. 3).

Let the big service time take value $y \gg 1$. Recall from (4.24) that it is enough to consider values of $n \geq h(x)$ only, where $h(x) \uparrow \infty$, $h(x) = o(x)$.

For any $k \geq 1$, assume $K = k$ and $y \leq na$, then the exceptional service is completed before or at time 0, and the situation (A) occurs. Hence, $X_0 \equiv N = n - j$ for some nonnegative $j \leq n$, and $y + jb/q \approx na$ because approximately j further customers leave the system prior to time 0. Then $U \sim Nb(1 - r^k)(1 - r)$, and $U > x$ is asymptotically equivalent to $Nb(1 - r^k)(1 - r) > x$, where the last inequality is identical with $(n - j)b > x_k$. This together with $j \approx (na - y)q/b$ implies that

$$y \gtrsim x_k/q + n(a - b/q).$$

Since $na \geq y$, this further implies that $n \gtrsim x_k/b$.

We next assume $K = k$ and $n < x_k/b$. Then, the contraposition of the above implication implies that $y > na$, and the situation (B) occurs. Therefore, we should take $y = z + na$, and $U_0 = nb + z$, where $N = n$. Since $U \sim (nb + z)(1 - r^k)/(1 - r)$, $U > x$ is equivalent to $nb + z \gtrsim x_k$, and therefore

$$y \gtrsim x_k + n(a - b).$$

Combining together both cases, we obtain the following result:

$$\mathbb{P}(U^0 > x) \sim q \sum_{k=1}^{\infty} p^{k-1} \left(\sum_{n=1}^{x_k/b-1} \mathbb{P}(\sigma > x_k + n(a - b)) \right.$$

$$\left. + \sum_{n=x_k/b}^{\infty} \mathbb{P}(\sigma > x_k/q + n(a - b/q)) \right). \quad (4.28)$$

Clearly, the second sum in the parentheses is equivalent to

$$\sum_{n=0}^{\infty} \mathbb{P}(\sigma > x_k a/b + n(a - b/q)) \sim (a - b/q)^{-1} \overline{G_I}(x_k a/b)$$

while the first sum in the parentheses is

$$\sum_{n=1}^{\infty} - \sum_{n=x_k/b}^{\infty} \sim (a - b)^{-1} \left(\overline{G_I}(x_k) - \overline{G_I}(x_k a/b) \right).$$

Hence, we have

$$\mathbb{P}(U^0 > x) \sim \frac{q}{a - b} \sum_{k=1}^{\infty} p^{k-1} \left(\overline{G_I}(x_k) + \frac{b(1 - q)}{aq - b} \overline{G_I}(x_k a/b) \right), \qquad (4.29)$$

where we recall that $x_k = \frac{x(1-r)}{1-r^k}$, for $k \geq 1$. Since $m_K^{(0)} = \mathbb{E}\left(\frac{1-r^K}{1-r}\right)$, we arrive at (2.19) and (2.20).

Appendix

A Properties of Heavy-Tailed Distributions

We revise basic properties of several classes of heavy-tailed distributions listed at the end of Sect. 1 (see [9] for the modern theory of heavy tailed distribution and other books [1,3] for more detail), and formulate a part of the main result from [7] that plays an important role in our analysis.

Let $\{\xi_n\}_{-\infty}^{\infty}$ be $i.i.d.$ r.v.'s with finite mean $\mathbb{E}\xi_1$ and with $\mathbb{P}(\xi_1 > 0) > 0$ and $\mathbb{P}(\xi_1 < 0) > 0$. Let $F(x) = \mathbb{P}(\xi_1 \leq x)$ be their common distribution and $\overline{F}(x) = 1 - F(x) = \mathbb{P}(\xi_1 > x)$ its tail. Let $m^+ \equiv m^+(F) = \mathbb{E}\max(0, \xi_1) = \int_0^{\infty} \overline{F}(t)dt$. Let $S_0 = 0$, $S_n = \sum_1^n \xi_i$, and $\tau = \min\{n \geq 1 : S_n \leq 0\}$.

If $F \in \mathcal{L}$ (long tailed), that is, (1.3) holds for some $y > 0$, then it holds for all y and, moreover, uniformly in $|y| \leq C$, for any fixed C. Therefore, if $F \in \mathcal{L}$, then there exists a positive function $h(x) \to \infty$ such that $\overline{F}(x - h(x)) \sim \overline{F}(x) \sim \overline{F}(x + h(x))$. In this case we say that the tail distribution $\overline{F}$ is h-insensitive.

