Queueing Syst (2008) 60: 203–226
DOI 10.1007/s11134-008-9094-5

Asymptotic analysis of Lévy-driven tandem queues

Pascal Lieshout · Michel Mandjes

Received: 28 May 2008 / Revised: 29 September 2008 / Published online: 8 November 2008

Abstract We analyze tail asymptotics of a two-node tandem queue with spectrally-positive Lévy input. A first focus lies in the tail probabilities of the type $\mathbb{P}(Q_1 > \alpha x, Q_2 > (1-\alpha)x)$, for $\alpha \in (0,1)$ and x large, and Q_i denoting the steady-state workload in the ith queue. In case of light-tailed input, our analysis heavily uses the joint Laplace transform of the stationary buffer contents of the first and second queue; the logarithmic asymptotics can be expressed as the solution to a convex programming problem. In case of heavy-tailed input we rely on sample-path methods to derive the exact asymptotics. Then we specialize in the tail asymptotics of the downstream queue, again in case of both light-tailed and heavy-tailed Lévy inputs. It is also indicated how the results can be extended to tandem queues with more than two nodes.

Keywords Lévy processes · Queueing asymptotics · Rare events · Large deviations

Mathematics Subject Classification (2000) 60K25

Part of this work was done while M.M. was at Stanford University, Stanford, CA 94305, USA.
This research has been funded by the Dutch BSIK/BRICKS (Basic Research in Informatics for Creating the Knowledge Society) project.

P. Lieshout · M. Mandjes (✉)
Korteweg-de Vries Institute for Mathematics, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands
e-mail: m.r.h.mandjes@uva.nl

P. Lieshout
e-mail: lieshout@cwi.nl

P. Lieshout · M. Mandjes
CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

M. Mandjes
Eurandom, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

1 Introduction

Recently, substantial progress has been made in the analysis of Lévy-driven queueing networks. For the important case of spectrally-positive Lévy input (that is, the driving Lévy process does not have negative jumps), the joint Laplace transform of the stationary buffer contents has been found for a broad class of network structures, including tandem queues [18, 19, 29]. With the transforms being available, one may attempt to use these in order to explicitly find the joint distribution of the stationary buffer contents. So far this turned out to be possible in just a few cases, see for instance [4, 31] in the two-node tandem case.

To circumvent the problem of explicit inversion of the joint Laplace transform, one may settle for finding the *joint asymptotics* of both queues, that is, characterize for $\alpha \in (0, 1)$ and x large

$$\pi_\alpha(x) := \mathbb{P}\big(Q_1 > \alpha x, Q_2 > (1-\alpha)x\big),$$

where Q_i denotes the steady-state workload in the ith queue of the two-node tandem. In the one-dimensional setting, it is well understood how Laplace transforms reveal the tail asymptotics of the underlying random variable (for instance by application of Tauberian theorems), but considerably less attention has been paid to developing such relations in a multivariate context. An inherent problem is that there are hardly techniques available to derive the joint asymptotics $\pi_\alpha(x)$ from the bivariate Laplace transform $\mathbb{E}\exp(-sQ_1 - tQ_2)$; more specifically, a significant difficulty at the technical level is that it is not clear how the Laplace transform of $\pi_\alpha(x)$, i.e.,

$$\int_0^\infty e^{-rx}\pi_\alpha(x)\,\mathrm{d}x,$$

can be found from $\mathbb{E}\exp(-sQ_1 - tQ_2)$.

This paper is devoted to finding the joint asymptotics, as introduced above, for two-node tandem networks with spectrally-positive Lévy input. A second goal of the paper is to derive tail asymptotics of the downstream queue, thus complementing earlier results [19]. In more detail, the contributions are the following.

- For the case of light-tailed Lévy input we derive the *logarithmic asymptotics* of $\pi_\alpha(x)$:

$$\lim_{x\to\infty}\frac{1}{x}\log\pi_\alpha(x).$$

  The proof is along the following lines. Relying on the classical Chernoff bound, we find an upper bound to this decay rate in the form of the solution of a convex programming problem. Relying on sample-path large deviations for Lévy processes [2], it is shown that this upper bound is actually tight. To this end, we construct a trajectory whose rate function coincides with the solution of the above-mentioned convex programming problem; as this trajectory is 'feasible' (in that it is such that indeed queue 1 exceeds αx and queue 2 exceeds $(1-\alpha)x$), this yields the desired result. The solution has three different shapes, as could be expected in view of e.g. [14, 28]. Our results on the light-tailed case can be found in Sect. 3.1.

- In the case of heavy-tailed Lévy input the above line of reasoning does not apply. The rare event is typically the result of just a single big jump, rather than a sequence of somewhat unlikely outcomes. This idea leads to a procedure that provides us the exact asymptotics of $\pi_\alpha(x)$ in the heavy-tailed case: in Sect. 3.2, a function $f_\alpha(x)$ is presented such that $\pi_\alpha(x)/f_\alpha(x) \to 1$ as $x \to \infty$, in the sequel denoted by $\pi_\alpha(x) \sim f_\alpha(x)$. The proof consists of a lower bound that identifies a most likely scenario, and an upper bound that shows that all other scenarios lead to asymptotically negligible contributions; the line of reasoning resembles that of earlier papers, e.g. [7, 44].
- In Sect. 4.1, exact tail asymptotics for the workload of the downstream queue are given for the case of light-tailed Lévy input, generalizing earlier results in [19, Sect. 4]. Interestingly, multiple regimes are identified: one in which the first queue hardly affects the tail asymptotics of the downstream queue (corresponding to relatively large values of the service rate of the first queue), and one in which the first queue does play an explicit role in delaying and reshaping the traffic before entering the second queue (corresponding to relatively small values of the service rate of the first queue).
- Section 4.2 generalizes [19, Sect. 5] by presenting the exact tail asymptotics of the downstream queue for heavy-tailed Lévy input by one single theorem that covers both the compound Poisson case and the α-stable case. The analysis relies on the application of Tauberian theorems.
- We finish the paper by indicating in Sect. 5 how our results generalize to a multi-link setting. We also identify a number of directions for future research.

We finish this introduction by mentioning a number of relevant related results from the literature, which perhaps started off with the pioneering work of Dobrushin and Pecherskii, see e.g. [36]. In [40] the tail probabilities of the *total* network population are studied for a series of M/M/1 queues; cf. also [27]. [26] considers a series of queues with exponential service times, and uses sample-path large deviations to characterize the queueing asymptotics. An early reference on the concept of effective bandwidths, and particularly those of departure processes, is [15]; related results are presented in [9, 10, 19]. Results for sojourn times in the heavy-tailed case are given in [6], and see also [23] for recent results on a specific class of light-tailed distributions.

2 Model and preliminaries

In this paper we consider a two-node tandem queue, where the first (second) node has constant service capacity c_1 (c_2). With $A(\cdot) = \{A(t), t \in \mathbb{R}\}$ we associate the input process of the tandem, which is assumed to be a *Lévy process*, that is, a process with stationary independent increments. In case $t \geq 0$, $A(t)$ denotes the amount of traffic entering the system in the interval $(0, t]$, whereas for $t < 0$ we follow the convention that $-A(-t, 0)$ corresponds to the traffic generated in $(t, 0]$. Also, let $A(s, t) = A(t) - A(s)$ denote the amount of traffic generated in the interval $(s, t]$. In the remainder of this paper we focus on an important subclass of Lévy processes, viz. *spectrally positive* Lévy processes, that is, Lévy processes which do not have negative jumps. This class covers Brownian motion and compound Poisson input as important special cases.

Despite the fact that the input process is not necessarily increasing (for instance in the case of Brownian input), we can define a workload process; for the first queue of the tandem system the workload at time t is given through

$$Q_1(t) = \sup_{0\le s\le t}\big(A(s,t) - c_1(t-s)\big), \quad t\ge 0,$$

given that $Q_1(0)=0$. Bearing in mind that the *total* queue behaves as a single queue, fed by $A(\cdot)$, and emptied at rate c_2, we also have

$$Q_1(t)+Q_2(t) = \sup_{0\le s\le t}\big(A(s,t) - c_2(t-s)\big), \quad t\ge 0,$$

assuming that $Q_1(0)=Q_2(0)=0$. Then we can recover the workload at queue 2 by subtracting $Q_1(t)$ from $Q_1(t)+Q_2(t)$. With Q_1 (Q_2, respectively) we denote the stationary version of $Q_1(t)$ ($Q_2(t)$). Define $\mu := \mathbb{E}A(1) > 0$, and assume that both service rates (i.e., c_1 and c_2) are larger than μ to ensure stability. Here we assume that $c_1 > c_2$ to avoid the trivial situation that the second queue is always empty.

Spectrally-positive Lévy processes are uniquely given through their Laplace exponent $\kappa(\cdot)$:

$$\mathbb{E}e^{-sA(t)} = e^{t\kappa(s)}, \quad s\ge 0.$$

If $\kappa(s)$ also exists for negative s, then the Lévy process could be called light-tailed, as the tail of the distribution of $A(1)$ decays exponentially or faster. If $\kappa(s)$ is only defined for non-negative s, then the process could be called heavy-tailed, as the tail of the distribution of $A(1)$ tends to decay more slowly than any exponential. Important examples of spectrally-positive Lévy processes are the following. (1) *Brownian motion with drift*. We write $A\in\mathbb{B}\mathrm{m}(\mu,\sigma^2)$ when $\kappa(s) = -s\mu + \frac{1}{2}s^2\sigma^2$. (2) *Compound Poisson*. Jobs arrive according to a Poisson process of rate λ; the jobs $B_1, B_2, \ldots$ are i.i.d. samples from a distribution with Laplace transform $\beta(s) := \mathbb{E}e^{-sB}$. We write $A\in\mathbb{C}\mathrm{P}(\lambda,\beta(\cdot))$; it can be verified that $\kappa(s) = -\lambda + \lambda\beta(s)$.

