*Probability in the Engineering and Informational Sciences*, **21**, 2007, 83–107. Printed in the U.S.A.

# A TANDEM QUEUE WITH LÉVY INPUT

## *A NEW REPRESENTATION OF THE DOWNSTREAM QUEUE LENGTH*

KRZYSZTOF DĘBICKI
*Mathematical Institute*
*University of Wrocław*
*50-384 Wrocław, Poland*
*E-mail: debicki@math.uni.wroc.pl*

MICHEL MANDJES
*CWI*
*1090 GB Amsterdam, The Netherlands*
*Korteweg-de Vries Institute*
*University of Amsterdam*
*1018 TV Amsterdam, The Netherlands*
*and*
*EURANDOM, 5600 MB Eindhoven, The Netherlands*
*E-mail: michel@ewi.nl*

MIRANDA VAN UITERT
*The Netherlands Cancer Institute*
*1066 CX Amsterdam, The Netherlands*
*and*
*Information and Communication Theory Group*
*Delft University of Technology*
*2600 GA Delft, The Netherlands*
*E-mail: m.uitert@nki.nl*

In this article we present a new representation for the steady-state distribution of the workload of the second queue in a two-node tandem network. It involves the difference of two suprema over two adjacent intervals. In the case of spectrally positive Lévy input, this enables us to derive the Laplace transform and Pollaczek–Khintchine representation of the workload of the second queue. Additionally, we

© 2007 Cambridge University Press 0269-9648/07 $16.00 

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

obtain the exact distribution of the workload in the case of Brownian and Poisson input, as well as some insightful formulas representing the exact asymptotics for $\alpha$-stable Lévy inputs.

## 1. INTRODUCTION

Traffic engineering greatly benefits from models that are capable of accurately describing and predicting the performance of the system. The network nodes are usually modeled as *queues*, and queuing theory can be used to analyze the performance (in terms of loss, delay, throughput, etc.) of these nodes. However, most studies address performance issues for single nodes, which is evidently an oversimplification of reality. This justifies the growing interest in the performance of traffic streams traversing *concatenations* of queues. Measurement studies have provided empirical evidence that, in particular cases, network traffic can be approximated by Gaussian processes [e.g., Brownian motion or fractional Brownian motion (fBm)]. Mikosch, Resnick, Rootzén, and Stegeman [27] and Kaj and Taqqu [22] have focused on the cases in which it is *not* accurate to assume Gaussianity; in [27] it is argued that if the rate at which connections are initiated is modest relative to the tail of the connection-length distribution (which is assumed regularly varying), it is preferred to approximate the traffic stream by a stable Lévy motion. More justification for the use of Lévy motions in a communication networking context can be found in [10].

This article focuses on a two-node tandem network with Lévy input; we note, however, that it is straightforward to extend the analysis to tandem networks with more than two nodes. Our purpose is to analyze the buffer content distribution of the *downstream* (i.e., second) queue. Some results are known for special cases of this model. The tandem queue with Poisson arrivals, for instance, was analyzed in detail, but a considerable part of the results is in terms of Laplace transforms (see, e.g., [33,37,40]). For Brownian motion input, the exponential decay rate of the queue-size distribution of the second queue is given in [26].

Our article is related to earlier works by Kella and Whitt [25] and Kella [23]. These articles also consider tandem queues with Lévy input, but there it is required that the input be *nondecreasing*. This assumption, however, excludes classes of input processes such as Brownian motion and $\alpha$-stable Lévy motions (with $\alpha < 2$). The approach taken in this article covers these important cases. Moreover, in [25] and [23], the emphasis is on moments and Laplace transforms, whereas a substantial part of our results is in terms of explicit expressions for the probability distribution of the downstream queue (or its tail behavior).

This article is organized as follows. Section 2 finds a remarkable new representation of the queue size of the downstream queue. The new formula holds for general (not only Lévy) input processes and involves the difference of two suprema over two adjacent intervals. The derivation relates to results of [26]. The specifi-

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

cation of this representation for spectrally positive Lévy input enables us to analyze (in Section 3) the downstream queue in terms of Laplace transforms. This leads to the analogue of the result by Zolotarev [42] for the single queue. Additionally, we derive the Pollaczek–Khintchine representation of the distribution of the downstream queue that extends Theorem 4.1 in [25] from "subordinator Lévy input" to spectrally positive Lévy input.