In what follows, we make use of the following characteristic result (see Theorem 2.47 in [9]):

$$F \in \mathcal{IRV} \text{ if and only if } F \text{ is } h - \text{insensitive, for any } h(x) = o(x). \qquad (A.1)$$

In particular, if F is h-insensitive, then F is h_c-insensitive for any $c > 0$, where $h_c(x) = h(cx)$.

We also use another characteristic result which is a straightforward minor extension of Theorem 2.48 from [9]:

$$F \in \mathcal{IRV} \text{ if and only if } \overline{F}(V_n) \sim \overline{F}(v_n), \qquad (A.2)$$

for any sequence of non-negative random variables V_n with corresponding means $v_n = \mathbb{E}V_n$ satisfying

$$V_n \to \infty \text{ and } V_n/v_n \to 1 \text{ in probability.}$$

Here is another good property of $\mathcal{IRV}$ distributions. Let random variables X and Y have arbitrary joint distribution, with the distribution of X being $\mathcal{IRV}$ and $\mathbb{P}(|Y| > x) = o(\mathbb{P}(X > x))$. Then

$$\mathbb{P}(X + Y > x) \sim \mathbb{P}(X > x) \text{ as } x \to \infty. \tag{A.3}$$

If F is an $\mathcal{IRV}$ distribution with finite mean, then the distribution with the integrated tail $\overline{F_I}(x) = \min(1, \int_x^\infty \overline{F}(y)dy)$ is also $\mathcal{IRV}$ and $\overline{F}(x) = o(\overline{F_I}(x))$ and, moreover, $\int_x^{x+h(x)} \overline{F}(y)dy = o(\overline{F_I}(x))$ if F_I is h-insensitive.

We use the following well-known result: if $\{\sigma_{1,j}\}$ is an $i.i.d.$ sequence of random variables with common subexponential distribution F and if the counting random variable K does not depend on the sequence and has a light-tailed distribution, then

$$\left\{\sum_1^K \sigma_{1,j} > x\right\} \simeq \cup_1^K \{\sigma_{1,j} > x\} \text{ and } \mathbb{P}\left(\sum_1^K \sigma_{1,j} > x\right) \sim \mathbb{E}K\overline{F}(x), \quad x \to \infty. \tag{A.4}$$

Here is the principle of a single big jump again: the sum is large when one of the summands is large.

Let $M = \sup_{n \geq 0} \sum_{i=1}^n \xi_i$ where $\{\xi_i\}$ are $i.i.d.$ r.v.'s with negative mean $-m$ and with common distribution function F such that F_I is subexponential. Then

$$\{M > x\} \simeq \cup_{n \geq 1}\{\xi_n > x + mn\},$$

$$\mathbb{P}(M > x) \sim \sum_{n \geq 1} \mathbb{P}(\xi_1 > x + mn) \sim \frac{1}{m}\overline{F_I}(x). \tag{A.5}$$

Further, if F_I is subexponential, then, for any sequence $m_n \to m > 0$ and any function $h(x) = o(x)$,

$$\sum_{n \geq 0} \overline{F}(x + m_n n + h(x)) \sim \frac{1}{m}\overline{F_I}(x) \tag{A.6}$$

and, for any sequence $c_n \to 0$,

$$\sum_{n \geq 0} c_n \overline{F}(x + m_n n + h(x)) = o\left(\overline{F_I}(x)\right). \tag{A.7}$$

Remark A.1 Let $\mathcal{K}$ be any of the classes $\mathcal{L}, \mathcal{RV}, \mathcal{IRV}, \mathcal{D}, \mathcal{S}, \mathcal{S}^*$. The property of belonging to class $\mathcal{K}$ is a *tail property*: if $F \in \mathcal{K}$ and if $\overline{G}(x) \sim C\overline{F}(x)$ where C is a positive constant, then $G \in \mathcal{K}$. In particular,

 (i) if $F \in \mathcal{K}$, then $F_+ \in \mathcal{K}$;
 (ii) if the random variable ξ has distribution $F \in \mathcal{K}$ and $c_1 > 0$ and c_2 are any constants, then the distribution of the random variable $\eta = c_1\xi + c_2$ also belongs to $\mathcal{K}$;
 (iii) if the random variable ξ may be represented as $\xi = \sigma - t$ where σ and t are mutually independent random variables and t is non-negative (or, slightly more generally, bounded from below), and if the distribution of σ belongs to class $\mathcal{K}$, then $\mathbb{P}(\xi > x) \sim \mathbb{P}(\sigma > x)$, so the distribution of ξ belongs to $\mathcal{K}$ too.