We now recapitulate a number of results on the (joint) distribution of Q_1 and Q_2. The Laplace transform of Q_1 dates back to, at least, Zolotarev [42]: with $\vartheta(s) := \kappa(s) + c_1 s$, the so-called *generalized Pollaczek–Khinchine formula* states that, for $s\ge 0$,

$$\mathbb{E}e^{-sQ_1} = \frac{\vartheta'(0)s}{\vartheta(s)} = \frac{(c_1-\mu)s}{\vartheta(s)};$$

this is a generalization of the classical result for compound Poisson inputs. This result for a single queue has been extended more recently to the network setting [18, 29]; it was found that for $(s,t)\in\mathbb{R}_+^2$,

$$\mathbb{E}e^{-sQ_1-tQ_2} = \frac{(c_2-\mu)t}{t-\vartheta^{-1}((c_1-c_2)t)} \times \frac{\vartheta^{-1}((c_1-c_2)t) - s}{(c_1-c_2)t - \vartheta(s)}. \tag{1}$$

By plugging in $s=0$, one retrieves the Laplace transform of the downstream queue Q_2 [19]: for all $t\geq 0$,

$$\mathbb{E}e^{-tQ_2}=\frac{c_2-\mu}{c_1-c_2}\frac{\vartheta^{-1}(t(c_1-c_2))}{t-\vartheta^{-1}(t(c_1-c_2))}. \tag{2}$$

3 Joint asymptotics

In this section we consider *joint asymptotics*, that is, for a given $\alpha\in(0,1)$,

$$\pi_\alpha(x):=\mathbb{P}\big(Q_1>\alpha x, Q_2>(1-\alpha)x\big),$$

for x large. Section 3.1 treats the case that $A(\cdot)$ corresponds to a light-tailed input, whereas Sect. 3.2 deals with the heavy-tailed case.

3.1 Light-tailed input

In this subsection we derive the logarithmic asymptotics of $\pi_\alpha(x)$ for a light-tailed input. We do this by first finding an upper bound on the corresponding exponential decay rate, and then applying sample-path large deviations to prove that this upper bound is actually tight.

The analysis of the upper bound is based on the joint Laplace transform $\mathbb{E}e^{-sQ_1-tQ_2}$, as given in (1), for $(s,t)\in\mathbb{R}_+^2$. In the light-tailed case, however, this expression is valid for some $(s,t)\notin\mathbb{R}_+^2$ as well. As we will argue below, these (s,t) provide us with the crucial information to identify an upper bound on the decay rate.

Let

$$F:=\big\{(s,t)\in\mathbb{R}^2:\mathbb{E}e^{-sQ_1-tQ_2}<\infty\big\},$$

and $\bar{F}:=F\cap\mathbb{R}_-^2$.

Lemma 3.1 *F and $\bar{F}$ are convex.*

Proof Take (s_1,t_1) and (s_2,t_2) in F. Take a $\lambda\in(0,1)$. Then

$$\begin{aligned}&\mathbb{E}\exp\big(-\big(\lambda s_1+(1-\lambda)s_2\big)Q_1-\big(\lambda t_1+(1-\lambda)t_2\big)Q_2\big)\\&\quad\leq\lambda\mathbb{E}e^{-s_1Q_1-t_1Q_2}+(1-\lambda)\mathbb{E}e^{-s_2Q_1-t_2Q_2}<\infty,\end{aligned}$$

due to straightforward convexity arguments. The statement on $\bar{F}$ follows immediately. □

Proposition 3.2 *The following logarithmic asymptotic upper bound applies*:

$$\limsup_{x\to\infty}\frac{1}{x}\log\pi_\alpha(x)\leq\min_{s,t\in\bar{F}}\big(\alpha s+(1-\alpha)t\big).$$

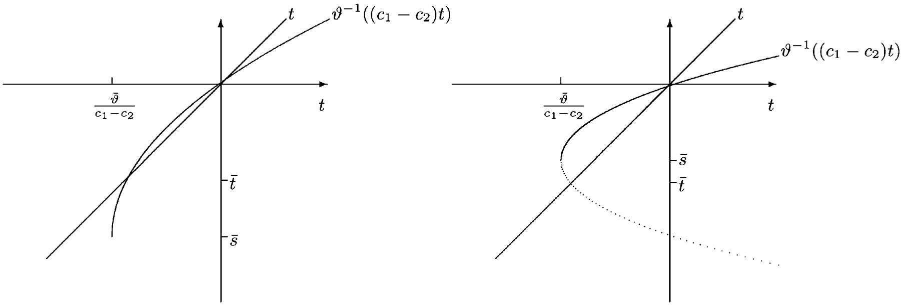

Fig. 1 *Left* picture: $\bar{s} < \bar{t}$; *right* picture: $\bar{s} \geq \bar{t}$

Proof Due to the Chernoff bound, we have for all $(s, t) \in \bar{F}$:

$$\mathbb{P}\big(Q_1 > \alpha x, Q_2 > (1-\alpha)x\big) \leq \mathbb{E}e^{-sQ_1 - tQ_2} e^{\alpha s x + (1-\alpha)tx}.$$

The stated follows by taking logs of both sides, dividing by x, and tending x to ∞. □

We conclude that finding an upper bound on the decay rate reduces to a *convex programming problem*, by virtue of Lemma 3.1 and Proposition 3.2. These have attractive numerical properties.

Let us now analyze $\bar{F}$ in greater detail.

- Observe that, trivially, $(c_2 - \mu)t$ is negative for any negative t.
- We then wonder when $t - \vartheta^{-1}((c_1 - c_2)t)$ is negative. Realize that $\vartheta(\cdot)$ is not bijective, and hence $\vartheta^{-1}((c_1 - c_2)t)$ is not always well-defined. Let $\bar{\vartheta}$ be $\inf_s \vartheta(s)$, and $\bar{s}$ the minimizing argument; from $\mu < c_1$ it follows immediately that both $\bar{\vartheta}$ and $\bar{s}$ are negative. We have to require that t be larger than $\bar{\vartheta}/(c_1 - c_2)$.

  Let $\bar{t}$ be the (non-zero) root of $\vartheta(t) = (c_1 - c_2)t$ (or, equivalently, $\kappa(t) = -c_2 t$; note that $\bar{t}$ does not depend on c_1). From the assumption $c_1 > c_2 > \mu$ it follows that this root is necessarily negative. It is easily checked that there are now two possibilities: (1) if $\bar{s} < \bar{t}$, then $t - \vartheta^{-1}((c_1 - c_2)t)$ is negative for all $t \in (\bar{t}, 0)$; (2) if $\bar{s} \geq \bar{t}$, then $t - \vartheta^{-1}((c_1 - c_2)t)$ is negative for all $t \in (\bar{\vartheta}/(c_1 - c_2), 0)$. The two cases are illustrated in Fig. 1.
- Now consider the factor

$$\frac{\vartheta^{-1}((c_1 - c_2)t) - s}{(c_1 - c_2)t - \vartheta(s)}; \tag{3}$$

  we again have to impose that $t \geq \bar{\vartheta}/(c_1 - c_2)$. It is readily verified that for $s \geq \bar{s}$ the numerator and denominator of (3) are either both negative, or both positive. For $s < \bar{s}$ the numerator is positive for $t \in (\bar{\vartheta}/(c_1 - c_2), 0)$, whereas the denominator is positive for $t \geq \vartheta(s)/(c_1 - c_2)$; using the definition of $\bar{\vartheta}$, conclude that the ratio is positive for $t \geq \vartheta(s)/(c_1 - c_2)$.

These observations are summarized by the following lemma; a graphical representation is given in Fig. 2.

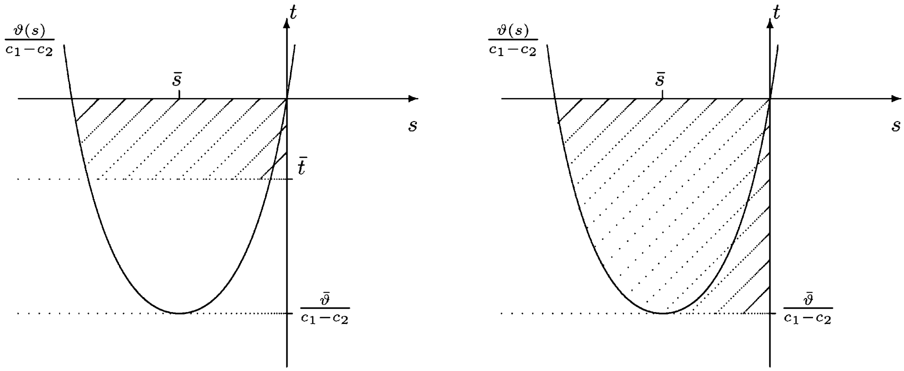

Fig. 2 The *shaded area* is $\bar{F}$. *Left* picture: $\bar{s} < \bar{t}$; *right* picture: $\bar{s} \geq \bar{t}$

Lemma 3.3 *$\bar{F}$ is described by the following convex region*:

$$\left\{(s,t) \in \mathbb{R}^2_- : s < \bar{s}, t > \max\left\{\bar{t}, \frac{\vartheta(s)}{c_1 - c_2}\right\}\right\} \cup \left\{(s,t) \in \mathbb{R}^2_- : s \geq \bar{s}, t > \bar{t}\right\}$$

in case $\bar{s} < \bar{t}$, and

$$\left\{(s,t) \in \mathbb{R}^2_- : s < \bar{s}, t > \frac{\vartheta(s)}{c_1 - c_2}\right\} \cup \left\{(s,t) \in \mathbb{R}^2_- : s \geq \bar{s}, t > \frac{\bar{\vartheta}}{c_1 - c_2}\right\}$$

in case $\bar{s} \geq \bar{t}$.

In case $\bar{s} < \bar{t}$, it now follows that there are three possible solutions to the convex programming problem. In the first place, the minimum can be attained at $(s_-, 0)$, where s_- solves $\vartheta(s) = 0$. In the second place, the minimum can be attained at $(s, (c_1 - c_2)^{-1}\vartheta(s))$ for $s \in (s_-, s_+)$; here $s_+ < 0$ is the smaller solution to $\vartheta(s)/(c_1 - c_2) = \bar{t}$, which is smaller than $\bar{s}$ (we remark that the larger solution to this equation is $\bar{t} > \bar{s}$). Finally, the minimum can be attained at $(s_+, \bar{t})$. It is immediate that the first solution comes out if

$$\frac{\vartheta'(s_-)}{c_1 - c_2} > -\frac{\alpha}{1 - \alpha}, \tag{4}$$

the third solution if

$$\frac{\vartheta'(s_+)}{c_1 - c_2} < -\frac{\alpha}{1 - \alpha}, \tag{5}$$

and otherwise the second solution.

In case $\bar{s} \geq \bar{t}$ the third solution cannot occur: we obtain the first solution if (4) applies, and otherwise the second solution.