Section 4 applies the general results to a number of examples with *light-tailed input*. For both Brownian and Poisson input, we give the exact distribution of the workload of the second queue. Our analysis indicates that in this light-tailed case, there are two regimes: there is a critical service rate $c_1^\star$ such that (1) for $c_1 > c_1^\star$, all traffic entering the first queue is essentially served immediately and goes directly into the second queue, such that traffic is not "reshaped" by the first queue, and (2) for $c_1 \leq c_1^\star$, the first queue *does* play a role in delaying and reshaping the traffic before entering the second queue. This behavior is in line with earlier results in [11,26].

Section 5 deals with examples featuring *heavy-tailed input*. We find asymptotics and bounds of the tail distribution when the input is $\alpha$-stable Lévy motion and compound Poisson (with regularly varying jumps).

## 2. MODEL AND FUNDAMENTAL RESULT

In this section we first describe our tandem queuing model. Then we present our fundamental result.

### 2.1. Model

In this article we consider a two-node tandem network with service rate $c_1$ at the first node and service rate $c_2 < c_1$ at the second node; the buffering capacity is infinite at both nodes. Although traffic can enter the first queue in "quanta" of positive size, we assume that it leaves the first queue as "fluid." In the standard example of packets fed into a communication network, it would mean that during the service of packets in the first queue, the second queue fills at the (net) rate $c_1 - c_2$.

Let $\{A(t), t \in \mathbb{R}\}$ be the input process; then the amount of traffic generated in the interval $(s,t]$, where $s < t$ and $s, t \in \mathbb{R}$, is denoted by $A(s,t) := A(t) - A(s)$. We first assume that $\{A(t), t \in \mathbb{R}\}$ has stationary increments; later we focus on the case of stationary independent increments. We denote by $Q_1$ and $Q_2$ the steady-state workload at the first and second queue, respectively. Analogously, we denote by $Q$ the total steady-state workload in the tandem network. We assume that $\mu := \mathbb{E}\{A(1)\} < c_2$ to ensure that both $Q_1$ and $Q_2$ are finite a.s.

In this article we focus on the probability that the stationary buffer content at the downstream (i.e., second) node exceeds level $u$:

$$\mathbb{P}(Q_2 > u) \quad \text{for } u > 0.$$

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

Following Reich's representation [30], we can write for the first queue, using the fact that $A(\cdot)$ has stationary increments,

$$Q_1 =_{\mathrm{d}} \sup_{t\geq 0}\{A(-t,0) - c_1 t\} =_{\mathrm{d}} \sup_{t\geq 0}\{A(t) - c_1 t\}. \tag{1}$$

Now note that the following "reduction property" applies, which was also used in [25]. Consider a single queue fed by the original input process $\{A(t), t \in \mathbb{R}\}$ and drained at the slowest of the service rates $\min\{c_1, c_2\}$. Then it is not hard to see that the buffer content of this queue is equal to the *total* buffer content of our tandem system (see, e.g., Avi-Itzhak [5] and Friedman [18]). Hence, we have for the total workload $Q$ that

$$Q =_{\mathrm{d}} \sup_{t\geq 0}\{A(-t,0) - c_2 t\} =_{\mathrm{d}} \sup_{t\geq 0}\{A(t) - c_2 t\}, \tag{2}$$

because we assumed that $c_2 < c_1$.

The above observations directly lead to the following representation for $Q_2$:

$$Q_2 =_{\mathrm{d}} \sup_{t\geq 0}\{A(t) - c_2 t\} - \sup_{t\geq 0}\{A(t) - c_1 t\}. \tag{3}$$

*Remark 2.1:* It is tempting to explicitly impose the assumption that $A(t)$ is *increasing* in $t$, bearing in mind that $A(t)$ is interpreted as the amount of traffic entering the queue in $[0, t)$. However, there are good reasons to *not* assume a priori that $A(\cdot)$ is increasing, the most prominent being the following.

Consider the superposition of a number of independent and identically distributed (i.i.d.) *on-off* sources (i.e., sources alternating between sending at some constant, positive rate and being silent); the corresponding input process is evidently increasing. When the number of sources grows large, however, there is, after some appropriate rescaling, convergence to "limit processes" that are *not* necessarily increasing (for instance, (fractional) Brownian motion (with drift) or Lévy input models; see, e.g., [15,41]). This motivates our choice to not restrict ourselves to increasing inputs.