The following result is a part of Theorem 1 in [7], see also [8] for a more general statement.

Theorem A.1 *Let $S_n = \sum_1^n \xi$, $S_0 = 0$ be a random walk with i.i.d. increments with distribution function F and finite negative mean*

$$\mathbb{E}\xi_1 = -m < 0. \tag{A.8}$$

Assume $F \in \mathcal{S}^*$. Let $T \le \infty$ be any stopping time (with respect to $\{\xi_n\}$). Let $M_T = \max_{0 \le n \le T} S_n$. Then

$$\lim_{x \to \infty} \frac{\mathbb{P}(M_T > x)}{\overline{F}(x)} = \mathbb{E}T. \tag{A.9}$$

B Proof of Corollary 3.1

We first note that the event $\{\tau^H \ge n\}$ is independent of σ_n^H and ξ_n^H because $\{\tau^H \ge n\} = \{\tau^H \le n-1\}^c$ is $\sigma(\{\xi_\ell^H; 1 \le \ell \le n-1\})$-measurable. Hence,

$$\mathbb{P}(\tau^H \ge n, \sigma_n^H > x(1-\rho)) = \mathbb{P}(\tau^H \ge n)\mathbb{P}(\sigma_n^H > x(1-\rho)),$$

and therefore the equivalence (3.4) is immediate from (3.3) of Theorem 3.1, while (3.5) easily follows from (3.4).

Thus, it remains to prove (3.6). For this, we introduce some notation. Let $S_n^H = \sum_{i=1}^n \xi_i^H$ and let $S_n^{\sigma^H} = \sum_{i=1}^n \sigma_i^H$. Define a sequence of events E_n, $n = 0, 1, \ldots$, as

$$E_n = \cap_{\ell \ge 1} \left\{ \left(S_{\ell+n}^H - S_n^H \right) \ge (b_H - a)\ell - \delta_\ell \ell - C, \left(S_{\ell+n}^{\sigma^H} - S_n^{\sigma^H} \right) \ge b_H \ell - \delta_\ell \ell - C \right\}$$

which is stationary in n (here, by convention, $S_0^H = S_0^{\sigma^H} = 0$). Due to the SLLN, there exists a sequence $\delta_\ell \downarrow 0$ such that

$$\mathbb{P}\left(\left| S_\ell^H/\ell - (b_H - a) \right| \le \delta_\ell \text{ and } \left| S_\ell^{\sigma^H}/\ell - b_H \right| \le \delta_\ell, \text{ for all } \ell \ge n \right) \to 1,$$

as $n \to \infty$. Therefore, for any $\varepsilon > 0$, there exists $C = C_\varepsilon > 0$, for which E_n is denoted by $E_{n,\varepsilon}$, such that

$$\mathbb{P}(E_{n,\varepsilon}) = \mathbb{P}(E_{0,\varepsilon}) \ge 1 - \varepsilon. \tag{B.10}$$

Introduce a function $h_\varepsilon(x)$ by

$$h_\varepsilon(x) = \max_{i \le [x/a]} (i\delta_i) + C_\varepsilon + b_H,$$

where $[x/a]$ is the integer part of the ratio x/a. Then, for this ε and $n \ge 1$, define $J_{n,\varepsilon}(x)$ as

$$J_{n,\varepsilon}(x) = \left\{ \tau^H \ge n, \xi_n^H > x(1-\rho) + h_\varepsilon(x) \right\}, \qquad x > 0.$$

Then, on the event $J_{n,\varepsilon}(x) \cap E_{n,\varepsilon}$, we have $S_{n-1}^H > 0$, $S_n^H > \xi_n^H > x(1-\rho) + h_\varepsilon(x)$, and therefore

$$S_{n+\ell}^H > x(1-\rho) + h_\varepsilon(x) - (1-\rho)a\ell - \delta_\ell \ell - C_\varepsilon$$
$$= (1-\rho)(x - a\ell) + \max_{i \le [x/a]} (i\delta_i) - \delta_\ell \ell + b_H > 0, \qquad 0 \le \ell \le [x/a].$$