We have arrived at the following result. Define

$$\alpha_+ := -\frac{\vartheta'(s_-)}{c_1 - c_2 - \vartheta'(s_-)}, \qquad \alpha_- := -\frac{\vartheta'(s_+)}{c_1 - c_2 - \vartheta'(s_+)}.$$

Also, let $s(\alpha)$ be the (unique, as $\vartheta'(\cdot)$ is monotone) solution to $\vartheta'(s) = -(c_1 - c_2)\alpha/(1-\alpha)$; $t(\alpha)$ is defined as $\vartheta(s(\alpha))/(c_1 - c_2)$.

Proposition 3.4 *If* $\alpha < \alpha_-$, *then*

$$\limsup_{x\to\infty} \frac{1}{x} \log \pi_\alpha(x) \le \alpha s_+ + (1-\alpha)\bar{t};$$

if $\alpha_- < \alpha < \alpha_+$, *then*

$$\limsup_{x\to\infty} \frac{1}{x} \log \pi_\alpha(x) \le \alpha s(\alpha) + (1-\alpha) t(\alpha);$$

if $\alpha > \alpha_+$, *then*

$$\limsup_{x\to\infty} \frac{1}{x} \log \pi_\alpha(x) \le \alpha s_-.$$

We conclude from the above that straightforward Chernoff-bound arguments lead to an upper bound for the decay rate. We now address the question whether the identified decay rate is actually tight. We do so by relying on sample-path large deviations.

It is well known that the steady-state queue length of the first queue, that is, Q_1, is distributed as $\sup_{\tau>0}(A(-\tau,0) - c_1\tau)$, cf. [37], whereas the total queue, $Q_1 + Q_2$ is distributed as a single queue with service rate c_2, i.e., $\sup_{\sigma>0}(A(-\sigma,0) - c_2\sigma)$, cf. [3, 22, 33, 39]. Representing the downstream queue as the difference between the total queue and the first queue, it is evident that $\pi_\alpha(x)$ can be rewritten as

$$\mathbb{P}\left(\begin{matrix}\sup_{\tau>0}\big(A(-\tau,0) - c_1\tau\big) > \alpha x,\\ \sup_{\sigma>0}\big(A(-\sigma,0) - c_2\sigma\big) - \sup_{\tau>0}\big(A(-\tau,0) - c_1\tau\big) > (1-\alpha)x\end{matrix}\right).$$

In the latter event (that is, the event concerning the downstream queue), the optimizing τ can be proven to be smaller than the optimizing σ, see [33, Lemma 2.4]. Also performing a time scaling $\tau \to \tau x$ and $\sigma \to \sigma x$, we obtain

$$\pi_\alpha(x) = \mathbb{P}\left(\begin{matrix}\exists\tau > 0 : x^{-1}\cdot A(-\tau x, 0) > c_1\tau + \alpha,\\ \exists\sigma > 0 : \forall\tau \in (0,\sigma) : x^{-1}\cdot A(-\sigma x, -\tau x) > c_2\sigma - c_1\tau + (1-\alpha)\end{matrix}\right),$$

which we can bound from below by, for any $T > 0$,

$$\mathbb{P}\left(\begin{matrix}\exists\tau \in (0,T) : x^{-1}\cdot A(-\tau x, 0) > c_1\tau + \alpha,\\ \exists\sigma \in (0,T) : \forall\tau \in (0,\sigma) : x^{-1}\cdot A(-\sigma x, -\tau x) > c_2\sigma - c_1\tau + (1-\alpha)\end{matrix}\right);$$

below, in Remark 3.6, we will select an appropriate value for T.

Now consider the set of paths corresponding to the event in the previous display:

$$\mathcal{A} := \left\{ f : \begin{matrix}\exists\tau \in (0,T) : -f(-\tau) > c_1\tau + \alpha,\\ \exists\sigma \in (0,T) : \forall\tau \in (0,\sigma) : f(-\tau) - f(-\sigma) > c_2\sigma - c_1\tau + (1-\alpha)\end{matrix}\right\}.$$

It is readily seen that, under the supremum metric, $\mathscr{A}$ is open. Now Theorem 5.1 of de Acosta [2] can be applied to obtain

$$\liminf_{x\to\infty} \frac{1}{x} \log \pi_\alpha(x) \geq -\mathbb{I}(f), \tag{6}$$

for any $f \in \mathscr{A}$, where

$$\mathbb{I}(f) := \int_{-T}^{0} \sup_s \big(sf'(\tau) - \kappa(-s)\big)\, \mathrm{d}\tau.$$

We also introduce

$$\mathscr{A}^\star := \left\{ f : \begin{array}{l} \exists\tau \in (0,T) : -f(-\tau) \geq c_1\tau + \alpha, \\ \exists\sigma \in (0,T) : \forall\tau \in (0,\sigma) : f(-\tau) - f(-\sigma) \geq c_2\sigma - c_1\tau + (1-\alpha) \end{array} \right\}.$$

Proposition 3.5 *If $\alpha < \alpha_-$, then*

$$\liminf_{x\to\infty} \frac{1}{x} \log \pi_\alpha(x) \geq \alpha s_+ + (1-\alpha)\bar{t};$$

if $\alpha_- < \alpha < \alpha_+$, then

$$\liminf_{x\to\infty} \frac{1}{x} \log \pi_\alpha(x) \geq \alpha s(\alpha) + (1-\alpha)t(\alpha);$$

if $\alpha > \alpha_+$, then

$$\liminf_{x\to\infty} \frac{1}{x} \log \pi_\alpha(x) \geq \alpha s_-.$$

Proof First consider the path f_+ given through $f_+(\tau) = -\kappa'(s_-)\tau$ for $\tau \in (-\tau_+, 0)$, with τ_+ defined by $-\alpha/(\kappa'(s_-) + c_1) = -\alpha/\vartheta'(s_-)$; $f_+(\tau) = \mu\tau$ for $\tau \geq 0$ and $f_+(\tau) = \kappa'(s_-)\tau_+ + \mu(\tau + \tau_+)$ for $\tau \leq -\tau_+$. Recall that $-\kappa'(s_-)$ is a positive number, larger than c_1. Let us check under which conditions this path lies in $\mathscr{A}^\star$. It is readily checked that

$$\begin{aligned} f_+(-\tau_+) &= f_+\left(\frac{\alpha}{\kappa'(s_-) + c_1}\right) = -\frac{\alpha\kappa'(s_-)}{\kappa'(s_-) + c_1} = -\left(-\frac{\alpha c_1}{\kappa'(s_-) + c_1} + \alpha\right) \\ &= -c_1\tau_+ - \alpha, \end{aligned}$$

so the path f_+ is such that queue 1 attains the value α at time 0. Observing that the total queue attains value

$$\frac{-\kappa'(s_-) - c_2}{-\kappa'(s_-) - c_1} \cdot \alpha, \tag{7}$$

it is concluded that the path is in $\mathscr{A}^\star$ if (7) exceeds $\alpha + (1-\alpha) = 1$, a condition which reduces to $\alpha \geq \alpha_+$. However, (6) required that the path f_+ lies in $\mathscr{A}$ rather than $\mathscr{A}^\star$, but clearly f_+ can be approximated arbitrarily closely by a path in $\mathcal{A}$; this reasoning is standard and omitted. Also, realizing that s_- solves $\kappa(s) = -c_1 s$,

$$\begin{aligned}\mathbb{I}(f) &= \int_{-\tau_+}^{0} \sup_s \bigl(-s\kappa'(s_-) - \kappa(-s)\bigr)\,\mathrm{d}\tau \\ &= -\frac{\alpha}{\kappa'(s_-) + c_1} \sup_s \bigl(-s\kappa'(s_-) - \kappa(-s)\bigr) = -\alpha s_-.\end{aligned}$$

This proves the lower bound for $\alpha \geq \alpha_+$.

Now consider the path that generates traffic in the following way: f_0 is given through $f_0(\tau) = r\tau$ for $\tau \in (-\tau_0, 0)$, with

$$r := \frac{\alpha}{1-\alpha}(c_1 - c_2) + c_1,$$

and $\tau_0 := (1-\alpha)/(c_1 - c_2)$; $f_0(\tau) = \mu\tau$ for $\tau \geq 0$ and $f_0(\tau) = -r\tau_0 + \mu(\tau + \tau_0)$ for $\tau \leq -\tau_0$. It is observed that this path is such that the first queue has content $(r - c_1)\tau_0 = \alpha$ at time 0, while the total queue has content $(r - c_2)\tau_0 = 1$ at time 0 (both statements, irrespective of the value of $\alpha \in (0, 1)$); in other words, the path f_0 lies in $\mathcal{A}^\star$. Hence we have, *for all* $\alpha \in (0, 1)$,

$$\liminf_{x\to\infty} \frac{1}{x} \log \pi_\alpha(x) \geq -\int_{-\tau_0}^{0} \sup_s \bigl(sr - \kappa(-s)\bigr)\,\mathrm{d}\tau.$$

After elementary calculus, we obtain that the right-hand side of the previous display equals $\alpha s(\alpha) + (1-\alpha)t(\alpha)$.