Two additional comments need to be made. First, we mention that in many cases, the above-mentioned convergence of traffic aggregates carries over to the stationary buffer-content process. Loosely speaking, the queue fed by the aggregate input process converges to the queue fed by the "limit process" (see, e.g., [10,14,15]). Second, notice that when using representation (3), the queue length in the downstream node is necessarily positive.

### 2.2. A Representation for the Workload Distribution

In this subsection we prove a fundamental result that simplifies the representation (3) for the queue length in the downstream queue $Q_2$. In particular, we show that the set over which both suprema in (3) should be taken can be decomposed into two

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

adjacent intervals. Define $t_u := u/(c_1 - c_2)$, which can be interpreted as the minimum time it takes for the second queue to exceed level $u$, starting empty.

PROPOSITION 2.2: *Let $c_1 > c_2 > 0$ and let $\{X(t), t \geq 0\}$ be a stochastic process satisfying $\lim_{t\to\infty}\{X(t) - c_2 t\} = -\infty$ a.s. Then, for each $u \geq 0$,*

$$\mathbb{P}\Big(\sup_{t\geq 0}\{X(t) - c_2 t\} - \sup_{t\geq 0}\{X(t) - c_1 t\} > u\Big)$$
$$= \mathbb{P}\Big(\sup_{t\in[t_u,\infty)}\{X(t) - c_2 t\} - \sup_{t\in[0,\,t_u]}\{X(t) - c_1 t\} > u\Big).$$

PROOF: We use the short notation $Y_S^{(i)} := \sup_{t\in S}\{X(t) - c_i t\}$, for $i = 1,2$ and $S \subseteq \mathbb{R}^+$. Define also $t_i^\star := \arg\sup_{t\geq 0}\{X(t) - c_i t\}$. We have to prove that

$$\mathbb{P}(Y^{(2)}_{[0,\infty)} - Y^{(1)}_{[0,\infty)} > u) = \mathbb{P}(Y^{(2)}_{[t_u,\infty)} - Y^{(1)}_{[0,\,t_u]} > u).$$

Our proof is in four steps.

- We first prove that $Y^{(2)}_{[0,\infty)} - Y^{(1)}_{[0,\infty)} > u$ implies that $t_2^\star \geq t_u$. To show this, suppose that $t_2^\star < t_u$; then we have that

$$Y^{(2)}_{[0,\infty)} - Y^{(1)}_{[0,\infty)} = \sup_{t\in[0,\,t_u)}\Big\{X(t) - c_2 t - \sup_{s\geq 0}\{X(s) - c_1 s\}\Big\}$$
$$\leq \sup_{t\in[0,\,t_u)}\{X(t) - c_2 t - \{X(t) - c_1 t\}\} = u.$$

- Because of the first step, we can restrict ourselves to $t_2^\star \geq t_u$. Consequently, we can decompose our probability into

$$\mathbb{P}(Y^{(2)}_{[0,\infty)} - Y^{(1)}_{[0,\infty)} > u)$$
$$= \mathbb{P}(Y^{(2)}_{[t_u,\infty)} - Y^{(1)}_{[0,\infty)} > u)$$
$$= \mathbb{P}(Y^{(2)}_{[t_u,\infty)} - Y^{(1)}_{[0,\infty)} > u; t_1^\star > t_u) + \mathbb{P}(Y^{(2)}_{[t_u,\infty)} - Y^{(1)}_{[0,\infty)} > u; t_1^\star \leq t_u);$$

  call the latter probabilities $P_1$ and $P_2$, respectively.
- First consider $P_1$. Note that under $t_1^\star > t_u$, it holds that

$$Y^{(2)}_{[t_u,\infty)} - Y^{(1)}_{[0,\infty)} \geq (X(t_1^\star) - c_2 t_1^\star) - (X(t_1^\star) - c_1 t_1^\star) > u,$$

  which combined with the fact that

$$\{Y^{(2)}_{[t_u,\infty)} - Y^{(1)}_{[0,\infty)} > u\} \subseteq \{Y^{(2)}_{[t_u,\infty)} - Y^{(1)}_{[0,\,t_u]} > u\}$$

  implies that

$$P_1 = \mathbb{P}(Y^{(2)}_{[t_u,\infty)} - Y^{(1)}_{[0,\,t_u]} > u; t_1^\star > t_u).$$

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

- We trivially have

$$P_2 = \mathbb{P}(Y^{(2)}_{[t_u,\infty)} - Y^{(1)}_{[0,\,t_u]} > u; t_1^\star \le t_u).$$

Thus,

$$P_1 + P_2 = \mathbb{P}(Y^{(2)}_{[t_u,\infty)} - Y^{(1)}_{[0,\,t_u]} > u).$$

This completes the proof. ■

We now return to the class of arrival processes with stationary increments, that was introduced in Section 2.1. For this class, the above result enables us to rewrite (3) as follows.