Hence, letting $\ell_0 = [x/a]$, we have, on the same event,

$$B \ge \sum_{i=n}^{n+\ell_0} \sigma_i^H = \sigma_n^H + S_{n+\ell_0}^{\sigma^H} - S_n^{\sigma^H} > x(1-\rho) + h_\varepsilon(x) + b_H \ell_0 - \delta_{\ell_0} \ell_0 - C_\varepsilon > x.$$

For any integer $N \geq 1$, let

$$L_{N,\varepsilon}(x) = \cup_{n=1}^{N} \left\{ \max_{i<n} \xi_i^H \leq x(1-\rho) \right\} \cap J_{n,\varepsilon}(x) \cap E_{n,\varepsilon},$$

then we have

$$\{B > x\} \supseteq L_{N,\varepsilon}(x). \tag{B.11}$$

Let F^H be the distribution of ξ^H, and recall that H is the distribution of σ^H. Both of them are intermediate regularly varying, they are tail-equivalent and h-insensitive [see Remark A.1 and (A.1)]. Since $L_{N,\varepsilon}(x)$ is a union of N disjoint events and $\{\tau^H \geq n, \max_{i<n} \xi_i^H \leq x(1-\rho)\}$ is independent of ξ_n^H and $E_{n,\varepsilon}$, (B.10) yields, as $x \to \infty$,

$$\mathbb{P}\big(L_{N,\varepsilon}(x)\big) = \sum_{n=1}^{N} \mathbb{P}\left(\tau^H \geq n, \max_{i<n} \xi_i^H \leq x(1-\rho) \right) \cdot \overline{F^H}(x(1-\rho) + h_\varepsilon(x)) \cdot \mathbb{P}(E_{n,\varepsilon})$$

$$\sim \sum_{n=1}^{N} \mathbb{P}(\tau^H \geq n) \overline{H}(x(1-\rho)) \mathbb{P}(E_{n,\varepsilon})$$

$$\geq (1-\varepsilon) \overline{H}(x(1-\rho)) \sum_{n=1}^{N} \mathbb{P}(\tau^H \geq n). \tag{B.12}$$

Let $L_\varepsilon(x) = \lim_{N \to \infty} L_{N,\varepsilon}(x)$, then $L_\varepsilon(x) \subset \{B > x\}$ by (B.11), and, for any N, by (3.3) of Theorem 3.1,

$$\mathbb{P}(B > x) - \mathbb{P}(L_\varepsilon(x)) \lesssim \overline{H}(x(1-\rho)) \left(\varepsilon \sum_{n=1}^{N} \mathbb{P}(\tau^H \geq n) + \sum_{n=N+1}^{\infty} \mathbb{P}(\tau^H > n) \right).$$

Choosing N such that $\sum_{n=N+1}^{\infty} \mathbb{P}(\tau^H > n) \leq \varepsilon \mathbb{E}\tau^H$, we get

$$0 \leq \mathbb{P}(B > x) - \mathbb{P}(L_\varepsilon(x)) \lesssim 2\overline{H}(x(1-\rho))\varepsilon \mathbb{E}\tau^H. \tag{B.13}$$

For $x > 0$, define events $J(x)$ and $\overline{J}_\varepsilon(x)$ as

$$J(x) = \bigcup_{n=1}^{\infty} \{\tau^H \geq n, \xi_n^H > x(1-\rho)\},$$

$$\overline{J}_\varepsilon(x) = \bigcup_{n=1}^{\infty} \{\max_{i<n} \xi_i^H \leq x(1-\rho)\} \cap J_{n,\varepsilon}(x).$$