Finally, consider the piecewise linear path f_- with slope $-\kappa'(\bar{t})$ in $(-\tau_- - \tau_-', -\tau_-)$, with slope $-\kappa'(s_+)$ in $(-\tau_-, 0)$, and slope μ elsewhere; here

$$\tau_-' := \frac{1}{-\kappa'(\bar{t}) - c_2}\left(1 - \alpha - \frac{\alpha}{-\vartheta'(s_+)}(c_1 - c_2)\right); \qquad \tau_- := -\frac{\alpha}{\vartheta'(s_+)}.$$

It is seen that τ_-' is non-negative (and hence the path is well-defined) for all $\alpha < \alpha_-$. The content of queue 1 at time 0 is, due to $-\kappa'(\bar{t}) < c_1$, equal to $(-\kappa'(s_+) - c_1)\tau_- = \alpha$. The content of the total queue at time 0 is, due to $-\kappa'(\bar{t}) \geq c_2$,

$$\bigl(-\kappa'(\bar{t}) - c_2\bigr)\tau_-' + \bigl(-\kappa'(s_+) - c_2\bigr)\tau_- = 1,$$

and hence the path is in $\mathcal{A}^\star$. It is readily checked that, using $\kappa(\bar{t}) = -c_2\bar{t}$ and $\vartheta(s_+)/(c_1 - c_2) = \bar{t}$,

$$\begin{aligned}\mathbb{I}(f_-) &= \int_{-\tau_- - \tau_-'}^{\tau_-} \sup_s \bigl(-s\kappa'(\bar{t}) - \kappa(-s)\bigr)\,\mathrm{d}\tau + \int_{-\tau_-}^{0} \sup_s \bigl(-s\kappa'(s_+) - \kappa(-s)\bigr)\,\mathrm{d}\tau \\ &= \alpha\left(\frac{c_2 - c_1}{-\vartheta'(s_+)} \cdot \frac{\bar{t}\kappa'(\bar{t}) - \kappa(\bar{t})}{-\kappa'(\bar{t}) - c_2} + \frac{s_+\kappa'(s_+) - \kappa(s_+)}{-\vartheta'(s_+)}\right) + (1-\alpha)\frac{\bar{t}\kappa'(\bar{t}) - \kappa(\bar{t})}{-\kappa'(\bar{t}) - c_2} \\ &= -\alpha s_+ - (1-\alpha)\bar{t}.\end{aligned}$$

This proves the stated. We remark that, in the Brownian case, a pictorial illustration of the paths to overflow is given in [31, Fig. 3]; the paths in the non-Brownian case look similar. □

Remark 3.6 Above we mentioned that we had to select an appropriate value of the 'time horizon' T. From the proof of Proposition 3.5 it is clear that any T larger than $\max\{\tau_+, \tau_0, \tau_- + \tau'_-\}$ can be chosen.

The following theorem is a direct consequence of Propositions 3.4 and 3.5.

Theorem 3.7 *If $\alpha < \alpha_-$, then*

$$\lim_{x\to\infty} \frac{1}{x} \log \pi_\alpha(x) = \alpha s_+ + (1-\alpha)\bar{t};$$

if $\alpha_- < \alpha < \alpha_+$, then

$$\lim_{x\to\infty} \frac{1}{x} \log \pi_\alpha(x) = \alpha s(\alpha) + (1-\alpha)t(\alpha);$$

if $\alpha > \alpha_+$, then

$$\lim_{x\to\infty} \frac{1}{x} \log \pi_\alpha(x) = \alpha s_-.$$

Example 3.8 Let $A \in \mathbb{B}\mathrm{m}(0,1)$; below we explain how to translate the results to the case $A \in \mathbb{B}\mathrm{m}(\mu, \sigma^2)$ for any $\mu < c_2$ and $\sigma^2 > 0$. Hence $\vartheta(s) = \frac{1}{2}s^2 + c_1 s$, and $\bar{s} = -c_1$, $\bar{\vartheta} = -\frac{1}{2}c_1^2$, and $\bar{t} = -2c_2$. The condition $\bar{s} < \bar{t}$ is equivalent to $c_1 > 2c_2$. Also, $s_- = -2c_1$ and $s_+ = 2(c_2 - c_1)$ (which is smaller than $\bar{s}$ under $\bar{s} < \bar{t}$).

Hence we obtain from (4) that for α larger than $\alpha_+ := c_1/(2c_1 - c_2)$ the point $(s_-, 0)$ is optimal, with decay rate $\alpha s_- = -2\alpha c_1$. For $\alpha < \alpha_- := (c_1 - 2c_2)/(2c_1 - 3c_2)$ the point $(s_+, \bar{t})$ is optimal, with decay rate $-2\alpha(c_1 - c_2) - 2(1-\alpha)c_2$. For α in between α_- and α_+, the optimum is reached in $(s, (c_1 - c_2)^{-1}\vartheta(s))$ for $s \in (s_-, s_+)$; the optimal s equals $(\alpha c_2 - c_1)/(1-\alpha)$. Tedious calculations yield that the corresponding t equals

$$-\frac{(c_1 - \alpha c_2)(\alpha c_2 + (1-2\alpha)c_1)}{2(1-\alpha)(c_1 - c_2)},$$

yielding the decay rate

$$-\frac{1}{2}\frac{(c_1 - \alpha c_2)^2}{(1-\alpha)(c_1 - c_2)}.$$

Recall that in case $c_1 < 2c_2$ it is seen that $(s_+, \bar{t})$ cannot be optimal.

The above results can easily be extended to general Brownian input, i.e., for general $\mu < c_2$ and $\sigma^2 > 0$. It can be verified that in order to generalize the results one simply has to set $x \leftarrow x/\sigma$ and $c_i \leftarrow (c_i - \mu)/\sigma$, $i = 1, 2$, in the above asymptotics.

Example 3.9 Let us now consider the compound Poisson case with exponential jobs, i.e., $A \in \mathbb{CP}(\lambda, \beta(\cdot))$, with $\beta(s) = \nu/(\nu+s)$. It can be verified that $\kappa(s) = \lambda(\beta(s)-1)$. To ensure stability, we assume that $\lambda\nu^{-1} < c_2 < c_1$. We find

$$\kappa(s) = -\frac{s\lambda}{\nu+s}; \qquad \vartheta(s) = -\frac{s\lambda}{\nu+s} + c_1 s; \qquad \vartheta'(s) = -\frac{\lambda\nu}{(\nu+s)^2} + c_1.$$

Furthermore,

$$\bar{s} = \sqrt{\frac{\lambda\nu}{c_1}} - \nu; \qquad \bar{\vartheta} = -\left(\sqrt{c_1\nu} - \sqrt{\lambda}\right)^2; \qquad \bar{t} = \frac{\lambda}{c_2} - \nu.$$

It is easily verified that the condition $\bar{s} < \bar{t}$ is equivalent to $c_1 > (\nu/\lambda) \cdot c_2^2$, and that $s_- = \lambda/c_1 - \nu$ and $s_+ = -\nu(c_1 - c_2)/c_1$, where s_+ is smaller than $\bar{s}$ under $\bar{s} < \bar{t}$. We next deduce that

$$\alpha_+ = \frac{c_1^2\nu - c_1\lambda}{c_1^2\nu - c_2\lambda}; \qquad \alpha_- = \frac{c_1^2\lambda - c_1c_2^2\nu}{c_1^2\lambda - c_2^3\nu}.$$

Using Theorem 3.7, we find the following decay rates.

- For $\alpha < \alpha_-$:

$$\lambda\frac{1-\alpha}{c_2} + \nu\frac{\alpha c_2 - c_1}{c_1}.$$

- For $\alpha \in (\alpha_-, \alpha_+)$:

$$-\frac{(-\nu\sqrt{c-\alpha c_2} + \sqrt{(1-\alpha)\lambda\nu})((1-\alpha)\lambda\sqrt{c_1-\alpha c_2} - (c_1-\alpha c_2)\sqrt{(1-\alpha)\lambda\nu})}{(c_1-c_2)\sqrt{c_1-\alpha c_2}\sqrt{(1-\alpha)\lambda\nu}};$$

  this requires straightforward, though tedious, calculus, and uses

$$s(\alpha) = \frac{\nu(\alpha c_2 - c_1) + \sqrt{(1-\alpha)(c_1-\alpha c_2)\lambda\nu}}{c_1 - \alpha c_2},$$

$$t(\alpha) = -\frac{c_1\nu((\alpha-2)\lambda + \sqrt{(1-\alpha)(c_1-\alpha c_2)\lambda\nu}) - \lambda(\alpha c_2\nu + \sqrt{(1-\alpha)(c_1-\alpha c_2)\lambda\nu})}{(c_1-c_2)\sqrt{(1-\alpha)(c_1-\alpha c_2)\lambda\nu}}.$$

- For $\alpha > \alpha_+$:

$$\alpha\left(\frac{\lambda}{c_1} - \nu\right).$$

3.2 Heavy-tailed input

In this subsection we identify the exact asymptotics of $\pi_\alpha(x)$ in the case of $\mathbb{CP}$ input with regularly varying jobs; we have that $\mathbb{P}(B > x) = x^{-\delta}L(x)$, for some $\delta > 1$ and $L(\cdot)$ being a slowly varying function [12]. We do so by relying on the 'principle of a single big jump,' the intuition underlying the proof being that the event of interest is essentially due to one large service requirement. In other words: in order for the workload of queue 1 to exceed αx, and for the workload of queue 2 to exceed $(1-\alpha)x$, with overwhelming probability this is due to a single job, whose size is roughly of the order x. This idea has been used in several papers before, see e.g. [24, 44], or, in a more complex model, [7].

The proof consists of a lower bound that focuses on the probability of the single most likely event, in conjunction with an upper bound that shows that all other scenarios (for instance those with multiple big jumps) yield negligible contributions. The lower bound is relatively straightforward, and provides us with interesting insights into the way the rare event under consideration occurs. The upper bound requires more work, but the 'standard recipe' from [43, pp. 37–39] applies directly here, as we will argue below.

Lower bound Suppose a job, arriving at the system in stationarity, enters at time $-t$. Then we may wonder how large it should be to make sure that queue 1 is larger than αx at time 0, and queue 2 larger than $(1-\alpha)x$. First focus on the first queue. As the first buffer, roughly, drains at rate $\mu - c_1$, it is clear that the job should be at least $\alpha x + (c_1 - \mu)t$. Now consider the second queue. As traffic leaves the first queue at a maximum rate c_1, it is readily seen that t should be at least $t_\alpha x$, with $t_\alpha := (1-\alpha)/(c_1 - c_2)$.