THEOREM 2.3: *For each* $u \ge 0$,

$$\mathbb{P}(Q_2 > u) = \mathbb{P}\Big(\sup_{t\in[t_u,\infty)}\{A(t) - c_2 t\} - \sup_{t\in[0,\,t_u]}\{A(t) - c_1 t\} > u\Big).$$

PROOF: The statement follows immediately by combining (3) and Proposition 2.2. ■

If the increments of $A(\cdot)$ are also independent, then we get the following representation.

THEOREM 2.4: *Let* $\{A(t), t \in \mathbb{R}\}$ *be a stochastic process with stationary independent increments and let* $\mu = \mathbb{E}A(1) < c_2$. *Then for each* $u \ge 0$ *and* $A_1(\cdot)$ *and* $A_2(\cdot)$ *independent copies of the process* $A(\cdot)$,

$$\mathbb{P}(Q_2 > u) = \mathbb{P}\Big(\sup_{t\in[0,\infty)}\{A_1(t) - c_2 t\} > \sup_{t\in[0,\,t_u]}\{-A_2(t) + c_1 t\}\Big). \tag{4}$$

PROOF: To get the result, use

$$\begin{aligned}
&\sup_{t\in[t_u,\infty)}\{A(t) - c_2 t\} - \sup_{t\in[0,\,t_u]}\{A(t) - c_1 t\} \\
&\quad = \sup_{t\in[t_u,\infty)}\{A(t) - A(t_u) - c_2(t - t_u)\} \\
&\qquad - \sup_{t\in[0,\,t_u]}\{A(t) - A(t_u) - c_1(t - t_u)\} + (c_1 - c_2)t_u.
\end{aligned}$$

Now the theorem follows immediately from the independence and stationarity of the increments of $A(\cdot)$, in conjunction with the fact that $(c_1 - c_2)t_u = u$. ■

## 3. DISTRIBUTION OF $\mathbb{P}(Q_2 > u)$

This section studies the distribution of $\mathbb{P}(Q_2 > u)$. Direct application of the new representation of the downstream queue, as derived in the previous section, yields the Laplace transform $\mathbb{E}e^{-xQ_2}$ and the Pollaczek–Khintchine representation of the distribution of $Q_2$.

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

### 3.1. Laplace Transform

In this subsection we assume that $A(\cdot) \equiv \{A(t), t \in \mathbb{R}\}$ is a spectrally positive Lévy process (see, e.g., [35]) (thus allowing positive jumps). Hence, its Laplace transform is uniquely determined by the Laplace exponent $\kappa(\cdot)$ through

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

$s \geq 0$; see, for instance, [4, Chap. IX.9].

We need the following additional notation:

$$S_1 := \sup_{t \geq 0}\{A_1(t) - c_2 t\},$$

$$S_{2,T} := \sup_{t \in [0,T]}\{-A_2(t) + c_1 t\},$$

where $A_1(\cdot)$ and $A_2(\cdot)$ are independent copies of the process $A(\cdot)$. So we can rewrite formula (4) in Theorem 2.4, with $t_u := u/(c_1 - c_2)$, as

$$\mathbb{P}(Q_2 \leq u) = \mathbb{P}(S_1 - S_{2,t_u} \leq 0). \tag{5}$$

It appears that the function $\theta_1(\cdot)$, with

$$\theta_1(s) := \log(\mathbb{E}e^{-sA(1)+sc_1}),$$

is of crucial importance in the further analysis; observe that $\theta_1(s) = \kappa(s) + c_1 s$. Additionally, let $R_y$ denote the first-passage time of level $y$: $R_y := \inf\{t \geq 0 : -A_2(t) + c_1 t > y\}$, so that

$$\mathbb{P}(S_{2,t_u} < y) = \mathbb{P}(R_y > t_u) \tag{6}$$

for each $y \geq 0$. We first summarize some useful properties of $\theta_1(\cdot)$ and $R_y$.