Since $L_\varepsilon(x) = \cup_{n=1}^{\infty} \{\max_{i<n} \xi_i^H \leq x(1-\rho)\} \cap J_{n,\varepsilon}(x) \cap E_{n,\varepsilon}$, we have

$$0 \leq \mathbb{P}(\overline{J}_\varepsilon(x)) - \mathbb{P}(L_\varepsilon(x)) = \sum_{n=1}^{\infty} \mathbb{P}(\tau^H \geq n, \max_{i<n} \xi_i^H$$

$$\leq x(1-\rho)) \overline{F^H}(x(1-\rho))(1 - \mathbb{P}(E_{n,\varepsilon}))$$

$$\leq \varepsilon \mathbb{E}\tau^H \overline{F^H}(x(1-\rho)).$$

Further, $\overline{J}_\varepsilon(x) \subset J(x)$ and

$$0 \leq \mathbb{P}(J(x)) - \mathbb{P}(\overline{J}_\varepsilon(x)) \leq \mathbb{E}\tau^H \mathbb{P}(\xi_1^H \in (x(1-\rho), x(1-\rho) + h_\varepsilon(x)])$$

$$= o(\overline{F^H}(x(1-\rho))).$$

Combining those with (B.13), we obtain that

$$
\begin{aligned}
\mathbb{P}\left(\{B > x\} \setminus J(x)\right) &\le \mathbb{P}(B > x) - \mathbb{P}(L_\varepsilon(x) \cap J(x)) \\
&= \mathbb{P}(B > x) - \mathbb{P}(L_\varepsilon(x)) \le 2\varepsilon \mathbb{E}\tau^H \overline{F^H}(x(1-\rho)),
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}\left(J(x) \setminus \{B > x\}\right) &\le \mathbb{P}\left(J(x)\right) - \mathbb{P}\left(L_\varepsilon(x)\right) \\
&\lesssim \varepsilon \mathbb{E}\tau^H \overline{F^H}(x(1-\rho)) + o(\overline{F^H}(x(1-\rho))).
\end{aligned}
$$

Since $\varepsilon > 0$ may be taken arbitrarily small, we arrive at (3.6).

C Alternative Proof of Corollary 2.1

In this section, we give an alternative proof of Corollary 2.1, which is based on the result from [7], not using PSBJ. Instead of it, our basic tools are Theorem A.1 and the law of large numbers. We also slightly generalise Corollary 2.1.

Theorem C.1 *For the stable $GI/GI/1$ feedback queue, assume that its service time distribution is intermediate regularly varying and has a finite mean. If the first customer arriving at the empty system has an exceptional first service time η instead of $\sigma_{1,0}$, that is, $U_1 = \eta$, such that*

- (I) η *has an intermediate regularly varying distribution with $\mathbb{E}(\eta) < \infty$;*
- (II) $\dfrac{\mathbb{P}(\eta > x)}{\mathbb{P}(\sigma > x)}$ *converges to a constant from $[0, \infty]$, as $x \to \infty$;*
- (III) $\mathbb{E}(X_{k-1}^{(0)}) < \infty$ *for $k \ge 1$;*

then, for each $k \ge 1$, as $x \to \infty$,

$$
\mathbb{P}\left(T_k > x\right) \sim \sum_{\ell=0}^{k-2} \mathbb{E}\left(1 + X_{k-\ell-1}^{(0)}\right) \mathbb{P}\left((1 + m_\ell^{(0)})\sigma > x\right) + \mathbb{P}\left((1 + m_{k-1}^{(0)})\eta > x\right).
$$
(C.14)

Remark C.1 If $\eta = \sigma_{1,0}$, then the conditions (I)–(III) are satisfied, and this theorem is just Corollary 2.1.

Proof Recall the definitions of $X_{\ell-1}$, U_ℓ under $\sigma_{1,0} = \eta$. We have $X_0 = u_0 = 0$, $U_1 = \eta$, and, for $\ell \ge 2$,

$$
X_{\ell-1}^{(0)} = N^e(T_{\ell-2} + U_{\ell-1}) - N^e(T_{\ell-2}) + N_{\ell-2}^B(X_{\ell-2}^{(0)}),
$$
(C.15)

$$
\sum_{j=1}^{\ell-1} X_j^{(0)} = N^e(T_{\ell-1}) + \sum_{j=2}^{\ell-1} N_{j-1}^B(X_{j-1}^{(0)}),
$$
(C.16)

$$
T_\ell = \sum_{j=1}^{\ell} U_j = \sum_{j}^{\ell} \sum_{i=0}^{X_{j-1}^{(0)}} \sigma_{j,i},
$$
(C.17)

where X_{j-1} and $\sigma_{j,i}$ are independent. Since (C.14) is an identity for $k = 1$, we assume that $k \ge 2$. We partition the event $\{T_k > x\}$ into the following k disjoint sets for each

$$y \equiv (y_1, y_2, \ldots, y_{k-1}) > 0.$$

$$I_k^{(\ell)}(y, x) = \begin{cases} \{T_k > x, U_j \le y_j, 1 \le j \le k-1\}, & \ell = 1, \\ \{T_k > x, U_j \le y_j, 1 \le j \le k-\ell, U_{k-\ell+1} > y_{k-\ell+1}\}, & 2 \le \ell \le k. \end{cases}$$