This idea can be used to construct a lower bound on $\pi_\alpha(x)$, as follows. Because of the (weak) law of large numbers, we can find, for any $\zeta, \varepsilon > 0$, an x_0 such that for all $t \geq t_\alpha x_0$,

$$\mathbb{P}\left(\frac{A(-t,0)}{t} > \mu - \varepsilon\right) > 1 - \zeta.$$

Jobs arriving at rate λ, this leads to the lower bound

$$\begin{aligned}\pi_\alpha(x) &\geq \int_{t_\alpha x}^{\infty} \lambda \mathbb{P}\big(B > \alpha x + (c_1 - \mu + \varepsilon)t\big)\mathbb{P}\big(A(-t,0) > (\mu - \varepsilon)t\big)\,\mathrm{d}t \\ &\geq (1-\zeta)\int_{t_\alpha x}^{\infty} \lambda \mathbb{P}\big(B > \alpha x + (c_1 - \mu + \varepsilon)t\big)\,\mathrm{d}t.\end{aligned}$$

After a change-of-variable $\alpha x + (c_1 - \mu + \varepsilon)t =: y$, and using that, due to Karamata's theorem,

$$\int_u^\infty x^{-\delta} L(x)\,\mathrm{d}x \sim \frac{1}{\delta - 1} u^{1-\delta} L(u),$$

we obtain

$$\pi_\alpha(x) \geq (1-\zeta)\frac{\lambda}{c_1 - \mu + \varepsilon}\frac{1}{\delta - 1}\left(\frac{c_1 - \mu + \varepsilon}{c_1 - c_2} - \alpha\left(\frac{c_2 - \mu + \varepsilon}{c_1 - c_2}\right)\right)^{1-\delta} \cdot x^{1-\delta} L(x).$$

Letting $\zeta, \varepsilon \downarrow 0$, we obtain

$$\liminf_{x\to\infty} \frac{\pi_\alpha(x)}{x^{1-\delta} L(x)} \geq \frac{\lambda}{c_1 - \mu}\frac{1}{\delta - 1}\left(\frac{c_1 - \mu}{c_1 - c_2} - \alpha\left(\frac{c_2 - \mu}{c_1 - c_2}\right)\right)^{1-\delta}. \tag{8}$$

Upper bound We have identified above the most likely scenario. Now we show that all other scenarios can be asymptotically neglected. We follow the same steps as in [43, pp. 38–39]; as many of the arguments are standard—and essentially identical to those in, for example, [44]—we chose to leave out some details.

- With $D(-\tau,0)$ the amount of traffic leaving from the first queue between $-\tau$ and 0, we have that

$$\pi_\alpha(x)=\mathbb{P}\big(\exists\sigma>0: A(-\sigma,0)-c_1\sigma>\alpha x, \exists\tau>0: D(-\tau,0)-c_2\tau>(1-\alpha)x\big).$$

The first step is to prove that

$$\limsup_{x\to\infty}\frac{1}{\pi_\alpha(x)}\cdot\mathbb{P}\begin{pmatrix}\exists\sigma\geq Mx: A(-\sigma,0)-c_1\sigma>\alpha x,\\ \exists\tau\geq Mx: D(-\tau,0)-c_2\tau>(1-\alpha)x\end{pmatrix}\to 0 \tag{9}$$

as $M\to\infty$, i.e., that we can restrict ourselves to considering just the time interval $[-Mx,0]$. This is done as follows. Observe that, with $E_x:=\{A(-Mx,0)<(\mu+\zeta)Mx\}$ and $\zeta>0$, the following lower bound applies:

$$\begin{aligned}
&\mathbb{P}\big(\exists\sigma\geq Mx: A(-\sigma,0)-c_1\sigma>\alpha x, \exists\tau\geq Mx: D(-\tau,0)-c_2\tau>(1-\alpha)x\big)\\
&\quad\leq\mathbb{P}\big(\exists\sigma\geq Mx: A(-\sigma,0)-c_1\sigma>\alpha x\big)\\
&\quad=\mathbb{P}\big(\exists\sigma\geq Mx: A(-\sigma,-Mx)-c_1(\sigma-Mx)>\alpha x+c_1Mx-A(-Mx,0)\big)\\
&\quad\leq\mathbb{P}\big(E_x;\exists\sigma\geq Mx: A(-\sigma,-Mx)-c_1(\sigma-Mx)>\alpha x+c_1Mx\\
&\qquad-A(-Mx,0)\big)+\mathbb{P}\big(E_x^c\big)\\
&\quad\leq\mathbb{P}\big(\exists\sigma\geq Mx: A(-\sigma,-Mx)-c_1(\sigma-Mx)>\alpha x+(c_1-\mu-\zeta)Mx\big)\\
&\qquad+\mathbb{P}\big(E_x^c\big).
\end{aligned}\tag{10}$$

Now consider these two probabilities separately. Due to the fact that Lévy processes have stationary independent increments, the first probability in (10) reads

$$\begin{aligned}
&\mathbb{P}\big(\exists\sigma\geq 0: A(-\sigma,0)-c_1\sigma>\alpha x+(c_1-\mu-\zeta)Mx\big)\\
&\quad=\mathbb{P}\big(Q_1>\big(\alpha+(c_1-\mu-\zeta)M\big)x\big),
\end{aligned}$$

which is asymptotically equal to, see e.g. [13, 16],

$$\frac{c_1}{c_1-\mu}\cdot\big(\alpha+(c_1-\mu-\zeta)M\big)^{1-\delta}\cdot x^{1-\delta}L(x).$$

Now consider the second probability in (10), i.e., $\mathbb{P}(E_x^c)$. Using the argumentation as in the proof of [32, Proposition 3.3],

$$\begin{aligned}
\mathbb{P}\big(E_x^c\big)&=\mathbb{P}\big(A(-Mx,0)-(\mu+\zeta/2)Mx\geq(\zeta/2)Mx\big)\\
&\leq\mathbb{P}\big(\exists\sigma>0: A(-\sigma,0)-(\mu+\zeta/2)\sigma\geq(\zeta/2)Mx\big)=\mathbb{P}\big(\bar{Q}\geq(\zeta/2)Mx\big),
\end{aligned}$$

where $\bar{Q}$ is defined as Q_1, but now with service rate $\mu + \zeta/2$ rather than c_1. Again applying e.g. [13, 16], the latter probability is asymptotically equal to

$$\frac{\mu + \zeta/2}{\zeta/2} \cdot (\zeta M/2)^{1-\delta} \cdot x^{1-\delta} L(x).$$

The claim (9) follows by first applying the asymptotic lower bound (8) to $\pi_\alpha(x)$, then letting x grow large, and finally tending $M \to \infty$.

- Due to the previous step, we only have to consider the time interval $[-Mx, 0]$. The next step is to prove that there is at least one big job in this interval, in order to realize the event of interest; here a big job is defined as a job larger than ϵx, with $\epsilon > 0$. This is done by considering the complementary probability of no big jobs in this interval. The argumentation is as in [43, p. 39]: the desired property follows directly by applying a result in [38], which is Lemma 2.4.1 in [43]. It yields that the probability of exceeding level αx (level $(1-\alpha)x$, respectively) in the first (second) queue, with no big jumps in the interval $[-Mx, 0]$ is negligible relative to $x^{1-\delta} L(x)$, for x large.
- Now consider the probability of multiple big jumps in $[-Mx, 0]$. As in [43, p. 39], it can be argued that this probability is regularly varying of index $1-2\delta$, and hence negligible relative to $x^{1-\delta} L(x)$, for x large.
- The last point is slightly different from the recipe in [43], as we want to rule out big jumps in $(-t_\alpha x, 0]$ as well. These can be excluded for the following reason. Suppose the big jump occurs after $-t_\alpha x$. Then, in order to make sure that enough traffic accumulates in the second queue, the first queue must have been non-empty for a time interval of length proportional to x. Busy periods have tail asymptotics that are regularly varying of index $-\delta$ [34], whereas the contribution of the big jump in $(-t_\alpha x, 0]$ is regularly varying of index $1-\delta$. Conclude that big jumps in $(-t_\alpha x, 0]$ lead to a contribution that is negligible with respect to $x^{1-\delta} L(x)$, for x large.

As indicated in [43, p. 39], now that we have made sure that only the 'dominant scenario' plays a role asymptotically, establishing the upper bound is a matter of a straightforward computation, which is fully analogous to the corresponding computation in the single-M/G/1 context. It means that the asymptotics of a single big jump between $-Mx$ and $-t_\alpha x$ are to be determined; these turn out to agree with the lower bound. We have thus arrived at the following result.

Theorem 3.10 *As $x \to \infty$,*

$$\pi_\alpha(x) \sim \frac{\lambda}{c_1 - \mu} \frac{1}{\delta - 1} \left(\frac{c_1 - \mu}{c_1 - c_2} - \alpha \left(\frac{c_2 - \mu}{c_1 - c_2} \right) \right)^{1-\delta} \cdot x^{1-\delta} L(x).$$

The above result agrees with the asymptotics of $\mathbb{P}(Q_1 > x)$ when letting $\alpha \uparrow 1$, see [13, 16]; indeed, we essentially automatically have $Q_2 > 0$ whenever $Q_1 > x$. When letting $\alpha \downarrow 0$, it is clear that we should *not* obtain the asymptotics of $\mathbb{P}(Q_2 > x)$, as there is a significant probability that $Q_2 > x$ occurs, but $Q_1 = 0$. It is easily checked that for analyzing $\mathbb{P}(Q_2 > x)$ the 'dominant scenario' has the following form. Suppose again that the large job, of size b, arrives at $-t$; for reasons explained above, t needs to be larger than t_0x. Then the first buffer is empty after about $b/(c_1-\mu)$ units of time; then the content of the second buffer is about $b(c_1-c_2)/(c_1-\mu)$. After that period, the second buffer drains at a rate $\mu-c_2$. Hence, b should be such that

$$\frac{b}{c_1-\mu}(c_1-c_2)+\left(t-\frac{b}{c_1-\mu}\right)(\mu-c_2)>x,$$

or, equivalently, $b>x+(c_2-\mu)t$. This leads to the expression

$$\begin{aligned}\mathbb{P}(Q_2>x)&\approx\int_{t_0x}^{\infty}\lambda\mathbb{P}\big(B>x+(c_2-\mu)t\big)\,\mathrm{d}t\\&\approx\frac{\lambda}{c_2-\mu}\frac{1}{\delta-1}\left(\frac{c_1-\mu}{c_1-c_2}\right)^{1-\delta}x^{1-\delta}L(x).\end{aligned}\tag{11}$$

In [19, Theorem 5.7] it was shown that this approximation is asymptotically exact (which can also be proven with sample-path lower and upper bounds, as was done above for $\pi_\alpha(x)$); in the next section we retrieve this relation as a corollary of a more general result, see Example 4.8.

4 Asymptotics of the downstream queue

In this section we focus on the tail asymptotics of the downstream (i.e., second) queue, rather than on the joint asymptotics. It turns out that we can derive exact asymptotics, both in the light-tailed and heavy-tailed cases. The results complement those in [19, Sects. 4–5]. As we will see, depending on whether the Lévy process has light tails or heavy tails, we need to apply two different methods to derive the asymptotics: for the light-tailed case specific techniques are available (cf. the 'Heaviside approach' in [1], relying on, e.g., [20]), which are intrinsically different from the Tauberian techniques to be used in the heavy-tailed case, see e.g. [11].