PROPOSITION 3.1: *Let $\{A(t), t \in \mathbb{R}\}$ be a spectrally positive Lévy process with Laplace exponent $\kappa(\cdot)$ and $\mu := \mathbb{E}A(1) < c_2$. Then the following hold:*

*(i)* *$\theta_1(s)$ is strictly increasing on $s \in [0,\infty)$.*
*(ii)* *$\theta_1(s)/s \to c_1 - \mu$ as $s \downarrow 0$.*
*(iii)* *$\mathbb{E}e^{-sR_y} = e^{-y\theta_1^{-1}(s)}$ for $s \geq 0$ and $y \geq 0$, where $\theta_1^{-1}(\cdot)$ is the inverse function of $\theta_1(\cdot)$.*

PROOF: Part (i) follows from the fact that $\theta_1'(0+) = -\mathbb{E}A(1) + c_1 = -\mu + c_1 > 0$, in conjunction with the well-known property that $\kappa(\cdot)$ is convex. Part (ii) is a consequence of $\theta_1(s)/s \to \theta_1'(0+)$ as $s \downarrow 0$. Part (iii) is an immediate implication of Sato [35, Thm. 46.3] (see also, e.g., [8,19]). ■

In the following theorem we obtain an explicit formula for the Laplace transform of the downstream queue $Q_2$.

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

THEOREM 3.2: *Let $\{A(t), t \in \mathbb{R}\}$ be a spectrally positive Lévy process with Laplace exponent $\kappa(\cdot)$ and $\mu := \mathbb{E}A(1) < c_2$. Then, for each $x > 0$,*

$$\mathbb{E}e^{-xQ_2} = \frac{c_2 - \mu}{c_1 - c_2} \cdot \frac{\theta_1^{-1}(x(c_1 - c_2))}{x - \theta_1^{-1}(x(c_1 - c_2))}.$$

PROOF: Let $x > 0$. In the first step, we prove that

$$\mathbb{E}e^{-xQ_2} = \mathbb{E}e^{-\theta_1^{-1}(x(c_1 - c_2))S_1}. \tag{7}$$

Note that integration by parts yields

$$\begin{aligned}\mathbb{E}e^{-xQ_2} &= \int_0^\infty e^{-xu}\, d\mathbb{P}(Q_2 \le u) = x\int_0^\infty e^{-xu}\mathbb{P}(Q_2 \le u)\, du \\ &= 1 - x\int_0^\infty e^{-xu}\mathbb{P}(Q_2 > u)\, du. \end{aligned} \tag{8}$$

Moreover, due to (5) and (6) we have

$$\begin{aligned}\int_0^\infty e^{-xu}\mathbb{P}(Q_2 > u)\, du &= \int_0^\infty e^{-xu}\mathbb{P}(S_1 - S_{2,t_u} > 0)\, du \\ &= \int_0^\infty e^{-xu}\int_0^\infty \mathbb{P}(S_{2,t_u} < y)\, d\mathbb{P}(S_1 \le y)\, du \\ &= \int_0^\infty e^{-xu}\int_0^\infty \mathbb{P}(R_y > t_u)\, d\mathbb{P}(S_1 \le y)\, du.\end{aligned}$$

Applying the change of variable $v := u/(c_1 - c_2)$, this leads to

$$\begin{aligned}&\int_0^\infty e^{-xu}\mathbb{P}(Q_2 > u)\, du \\ &= (c_1 - c_2)\int_0^\infty\int_0^\infty e^{-x(c_1-c_2)v}\mathbb{P}(R_y > v)\, dv\, d\mathbb{P}(S_1 \le y) \\ &= (c_1 - c_2)\int_0^\infty\left(\frac{1}{x(c_1 - c_2)} - \int_0^\infty e^{-x(c_1-c_2)v}\mathbb{P}(R_y \le v)\, dv\right) d\mathbb{P}(S_1 \le y) \\ &= \frac{1}{x}\left(1 - \int_0^\infty\int_0^\infty e^{-x(c_1-c_2)v}\, d\mathbb{P}(R_y \le v)\, d\mathbb{P}(S_1 \le y)\right) \\ &= \frac{1}{x}\left(1 - \int_0^\infty e^{-y\theta_1^{-1}(x(c_1-c_2))}\, d\mathbb{P}(S_1 \le y)\right) \qquad (9) \\ &= \frac{1}{x}\left(1 - \mathbb{E}e^{-\theta_1^{-1}(x(c_1-c_2))S_1}\right), \qquad (10)\end{aligned}$$

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

where (9) is due to Proposition 3.1(iii). Now (7) follows directly from combining (8) with (10).