We prove that

$$\mathbb{P}\left(I_k^{(\ell)}(y, x)\right) \sim \begin{cases} \mathbb{E}\left(1 + X_{k-\ell}^{(0)}\right) \mathbb{P}\left((1 + m_{\ell-1}^{(0)})\sigma > x\right), & 1 \le \ell \le k-1, \\ \mathbb{P}\left((1 + m_{k-1}^{(0)})\eta > x\right), & \ell = k, \end{cases} \tag{C.18}$$

as $x \to \infty$ then $y_1, \ldots, y_{k-1} \to \infty$. By the assumptions (I)–(III), these asymptotics yield (C.14), and therefore the theorem is obtained.

We prove (C.18) deriving upper and lower bounds. We first consider the case that $\ell = 1$. Since $U_j \le y_j$ for $1 \le j \le k-1$, we have that $T_j \le \sum_{j'=1}^{j} y_{j'}$ for $1 \le j \le k-1$, and

$$\sum_{j=1}^{\ell-1} X_j^{(0)} \le N^e \left(\sum_{j'=1}^{\ell-1} y_{j'}\right) + \sum_{j=2}^{\ell-1} N_{j-1}^B\left(X_{j-1}^{(0)}\right) \quad \text{on } I_k^{(1)}(y, x).$$

This inductively shows that $X_{k-1}^{(0)}$ has light tail on $I_k^{(\ell)}(y, x)$, and therefore (C.17) implies that

$$\limsup_{x \to \infty} \frac{\mathbb{P}\left(I_k^{(1)}(y, x)\right)}{\mathbb{P}(\sigma > x)} \le \limsup_{x \to \infty} \frac{\mathbb{P}\left(\sum_{i=0}^{X_{k-1}^{(0)}} \sigma_{k,i} > x - \sum_{j=1}^{k-1} y_j\right)}{\mathbb{P}(\sigma > x)} = \mathbb{E}\left(1 + X_{k-1}^{(0)}\right).$$

The corresponding lower bound is obvious. That is,

$$\liminf_{x \to \infty} \frac{\mathbb{P}\left(I_k^{(1)}(y, x)\right)}{\mathbb{P}(\sigma > x)} \ge \liminf_{x \to \infty} \frac{\mathbb{P}\left(\sum_{i=0}^{X_{k-1}^{(0)} \mathbb{1}(U_j < y_j, 1 \le j \le k-1)} \sigma_{k,i} > x\right)}{\mathbb{P}(\sigma > x)}$$

$$= \mathbb{E}\left(\mathbb{1}(U_j < y_j, 1 \le j \le k-1)\left(1 + X_{k-1}^{(0)}\right)\right).$$

Hence, letting $y_j \to \infty$ for $j = 1, 2, \ldots, k$, we obtain (C.18) for $\ell = 1$.

We next consider the case $\ell = 2$. Let c_1 be a positive constant, which will be appropriately determined for each sufficiently small $\varepsilon > 0$.

$$\mathbb{P}\left(I_k^{(2)}(y, x)\right) \le A_+^{(2)}(y_1, \ldots, y_{k-2}, x) + B_+^{(2)}(y_1, \ldots, y_{k-2}, x),$$

where

$$A_+^{(2)}(y_1, \ldots, y_{k-2}, x) = \mathbb{P}\left(U_j \le y_j, 1 \le j \le k-2, U_{k-1} > c_1 x\right),$$

$$B_+^{(2)}(y_1, \ldots, y_{k-2}, x) = \mathbb{P}\left(\sum_{i=0}^{X_{k-1}^{(0)}} \sigma_{k,i} > (1 - c_1)x - (y_1 + \cdots + y_{k-2}),\right.$$

$$\left. U_j \le y_j, 1 \le j \le k-2, y_{k-1} < U_{k-1} \le c_1 x\right).$$

Since $U_{k-1} = \sum_{i=1}^{X_{k-2}^{(0)}} \sigma_{k-1,i}$, the term $A_+^{(2)}(y_1, \ldots, y_{k-2}, x)$ has the same asymptotics as $I_{k-1}^{(1)}(\mathbf{y}, c_1 x)$. Hence, it follows from the asymptotics of $I_k^{(1)}(\mathbf{y}, x)$ that

$$A_+^{(2)}(y_1, \ldots, y_{k-2}, x) \sim \mathbb{E}\left(1 + X_{k-1}^{(0)}\right) 1(k = 2)\mathbb{P}(\eta > c_1 x) + 1(k \geq 3)\mathbb{P}(\sigma > c_1 x).$$
(C.19)