4.1 Light-tailed input

In this subsection we derive the asymptotics of $\mathbb{P}(Q_2>x)$ in case that the Lévy process has light tails, i.e., in case $\kappa(s)$ is also defined for some negative s. The following lemma turns out to be useful, and is straightforward to prove, using (2) and integration by parts.

Lemma 4.1 *For all $t\geq 0$,*

$$\int_0^\infty e^{-tx}\mathbb{P}(Q_2>x)\,\mathrm{d}x=\frac{1}{t}\left(\frac{(c_1-c_2)t-(c_1-\mu)\vartheta^{-1}((c_1-c_2)t)}{(c_1-c_2)(t-\vartheta^{-1}((c_1-c_2)t))}\right).\tag{12}$$

Now let us consider the $t<0$ for which the transform (12) is still well-defined. In the first place, in the notation of the previous section, we should have that t is larger than the *branching point* $\bar{\vartheta}/(c_1-c_2)$. Also, recall that $\bar{t}$, as introduced in the previous section, corresponds to a *pole* of (12), as it solves $(c_1-c_2)(t-\vartheta^{-1}((c_1-c_2)t))=0$. We arrive at the following lemma.

Lemma 4.2 *If $c_1 > -\kappa'(\bar{t})$ then the rightmost singularity of $\mathbb{E}e^{-tQ_2}$ is the pole $t_\mathrm{p} := \bar{t}$, else the rightmost singularity is the branching point $t_\mathrm{b} := \bar{\vartheta}/(c_1-c_2)$.*

Proof The proof follows from the above and the proof of Lemma 3.3, in conjunction with the fact that $\bar{t} > \bar{s}$ is equivalent to $\vartheta'(\bar{t})/(c_1-c_2) > 0$, which in turn is equivalent to $c_1 > -\kappa'(\bar{t})$. □

We now study the tail behavior of $\mathbb{P}(Q_2 > x)$; it turns out that Lemma 4.2 entails that there is a sharp dichotomy (or, 'trichotomy,' as there is a boundary case that needs to be treated separately).

- Let us first focus on the case that $c_1 > -\kappa'(\bar{t})$, i.e., the case that the pole dominates. Define

$$K := -\lim_{t\downarrow t_\mathrm{p}} \frac{t-\vartheta^{-1}((c_1-c_2)t)}{t-t_\mathrm{p}} = \frac{c_1-c_2}{\vartheta'(\vartheta^{-1}((c_1-c_2)t_\mathrm{p}))} - 1 = \frac{c_1-c_2}{\vartheta'(t_\mathrm{p})} - 1.$$

  Then evaluate the transform (12) for $t \downarrow t_\mathrm{p}$. From

$$(t-t_\mathrm{p})\int_0^\infty e^{-tx}\mathbb{P}(Q_2 > x)\,\mathrm{d}x \sim \left(\frac{c_2-\mu}{c_1-c_2}\right)\frac{1}{K},$$

  standard techniques for asymptotic inversion of Laplace transforms yield

$$\mathbb{P}(Q_2 > x) \sim \frac{c_2-\mu}{c_1-c_2}\frac{1}{K}e^{t_\mathrm{p}x}.$$

- Now consider the case $c_1 = -\kappa'(\bar{t})$, that is, the pole t_p and the branching point t_b coincide. It is readily checked that

$$\vartheta^{-1}\big((c_1-c_2)t\big) - t_\mathrm{p} \sim \sqrt{\frac{2(c_1-c_2)}{\vartheta''(t_\mathrm{p})}}\sqrt{t-t_\mathrm{p}},$$

  implying that (12) behaves as, with $W_1 := \sqrt{2(c_1-c_2)/\vartheta''(t_\mathrm{p})}$, K_1 denoting an (irrelevant) constant, and $t \downarrow t_\mathrm{b} = t_\mathrm{p}$,

$$\begin{aligned}
&\frac{1}{t}\left(\frac{(\mu-c_2)t_\mathrm{p} - (c_1-\mu)W_1\sqrt{t-t_\mathrm{p}}}{(c_1-c_2)(t-t_\mathrm{p}-W_1\sqrt{t-t_\mathrm{p}})}\right)\\
&\quad\sim \frac{1}{t_\mathrm{p}(c_1-c_2)}\left(\frac{(\mu-c_2)t_\mathrm{p} - (c_1-\mu)W_1\sqrt{t-t_\mathrm{p}}}{-W_1\sqrt{t-t_\mathrm{p}}}\right)\\
&\quad\sim K_1 + \frac{c_2-\mu}{c_1-c_2}\frac{1}{W_1}\frac{1}{\sqrt{t-t_\mathrm{p}}}.
\end{aligned}$$

Using techniques as those in, e.g., [1], this leads to

$$\mathbb{P}(Q_2 > x) \sim \frac{1}{\Gamma(\frac{1}{2})} \frac{c_2 - \mu}{c_1 - c_2} \sqrt{\frac{\vartheta''(t_\mathrm{p})}{2(c_1 - c_2)}} \frac{1}{\sqrt{x}} e^{t_\mathrm{p} x}$$
$$= \frac{1}{\sqrt{2\pi}} \frac{c_2 - \mu}{c_1 - c_2} \sqrt{\frac{\vartheta''(t_\mathrm{p})}{c_1 - c_2}} \frac{1}{\sqrt{x}} e^{t_\mathrm{p} x}.$$

- Finally consider the case $c_1 < -\kappa'(\bar{t})$. Now

$$\vartheta^{-1}\big((c_1 - c_2)t\big) - \bar{s} \sim \sqrt{\frac{2(c_1 - c_2)}{\vartheta''(\bar{s})}} \sqrt{t - t_\mathrm{b}},$$

so that (12) behaves as, with $W_2 := \sqrt{2(c_1 - c_2)/\vartheta''(\bar{s})}$, K_2 denoting an (irrelevant) constant, and $t \downarrow t_\mathrm{b}$,

$$\frac{1}{t}\left(\frac{(c_1 - c_2)t - (c_1 - \mu)(\bar{s} + W_2\sqrt{t - t_\mathrm{b}})}{(c_1 - c_2)(t - \bar{s} - W_2\sqrt{t - t_\mathrm{b}})}\right)$$
$$= \frac{1}{t(c_1 - c_2)}\left(\frac{(c_1 - c_2)t - (c_1 - \mu)(\bar{s} + W_2\sqrt{t - t_\mathrm{b}})}{t - \bar{s} - W_2\sqrt{t - t_\mathrm{b}}}\right)\left(\frac{t - \bar{s} + W_2\sqrt{t - t_\mathrm{b}}}{t - \bar{s} + W_2\sqrt{t - t_\mathrm{b}}}\right)$$
$$\sim K_2 - \frac{c_2 - \mu}{(t_\mathrm{b} - \bar{s})^2} \sqrt{\frac{2}{(c_1 - c_2)\vartheta''(\bar{s})}} \sqrt{t - t_\mathrm{b}},$$

and hence

$$\mathbb{P}(Q_2 > x) \sim -\frac{1}{\Gamma(-\frac{1}{2})} \frac{c_2 - \mu}{(t_\mathrm{b} - \bar{s})^2} \sqrt{\frac{2}{(c_1 - c_2)\vartheta''(\bar{s})}} \frac{1}{x\sqrt{x}} e^{t_\mathrm{b} x}$$
$$= \frac{1}{\sqrt{2\pi}} \frac{c_2 - \mu}{(t_\mathrm{b} - \bar{s})^2} \sqrt{\frac{1}{(c_1 - c_2)\vartheta''(\bar{s})}} \frac{1}{x\sqrt{x}} e^{t_\mathrm{b} x};$$

cf. again [1], but also the busy-period asymptotics in [17].

The following theorem states the asymptotics of $\mathbb{P}(Q_2 > x)$.

Theorem 4.3 *For $c_1 > -\kappa'(\bar{t})$,*

$$\mathbb{P}(Q_2 > x) \sim \frac{c_2 - \mu}{c_1 - c_2} \frac{\vartheta'(t_\mathrm{p})}{c_1 - c_2 - \vartheta'(t_\mathrm{p})} e^{t_\mathrm{p} x};$$

for $c_1 = -\kappa'(\bar{t})$, with $t_\mathrm{p} = t_\mathrm{b}$,

$$\mathbb{P}(Q_2 > x) \sim \frac{1}{\sqrt{2\pi}} \frac{c_2 - \mu}{c_1 - c_2} \sqrt{\frac{\vartheta''(t_\mathrm{p})}{c_1 - c_2}} \frac{1}{\sqrt{x}} e^{t_\mathrm{p} x};$$

for $c_1 < -\kappa'(\bar{t})$,

$$\mathbb{P}(Q_2 > x) \sim \frac{1}{\sqrt{2\pi}} \frac{c_2 - \mu}{(t_b - \bar{s})^2} \sqrt{\frac{1}{(c_1 - c_2)\vartheta''(\bar{s})}} \frac{1}{x\sqrt{x}} e^{t_b x}.$$

Example 4.4 Again consider $A \in \mathbb{B}\mathrm{m}(0, 1)$; as before, this can be translated in a straightforward way to $A \in \mathbb{B}\mathrm{m}(\mu, \sigma^2)$ for any $\mu < c_2$ and $\sigma^2 > 0$. We now use Theorem 4.3 to compute the asymptotics of $\mathbb{P}(Q_2 > x)$, also relying on the findings of Example 3.8. We find, after some simplification, for $c_1 > 2c_2$,

$$\mathbb{P}(Q_2 > x) \sim \frac{c_1 - 2c_2}{c_1 - c_2} e^{-2c_2 x};$$

for $c_1 = 2c_2$,

$$\mathbb{P}(Q_2 > x) \sim \frac{1}{\sqrt{2\pi c_2}} \frac{1}{\sqrt{x}} e^{-2c_2 x};$$

for $c_1 < 2c_2$,

$$\mathbb{P}(Q_2 > x) \sim \frac{1}{\sqrt{2\pi}} \left(\frac{c_1 - c_2}{x}\right)^{3/2} \frac{4c_2}{c_1^2 (c_1 - 2c_2)^2} \exp\left(-\frac{c_1^2}{2(c_1 - c_2)} x\right).$$

These findings are in agreement with [19, Corollary 4.4], where the asymptotics were derived by first explicitly calculating the full distribution function $\mathbb{P}(Q_2 > x)$.