In order to complete the proof, note that, following Zolotarev's result for spectrally positive Lévy processes [42], for $x > 0$

$$\mathbb{E}e^{-xS_1} = \frac{c_2 - \mu}{\theta_2(x)}\, x, \tag{11}$$

where $\theta_2(s) := \log(\mathbb{E}e^{-sA(1)+sc_2})$; observe that $\theta_2(s) = \kappa(s) + c_2 s = \theta_1(s) + (c_2 - c_1)s$. Combining (7) with (11) leads to

$$\begin{aligned}\mathbb{E}e^{-xQ_2} &= \frac{(c_2 - \mu)\theta_1^{-1}(x(c_1 - c_2))}{\theta_2(\theta_1^{-1}(x(c_1 - c_2)))} \\ &= \frac{c_2 - \mu}{c_1 - c_2}\, \frac{\theta_1^{-1}(x(c_1 - c_2))}{x - \theta_1^{-1}(x(c_1 - c_2))}.\end{aligned}$$

This completes the proof. ■

*Remark 3.3:* Theorem 3.2 can be considered as an analogue of the result of Zolotarev [42]; there, the Laplace transform of $\mathbb{P}(Q_1 < u)$ was given for $A(\cdot)$ being a spectrally positive Lévy process, whereas here we present the corresponding result for $\mathbb{P}(Q_2 < u)$.

*Remark 3.4:* Recall that $Q$ denotes the (steady state) of the total workload. Since $S_1 =_{\mathrm{d}} Q$, it follows that (7) is equivalent to

$$\mathbb{E}e^{-xQ_2} = \mathbb{E}e^{-\theta_1^{-1}(x(c_1-c_2))Q}, \tag{12}$$

for $x \geq 0$. This relation plays a crucial role in Proposition 5.1.

*Remark 3.5:* We note that for special cases, Theorem 3.2 can be derived in alternative ways. Consider for instance the case of *compound Poisson* input. We assume that $A(t) = \sum_{i=1}^{N(t)} X_i$, where $N(\cdot)$ is a Poisson process with rate $\lambda$. The random variables $X_i$ are i.i.d., nonnegative, and independent of $N(\cdot)$. We denote by $X$ the generic random variable and by $F(\cdot)$ its distribution function. Under the above notation, we say that $A(\cdot)$ is a compound Poisson process, characterized by $(\lambda, F)$. Let $f := \mathbb{E}X$.

Observe that in the tandem setting, the downstream queue is fed by an on–off process: During on-periods, which are distributed as the busy period of the first queue, the downstream queue fills at rate $c_1 - c_2$, whereas during off-periods, which have an exponential distribution (with mean $\lambda^{-1}$), it drains at rate $c_2$. Assume for ease that $\mu = \lambda f < 1 = c_2 < 2 = c_1$ (the analysis can be extended straightforwardly to the case of general service rates). Under this choice, our formula for $\mathbb{E}e^{-xQ_2}$ reduces to

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

$$\mathbb{E}e^{-xQ_2} = (1-\mu)\,\frac{\theta_1^{-1}(x)}{x-\theta_1^{-1}(x)}.$$

Denoting by $P$ the busy period of the first queue, and $\pi(s) := \mathbb{E}e^{-sP}$ and $\xi(s) := \mathbb{E}e^{-sX}$, then [2, Eq. (3)] gives

$$\mathbb{E}e^{-xQ_2} = \frac{1-\lambda\mathbb{E}P}{1+\lambda\mathbb{E}P}\,\frac{x+\lambda-\lambda\pi(x)}{x-\lambda+\lambda\pi(x)}.$$