Thus, if we show that $B_+^{(2)}(y_1, \ldots, y_{k-2}, x)$ is asymptotically negligible, then the right-hand side of (C.18) is obtained as an upper bound for $\ell = 2$. To see this, we consider $X_{k-1}^{(0)}$ on the event

$$E_{k-1}^U(\mathbf{y}, x) \equiv \{U_j \leq y_j, 1 \leq j \leq k - 2, y_{k-1} < U_{k-1} \leq c_1 x\},$$

on which we have

$$X_{k-1}^{(0)} = N^e(T_{k-2} + U_{k-1}) - N^e(T_{k-2}) + N_{k-1}^B(X_{k-2}^{(0)}) \lesssim N^e(c_1 x), \qquad x \to \infty.$$

Hence, letting

$$z_{k-2} = y_1 + \cdots + y_{k-2},$$

$$\widetilde{S}_n^\sigma = \sum_{i=1}^n (\sigma_{k,i} - (b + \varepsilon)), \qquad \widetilde{M}_n = \max_{1 \leq j \leq n} \widetilde{S}_j^\sigma, \qquad n \geq 1,$$

and applying Theorem A.1, we have

$$\begin{aligned}
B_+^{(2)}(y_1, \ldots, y_{k-2}, x) &= \mathbb{P}\Big(\widetilde{S}_{X_{k-1}^{(0)}+1}^\sigma + (b + \varepsilon)(X_{k-1}^{(0)} + 1) \\
&\qquad > (1 - c_1)x - z_{k-2}, E_{k-1}^U(\mathbf{y}, x)\Big) \\
&\leq \mathbb{P}\Big(\widetilde{M}_{X_{k-1}^{(0)}+1} + (b + \varepsilon)(X_{k-1}^{(0)} + 1) \\
&\qquad > (1 - c_1)x - z_{k-2}, E_{k-1}^U(\mathbf{y}, x)\Big) \\
&\leq \mathbb{P}\Big(\widetilde{M}_{X_{k-1}^{(0)}+1} > (1 - c_1(1 + (b + \varepsilon)(\lambda + \varepsilon))x, E_{k-1}^U(\mathbf{y}, x)\Big) \\
&\quad + \mathbb{P}\Big(X_{k-1}^{(0)} + 1 > c_1(\lambda + \varepsilon)x - z_{k-2}, E_{k-1}^U(\mathbf{y}, x)\Big) \\
&\lesssim \mathbb{E}\Big((X_{k-1}^{(0)} + 1)1_{E_{k-1}^U(\mathbf{y},x)}\Big) \mathbb{P}(\sigma > (1 - c_1(1 + (b + \varepsilon)(\lambda + \varepsilon))x) \\
&\quad + \mathbb{P}\Big(N^e(c_1 x) > c_1(\lambda + \varepsilon)x - z_{k-2}\Big),
\end{aligned}$$

where the last probability term decays super-exponentially fast, so it is negligible. Thus, if we choose $c_1 > 0$ such that $c_1(1 + (b + \varepsilon)(\lambda + \varepsilon)) < 1$, then $B_+^{(2)}(y_1, \ldots, y_{k-2}, x)$ is asymptotically negligible because

$$\lim_{x \to \infty} \mathbb{E}\left((X_{k-1}^{(0)} + 1)1_{E_{k-1}^U(\mathbf{y},x)}\right) = 0.$$

Consequently, we choose $c_1 = \frac{1-\varepsilon}{1+(b+\varepsilon)(\lambda+\varepsilon)}$, which converges to $(1 + \lambda b)^{-1}$ as $\varepsilon \downarrow 0$. Thus, we have proved that the right-hand side of (C.18) is an upper bound for $\ell = 2$.