Interestingly, observing that t_p does not depend on c_1, one could say that Theorem 4.3 implies that essentially two regimes exist: for $c_1 \geq -\kappa'(\bar{t})$ the first queue hardly affects the tail asymptotics of the downstream queue, as c_1 only affects the proportionality constant. For $c_1 < -\kappa'(\bar{t})$ the first queue plays a more explicit role in delaying and reshaping the traffic before entering the second queue. This sharp dichotomy is in line with those in, e.g., [15, 33].

4.2 Heavy-tailed case

In this subsection we assume that the Lévy input process is heavy-tailed, implying that the rightmost singularity of $\mathbb{E}e^{-tQ_2}$ is 0. We can then use Tauberian theorems to derive the asymptotics of $\mathbb{P}(Q_2 > x)$ as $x \to \infty$. Let us first present the following definition, see [21].

Definition 4.5 We say that $f(x) \in \mathcal{R}_\delta(n, \eta)$, with $\delta \in (n, n+1)$, for $x \downarrow 0$, if

$$f(x) = \sum_{i=0}^{n} \frac{f^{(i)}(0)}{i!} x^i + \eta x^\delta L(1/x), \quad x \downarrow 0,$$

for a slowly varying function L, i.e., $L(x)/L(tx) \to 1$ for $x \to \infty$, for any t.

The following lemma follows directly from Lemma 5.2 in [21]. Recall that $\vartheta'(0) = c_1 - \mu = 1/(\vartheta^{-1}(0))'$.

Lemma 4.6 *Suppose that $\vartheta(s) \in \mathcal{R}_\delta(n,\eta)$. Then*

$$\vartheta^{-1}(t) \in \mathcal{R}_\delta\left(n, -\frac{\eta}{(\vartheta'(0))^{\delta+1}}\right) = \mathcal{R}_\delta\left(n, -\frac{\eta}{(c_1-\mu)^{\delta+1}}\right);$$

$$\vartheta^{-1}\big((c_1-c_2)t\big) \in \mathcal{R}_\delta(n,\zeta), \quad \zeta := -\frac{\eta(c_1-c_2)^\delta}{(c_1-\mu)^{\delta+1}},$$

for $s \downarrow 0$.

In order to apply Tauberian theorems, we need to characterize the behavior of $\mathbb{E}^{-tQ_2}$ as $t \downarrow 0$. Due to Lemma 4.6 we have that $\vartheta^{-1}((c_1-c_2)t) \in \mathcal{R}_\delta(n,\zeta)$, so that, with $\xi := (c_2-\mu)/(c_1-c_2)$,

$$\mathbb{E}e^{-tQ_2} = \xi \cdot \frac{\vartheta^{-1}(t(c_1-c_2))}{t-\vartheta^{-1}(t(c_1-c_2))} = \xi \cdot \frac{\sum_{i=1}^n a_i t^i + \zeta t^\delta L(1/t)}{t(1-a_1) - \sum_{i=2}^n a_i t^i - \zeta t^\delta L(1/t)},$$

for appropriately chosen constants $a_1,\dots,a_n$ (where $a_1 = (c_1-c_2)/(c_1-\mu)$, as is easily verified). Dividing both numerator and denominator by $t(1-a_1)$, we obtain

$$\xi \cdot \frac{\sum_{i=0}^{n-1}(a_{i+1}/(1-a_1))t^i + (\zeta/(1-a_1))t^{\delta-1}L(1/t)}{1-\sum_{i=1}^{n-1}(a_{i+1}/(1-a_1))t^i - (\zeta/(1-a_1))t^{\delta-1}L(1/t)}.$$

Now applying the standard representation $1/(1-x) = \sum_{i=0}^\infty x^i$, we directly observe that we find two terms of the order $t^{\delta-1}$: one is proportional to $\zeta/(1-a_1)$, the other is proportional to $a_1\zeta/(1-a_1)^2$, and hence, for $t \downarrow 0$,

$$\mathbb{E}e^{-tQ_2} \in \mathcal{R}_{\delta-1}(n-1, \xi\cdot\bar{\zeta}), \quad \bar{\zeta} := \frac{\zeta}{1-a_1} + \frac{a_1\zeta}{(1-a_1)^2} = \zeta\left(\frac{c_1-\mu}{c_2-\mu}\right)^2.$$

The Tauberian theorem in Bingham, Goldie, and Teugels [12, Theorem 8.1.6] now yields the following result; see also [11].

Theorem 4.7 *If $\vartheta(s) \in \mathcal{R}_\delta(n,\eta)$, with $n \in \{1,2,\dots\}$ and $\delta \in \{n, n+1\}$, then, as $x \to \infty$,*

$$\mathbb{P}(Q_2 > x) \sim \frac{(-1)^n}{\Gamma(2-\delta)} \cdot (\xi\cdot\bar{\zeta}) \cdot x^{1-\delta}L(x)$$
$$= \frac{(-1)^{n+1}}{\Gamma(2-\delta)} \frac{\eta}{c_2-\mu}\left(\frac{c_1-\mu}{c_1-c_2}\right)^{1-\delta} x^{1-\delta}L(x).$$

Example 4.8 Suppose $\mathbb{P}(B > x) \sim x^{-\delta}L(x)$. From $\vartheta(s) = c_1 s + \lambda\beta(s) - \lambda$, it follows that $\theta(s) \in \mathcal{R}_\delta(n, \lambda\Gamma(1-\delta)(-1)^n)$ by applying 'Tauber.' Then the above theorem entails that the approximation given in (11) is asymptotically exact (use $(1-\delta)\cdot\Gamma(1-\delta) = \Gamma(2-\delta)$).

Example 4.9 Let us consider the case of spectrally-positive α-stable Lévy motion; this means that, in the notation of [30, p. 10] and [8, p. 217], $\beta \equiv 1$. In this case

$\kappa(s) = Cs^\delta$, for some positive constant C and $\delta \in (1,2)$; see e.g. [30, Exercise 3.7] and [25]. The above theorem can be applied instantly, with $n = 1$ and $\eta = C$, and we retrieve [19, Theorem 5.5].

5 Discussion and concluding remarks

Compound Poisson with subexponential jobs Restricting ourselves to the case of compound Poisson input, the light-tailed results cover the case in which the jobs have a finite moment generating function in a neighborhood of the origin: $\beta(\alpha) < \infty$ for some $\alpha < 0$; as a result, in this situation all moments are finite. On the other hand, the heavy-tailed results correspond to the situation in which just a finite number of moments are finite. In between, however, there is a third class of distributions: those for which all moments are finite, but without an analytic continuation for $\alpha < 0$ (that is $\beta(\alpha) = \infty$ for all $\alpha < 0$). Examples of the *subexponential distributions* in this class are the Weibull and Log-Normal distributions. It is likely that, just as is the case in the single-node situation, the results for regularly varying jobs carry over to subexponential jobs, see e.g. [35, 41].

Exact asymptotics In the light-tailed case, we found in Sect. 3.1 just logarithmic asymptotics. A subject for future research concerns the identification of the corresponding exact asymptotics. The results for the special case of Brownian input [31] indicate that it can be expected that for α in some specific range these have a purely exponential shape, whereas for other α in addition a factor $1/\sqrt{x}$ will appear. In [5, Sect. 5], which has been written simultaneously with the present paper, this has already been done for the compound Poisson case as well as the purely Brownian case. We remark that the techniques used in [5] substantially differ from ours; in addition, our approach also yields the 'most likely paths to overflow,' as determined in the proof of Proposition 3.5. Also, [5] does not consider the heavy-tailed case.

Tandem series with more than two nodes Let us now consider a three-node tandem queue with light-tailed Lévy input, and suppose we wish to find the asymptotics of

$$\mathbb{P}(Q_i > \alpha_i x, i = 1, 2, 3),$$

with $\alpha_i > 0$ and $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

Without loss of generality, assume that $c_1 > c_2 > c_3 > \mu$. Define $\vartheta_i(s) := \kappa(s) + c_i s$, and let $\vartheta_i^{-1}(\cdot)$ denote the inverse of $\vartheta_i(\cdot)$, $i = 1, 2, 3$. Then it is known [18] that, for $(s,t,u) \in \mathbb{R}_+^3$,

$$\begin{aligned}\mathbb{E}e^{-sQ_1 - tQ_2 - uQ_3} &= \frac{(c_3-\mu)u}{\vartheta_3(u)} \times \frac{\vartheta_1^{-1}((c_1-c_2)t + (c_2-c_3)u) - s}{\vartheta_1^{-1}((c_1-c_2)t + (c_2-c_3)u) - t} \\ &\quad \times \frac{\vartheta_2^{-1}((c_2-c_3)u) - t}{\vartheta_2^{-1}((c_2-c_3)u) - u} \\ &\quad \times \frac{(c_1-c_2)t + (c_2-c_3)u - \vartheta_1(t)}{(c_1-c_2)t + (c_2-c_3)u - \vartheta_1(s)} \times \frac{(c_2-c_3)u - \vartheta_2(u)}{(c_2-c_3)u - \vartheta_2(t)},\end{aligned}$$

which can be rewritten to

$$-\frac{(c_3-\mu)u}{\vartheta_2^{-1}((c_2-c_3)u)-u}\times\frac{\vartheta_2^{-1}((c_2-c_3)u)-t}{\vartheta_1^{-1}((c_1-c_2)t+(c_2-c_3)u)-t}$$
$$\times\frac{\vartheta_1^{-1}((c_1-c_2)t+(c_2-c_3)u)-s}{(c_1-c_2)t+(c_2-c_3)u-\vartheta_1(s)},$$

the product of three factors: one with just u, one with t and u, and one with s, t, and u. This relation holds for positive arguments, but again there may be a convex set of $(s,t,u)\in\mathbb{R}^3_-$ for which this is finite; the factorization above may help to further specify this convex set. It is now seen that, relying on the Chernoff bound, again an upper bound on the decay rate of $\mathbb{P}(Q_i>\alpha_i x, i=1,2,3)$ can be found by solving a convex programming problem. As before, we can then use sample-path large-deviations to show that this bound is tight; now it is expected that the most likely path consists, in some of the scenarios, of three linear segments (the third queue starting to build up first, then the second queue, and finally the first queue).