It is now left to check that the above two formulas coincide. As the input rate of the first queue equals the output rate of the first queue,

$$\mu = \lambda f = \frac{\mathbb{E}P}{1/\lambda+\mathbb{E}P}\cdot 2,$$

we obtain that $1-\mu = (1-\lambda\mathbb{E}P)/(1+\lambda\mathbb{E}P)$. Hence, it is left to show that

$$\frac{\theta_1^{-1}(x)}{x-\theta_1^{-1}(x)} = \frac{x+\lambda-\lambda\pi(x)}{x-\lambda+\lambda\pi(x)},$$

or, equivalently, $2\theta_1^{-1}(x) = x+\lambda-\lambda\pi(x)$ or

$$\theta_1\left(\frac{x}{2}+\frac{\lambda}{2}-\frac{\lambda}{2}\,\pi(x)\right) = x. \tag{13}$$

By definition, we have $\theta_1(x) = -\lambda+\lambda\xi(x)+2x$. Also, we have the well-known fixed-point equation for the busy period of the $M/G/1$ queue (with service rate 2):

$$\pi(x) = \xi\left(\frac{x}{2}+\frac{\lambda}{2}-\frac{\lambda}{2}\,\pi(x)\right).$$

Hence, we obtain (13):

$$\theta_1\left(\frac{x}{2}+\frac{\lambda}{2}-\frac{\lambda}{2}\,\pi(x)\right) = -\lambda+\lambda\xi\left(\frac{x}{2}+\frac{\lambda}{2}-\frac{\lambda}{2}\,\pi(x)\right)+x+\lambda-\lambda\pi(x) = x,$$

as desired. Related results can be found in [24,36].

### 3.2. Pollaczek–Khintchine Representation

Theorem 3.2 enables us to determine Pollaczek–Khintchine representation for the distribution function of $Q_2$. Define $\tilde{R}_y := (c_1-c_2)R_y$ and observe that, due to Proposition 3.1, $\{\tilde{R}_y : y \geq 0\}$ is a subordinator with $\mathbb{E}e^{-x\tilde{R}_y} = e^{-y\vartheta(x)}$, where $\vartheta(x) := \theta_1^{-1}((c_1-c_2)x)$ for $y \geq 0$ and $x \geq 0$. Recalling Proposition 3.1(ii), we have

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

$$\varrho := \lim_{x \downarrow 0} \frac{\vartheta(x)}{x} = \frac{c_1 - c_2}{c_1 - \mu},$$

with $\varrho \in (0,1)$.

Thus, following, for example, Takács [38, Eq. (23)], define a distribution function $H(\cdot)$ such that $H(x) = 0$ for $x < 0$ and its Laplace transform $\hat{\ell}_H(s)$ is of the form

$$\hat{\ell}_H(x) = \int_0^\infty e^{-xv}\, dH(v) = \frac{\vartheta(x)}{\varrho x} \tag{14}$$

for $x \geq 0$.

THEOREM 3.6: *Let $\{A(t), t \in \mathbb{R}\}$ be a spectrally positive Lévy process with Laplace exponent $\kappa(\cdot)$ and $\mu := \mathbb{E}A(1) < c_2$. Then*

$$\mathbb{P}(Q_2 \leq u) = (1 - \varrho) \sum_{i=1}^{\infty} \varrho^{i-1} H^{\star i}(u).$$

PROOF: Application of Theorem 3.2 immediately yields

$$\begin{aligned}
\mathbb{E}e^{-xQ_2} &= \frac{c_2 - \mu}{c_1 - c_2} \frac{\vartheta(x)}{x - \vartheta(x)} \\
&= \frac{c_2 - \mu}{c_1 - c_2} \frac{\varrho \dfrac{\vartheta(x)}{\varrho x}}{1 - \varrho \dfrac{\vartheta(x)}{\varrho x}} \\
&= \frac{c_2 - \mu}{c_1 - c_2} \sum_{i=1}^{\infty} (\varrho \hat{\ell}_H(x))^i \\
&= (1 - \varrho) \sum_{i=1}^{\infty} \varrho^{i-1} \hat{\ell}_H^{\,i}(x).
\end{aligned}$$

This implies Theorem 3.6 immediately. ■

*Remark 3.7:* Theorem 3.6 is related to Theorem 4.1 in [25], where it was assumed that the input process is a compound Poisson.