For the lower bound for $\ell = 2$, we take another decomposition. Let $d_1 = \frac{1+\varepsilon}{1+\lambda b} < 1$ for a sufficiently small $\varepsilon > 0$, then, for $d_1 x > y_{k-1}$,

$$\mathbb{P}\left(I_k^{(2)}(y_1, \ldots, y_{k-2}, d_1 x, x)\right) \geq \mathbb{P}(U_{k-1} > d_1 x, U_j \leq y_j, 1 \leq j \leq k - 2)$$

$$- \mathbb{P}(T_k \le x, U_j \le y_j, 1 \le j \le k-2, U_{k-1} > d_1 x).$$

Similar to (C.19), we have

$$\mathbb{P}(U_{k-1} > d_1 x, U_j \le y_j, 1 \le j \le k-2)$$
$$\sim \mathbb{E}\left(1 + X_{k-1}^{(0)}\right) 1(k=2)\mathbb{P}(\eta > d_1 x) + 1(k \ge 3)\mathbb{P}(\sigma > d_1 x).$$

On the other hand, by the law of large numbers,

$$\mathbb{P}(T_k \le x, U_j \le y_j, 1 \le j \le k-2, U_{k-1} > d_1 x)$$
$$\le \mathbb{P}\left(\sum_{i=0}^{X_{k-1}^{(0)}} \sigma_{k,i} \le (1 - d_1 x), U_j \le y_j, 1 \le j \le k-2, U_{k-1} > d_1 x\right)$$
$$\le \mathbb{P}\left(\sum_{i=0}^{N^e(z_{k-2}+d_1 x) - N^e(z_{k-2}) + N_{k-1}^B(X_{k-2}^{(0)})} \sigma_{k,i} \le (1 - d_1 x), U_j \le y_j, 1 \le j \le k-2, U_{k-1} > d_1 x\right)$$
$$= o(1)\mathbb{P}\left(U_{k-1} > d_1 x\right).$$

Hence, this term is asymptotically negligible, and therefore we have the asymptotic lower bound for $I_k^{(2)}(\mathbf{y}, x)$, which agrees with the upper bound, by letting $\varepsilon \downarrow 0$. Thus, we have proved (C.18) for $\ell = 2$. For $\ell = 3, \ldots, k$, (C.18) is similarly proved (we omit the details). Then the proof of the corollary is completed.   □

References

1. Asmussen, S.: Applied Probability and Queues, 2nd edn. Springer, Berlin (2003)
2. Baccelli, F., Foss, S.: Moments and tails in monotone-separable stochastic networks. Ann. Appl. Probab. **14**, 612–650 (2004)
3. Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling Extremal Events. Springer, Berlin (1997)
4. Foss, S., Korshunov, D.: On large delays in multi-server queues with heavy tails. Math. Oper. Res. **37**, 201–218 (2012)
5. Foss, S., Miyazawa, M.: Two-node fluid network with a heavy-tailed random input: the strong stability case. J. Appl. Probab. **51A**, 249–265 (2014)
6. Foss, S., Puhalskii, A.: On the limit law of a random walk conditioned to reach a high level. Stoch. Process. Appl. **121**, 288–313 (2011)
7. Foss, S., Zachary, S.: The maximum on a random time interval of a random walk with a long-tailed increments and negative drift. Ann. Appl. Probab. **13**, 37–53 (2003)
8. Foss, S., Palmowski, Z., Zachary, S.: The probability of exceeding a high boundary on a random time interval for a heavy-tailed random walk. Ann. Appl. Probab. **15**, 1936–1957 (2005)
9. Foss, S., Korshunov, D., Zachary, S.: An Introduction to Heavy-Tailed and Subexponential Distributions, 2nd edn. Springer, Berlin (2013)
10. Greenwood, P.: Asymptotics of randomly stopped sequences with independent increments. Ann. Probab. **1**, 317–321 (1973)
11. Jelenkovic, P., Momcilovic, P., Zwart, B.: Reduced load equivalence under subexponentiality. Queueing Syst. **46**, 97–112 (2004)
12. Takacs, L.: A single-server queue with feedback. Bell Syst. Tech. J. **42**, 505–519 (1963)
13. Zwart, A.P.: Queueing systems with heavy tails. PhD Thesis, Technische Universiteit Eindhoven, Eindhoven (2001)