This procedure can in fact be followed for any number of hops. Assuming $c_1>c_2>\cdots>c_{n-1}>c_n>\mu$ to ensure stability and non-triviality, and with $\vartheta_i(s):=\kappa(s)+c_i s$ and $\vartheta_i^{-1}(\cdot)$ defined as before. Exploiting [18], we find that, for $(s_1,\dots,s_n)\in\mathbb{R}^n_+$,

$$\mathbb{E}e^{-\sum_{i=1}^n s_iQ_i}=\frac{(c_n-\mu)s_n}{\theta_n(s_n)}\times\prod_{j=1}^{n-1}\frac{\vartheta_j^{-1}(\sum_{k=j+1}^n(c_{k-1}-c_k)s_k)-s_j}{\vartheta_j^{-1}(\sum_{k=j+1}^n(c_{k-1}-c_k)s_k)-s_{j+1}}$$
$$\times\prod_{j=1}^{n-1}\frac{\sum_{k=j+1}^n(c_{k-1}-c_k)s_k-\vartheta_j(s_{j+1})}{\sum_{k=j+1}^n(c_{k-1}-c_k)s_k-\vartheta_j(s_j)}. \tag{13}$$

Using that $\vartheta_i(s)=\vartheta_{i-1}(s)+(c_i-c_{i-1})s$, $i=2,\dots,n$, we find that (13) reduces to

$$-\frac{(c_n-\mu)s_n}{\sum_{k=2}^n(c_{k-1}-c_k)s_k-\vartheta_1(s_1)}\times\prod_{j=1}^{n-1}\frac{\vartheta_j^{-1}(\sum_{k=j+1}^n(c_{k-1}-c_k)s_k)-s_j}{\vartheta_j^{-1}(\sum_{k=j+1}^n(c_{k-1}-c_k)s_k)-s_{j+1}}. \tag{14}$$

Finally, rearranging (14) yields

$$\mathbb{E}e^{-\sum_{i=1}^n s_iQ_i}=-\frac{\vartheta_1^{-1}(\sum_{k=2}^n(c_{k-1}-c_k)s_k)-s_1}{\sum_{k=2}^n(c_{k-1}-c_k)s_k-\vartheta_1(s_1)}$$
$$\times\left(\prod_{j=2}^{n-1}\frac{\vartheta_j^{-1}(\sum_{k=j+1}^n(c_{k-1}-c_k)s_k)-s_j}{\vartheta_{j-1}^{-1}(\sum_{k=j}^n(c_{k-1}-c_k)s_k)-s_j}\right)$$
$$\times\frac{(c_n-\mu)s_n}{\vartheta_{n-1}^{-1}((c_{n-1}-c_n)s_n)-s_n}.$$

We observe the remarkable fact that the transform can be rewritten as the product of n fractions, where both the numerator and denominator of the ith fraction only depend on $s_i, \dots, s_n$, $i = 1, \dots, n$. This factorization may be useful in order to explicitly derive the boundary of the 'feasible region' (that is, the region in which the joint Laplace transform $\mathbb{E}\exp(-\sum_{i=1}^{n} s_i Q_i)$ is finite).

In the case of heavy-tailed input, the intuition is as in Sect. 3.2: the asymptotics of the probability $\mathbb{P}(Q_i > \alpha_i x, i = 1, 2, 3)$ are fully dominated by the scenario of one single big jump. Evidently, if this large job arrives at time $-t$, it must be at least $\alpha_1 x + (c_1 - \mu)t$, where t should be larger than both $\alpha_2 x/(c_1 - c_2)$ and $\alpha_3 x/(c_2 - c_3)$.

References

1. Abate, J., Whitt, W.: Asymptotics for M/G/1 low-priority waiting-time tail probabilities. Queueing Syst. **25**, 173–233 (1997)
2. de Acosta, A.: Large deviations for vector-valued Lévy processes. Stoch. Process. Their Appl. **51**, 75–115 (1994)
3. Avi-Itzhak, B.: A sequence of service stations with arbitrary input and regular service times. Manag. Sci. **11**, 565–571 (1965)
4. Avram, F., Palmowski, Z., Pistorius, M.: A two-dimensional ruin problem on the positive quadrant. Insur. Math. Econ. **42**, 227–234 (2008)
5. Avram, F., Palmowski, Z., Pistorius, M.: Exit problem of a two-dimensional risk process from the quadrant: exact and asymptotic result. Ann. Appl. Probab. (2008, to appear)
6. Baccelli, F., Foss, S.: Moments and tails in monotone-separable stochastic networks. Ann. Appl. Probab. **14**, 612–650 (2004)
7. Bekker, R., Mandjes, M.: A fluid model for a relay node in an ad hoc network: the case of heavy-tailed input. Math. Methods Oper. Res. (2008, to appear), url: http://ftp.cwi.nl/CWIreports/PNA/PNA-E0703.pdf
8. Bertoin, J.: Lévy Processes. Cambridge University Press, Cambridge (1996)
9. Bertsimas, D., Paschalidis, J., Tsitsiklis, J.: On the large deviation behavior in acyclic networks of G/G/1 queues. Ann. Appl. Probab. **8**, 1027–1069 (1998)
10. Bertsimas, D., Paschalidis, J., Tsitsiklis, J.: Large deviation analysis of the generalized processor sharing policy. Queueing Syst. **32**, 319–349 (1999)
11. Bingham, N., Doney, R.: Asymptotic properties of subcritical branching processes I: the Galton–Watson process. Adv. Appl. Probab. **6**, 711–731 (1974)
12. Bingham, N., Goldie, C., Teugels, J.: Regular Variation. Cambridge University Press, Cambridge (1987)
13. Borovkov, A.: Stochastic Processes in Queueing Theory. Springer, New York (1976)
14. Borovkov, A., Mogulskii, A.: Large deviations for Markov chains in the quarter plane. Usp. Mat. Nauk **56**, 3–115 (2001)
15. Chang, C.-S., Heidelberger, P., Juneja, S., Shahabuddin, P.: Effective bandwidth and fast simulation of ATM in-tree networks. Perform. Eval. **20**, 45–65 (1994)
16. Cohen, J.W.: Some results on regular variation for distributions in queueing and fluctuation theory. J. Appl. Probab. **10**, 343–353 (1973)
17. Cox, D., Smith, W.: Queues. Methuen, London (1961)
18. Dębicki, K., Dieker, A., Rolski, T.: Quasi-product form for Lévy-driven fluid networks. Math. Oper. Res. **32**, 629–647 (2007)
19. Dębicki, K., Mandjes, M., van Uitert, M.: A tandem queue with Lévy input: a new representation of the downstream queue length. Probab. Eng. Inf. Sci. **21**, 83–107 (2007)
20. Doetsch, G.: Introduction to the Theory and Applications of the Laplace Transformation. Springer, New York (1974)
21. Es-Saghouani, A., Mandjes, M.: On the correlation structure of a Lévy-driven queue. J. Appl. Probab. (2008, submitted), url: http://ftp.cwi.nl/CWIreports/PNA/PNA-R0711.pdf

22. Friedman, H.: Reduction methods for tandem queueing systems. Oper. Res. **13**, 121–131 (1965)
23. Foss, S.: On exact asymptotics for a stationary sojourn time distribution in a tandem of queues for a class of light-tailed distributions. Probl. Inf. Transm. **43**, 93–108 (2007)
24. Foss, S., Konstantopoulos, T., Zachary, S.: Discrete and continuous time modulated random walks with heavy-tailed increments. J. Theor. Probab. **20**, 581–612 (2007)
25. Furrer, H.: Risk processes perturbed by α-stable Lévy motion. Scand. Actuar. J. **1**, 59–74 (1998)
26. Ganesh, A., Anantharam, V.: Stationary tail probabilities in exponential server tandems with renewal arrivals. Queueing Syst. **22**, 203–247 (1996)
27. Ganesh, A.: Large deviations of the sojourn time for queues in series. Ann. Oper. Res. **79**, 3–26 (1998)
28. Ignatyuk, I., Malyshev, V., Scherbakov, V.: Boundary effects in large deviations problems. Russ. Math Surv. **49**, 41–99 (1994)
29. Kella, O., Whitt, W.: A tandem fluid network with Lévy input. In: Bhat, U.N., Basawa, I.V. (eds.) Queueing and Related Models, pp. 112–128. Oxford University Press, London (1992)
30. Kyprianou, A.: Introductory Lectures on Fluctuations of Lévy Processes with Applications. Springer, Berlin (2006)
31. Lieshout, P., Mandjes, M.: Tandem Brownian queues. Math. Methods Oper. Res. **66**, 275–298 (2007)
32. Mandjes, M., Borst, S.: Overflow behavior in queues with many long-tailed inputs. Adv. Appl. Probab. **32**, 1150–1167 (2000)
33. Mandjes, M., van Uitert, M.: Sample-path large deviations for tandem and priority queues with Gaussian inputs. Ann. Appl. Probab. **15**, 1193–1226 (2005)
34. De Meyer, A., Teugels, J.: On the asymptotic behavior of the distributions of the busy period and service time in $M/G/1$. J. Appl. Probab. **17**, 802–813 (1980)
35. Pakes, A.: On the tails of waiting-time distributions. J. Appl. Probab. **12**, 555–564 (1975)
36. Pecherskii, E., Sukhov, Ju.: Dobrushin's ideas in the theory of queueing networks. Russ. Math. Surv. **52**, 265–269 (1997)
37. Reich, E.: On the integro-differential equation of Takács I. Ann. Math. Stat. **29**, 563–570 (1958)
38. Resnick, S., Samorodnitsky, G.: Activity periods of an infinite server queue and performance of certain heavy tailed fluid queues. Queueing Syst. **33**, 43–71 (1999)
39. Rubin, I.: Path delays in communication networks. Appl. Math. Optim. **1**, 193–221 (1974)
40. Tsoucas, P.: Rare events in series of queues. J. Appl. Probab. **29**, 168–175 (1992)
41. Veraverbeke, N.: Asymptotic behavior of Wiener–Hopf factors of a random walk. Stoch. Process. Their Appl. **5**, 27–37 (1977)
42. Zolotarev, V.M.: The first passage time of a level and the behavior at infinity for a class of processes with independent increments. Theory Probab. Appl. **9**, 653–661 (1964)
43. Zwart, B.: Queueing Systems with Heavy Tails. Ph.D. Thesis, Eindhoven University of Technology (2001), url: http://alexandria.tue.nl/extra2/200112999.pdf
44. Zwart, B., Borst, S., Mandjes, M.: Exact asymptotics for fluid queues fed by multiple heavy-tailed on-off sources. Ann. Appl. Probab. **14**, 903–957 (2004)