*Remark 3.8:* The distribution $H(\cdot)$ has a natural representation in the language of the Lévy measure associated with $\{\tilde{R}_y : y \geq 0\}$. Indeed, since $\{\tilde{R}_y : y \geq 0\}$ is a subordinator, then, following [7, Sect. 3.1, Eq. (3)], the *Lévy–Khintchine* representation of the Laplace exponent $\vartheta(\cdot)$ has the following form:

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

$$\vartheta(x) = dx + x\int_0^\infty e^{-xs}\bar{\Pi}(s)\,ds,$$

where $d$ is the drift coefficient and $\bar{\Pi}(s) = \Pi((s,\infty))$ is the tail of the Lévy measure. Then

$$\hat{\ell}_H(x) = \frac{d}{\varrho} + \frac{1}{\varrho}\int_0^\infty e^{-xs}\bar{\Pi}(s)\,ds.$$

Hence,

$$H(t) = \frac{d}{\varrho} + \frac{1}{\varrho}\int_0^t \bar{\Pi}(s)\,ds$$

for all $t \geq 0$. Observe also that $\varrho = d + \int_0^\infty \bar{\Pi}(s)\,ds$.

## 4. EXAMPLES: LIGHT-TAILED CASE

In this section we apply results of previous sections to two special cases. The first focuses on Brownian input and the second focuses on Poisson input; hence, in all cases, the increments are independent and, consequently, (4) can be used. Before analyzing particular examples, let us first identify a "dichotomy" in the distribution of $Q_2$.

### 4.1. A Dichotomy

If $\kappa(s)$ exists for negative $s$, the Lévy process could be called "light-tailed," as the tail of the distribution of $A(1)$ decays exponentially or faster. In this section we use the results of the previous section to further investigate this light-tailed case, eventually leading to an interesting dichotomy. Throughout this subsection, we tacitly assume that the Lévy input is spectrally positive.

Bearing in mind Proposition 3.1(iii), it is seen that by analytic continuation, Theorem 3.2 holds for all $x$ for which $\theta_1^{-1}(x(c_1 - c_2))$ is well defined. For light-tailed Lévy processes, this means that we could pick any $x$ larger than $x_1 := (c_1 - c_2)^{-1}\min_{y\in\mathbb{R}}\theta_1(y)$ (use that $\theta_1(\cdot)$ is convex). Theorem 3.2 implies that if, in addition, $x > \theta_1^{-1}(x(c_1 - c_2))$, we have that $\mathbb{E}e^{-xQ_2}$ is finite and, hence, "Chernoff" implies that

$$\mathbb{P}(Q_2 > u) \leq Me^{xu} \tag{15}$$

for some finite constant $M$; see also [1, Lemma 3.2]. Let $x_2$ solve $x_2 = \theta_1^{-1}(x_2(c_1 - c_2))$, or, equivalently, $\theta_2(x) = 0$; if there is no solution, then take the infimum of all $x$ such that $\theta_2(x) \leq 0$ (this $x_2$ is necessarily negative). The above arguments immediately lead to the following result, yielding the tightest $x$ for which (15) holds.

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

PROPOSITION 4.1: *If $x_2 > \arg\min_y \theta_1(y)$ then (15) holds (for some $M < \infty$) for all $x > x_1$; else (15) holds (for some $M < \infty$) for all $x > x_2$.*

Assuming that $\kappa(\cdot)$ is differentiable, define $c_1^\star := -\kappa'(x_2)$; note that $x_2$ depends on $c_2$ only (not on $c_1$) and, consequently, $c_1^\star$ depends on $c_2$ only. The next result indicates that $c_1^\star$ can be viewed as a *critical service rate*, demarcating two regimes.

PROPOSITION 4.2: *If $c_1 < c_1^\star$, then (15) holds (for some $M < \infty$) for all $x > x_1$; else (15) holds (for some $M < \infty$) for all $x > x_2$.*

PROOF: Due to the convexity of $\theta_1(\cdot)$, requiring $x_2 > \arg\inf_y \theta_1(y)$ is equivalent to requiring $\theta_1'(x_2) > 0$. This implies the stated immediately. ■

The above indicates that there are in fact two regimes: There is essentially different behaviors for $c_1 > c_1^\star$ and $c_1 \le c_1^\star$. It is noted that in the first case, the exponential bound (with decay rate $x_2$) coincides with the traditional Kingman-type bound of a single queue with service rate $c_2$. We return to this phenomenon later; there we also give an interpretation of the two regimes.

### 4.2. Tandem with Brownian Motion Input

In this subsection we analyze representation (4) for the case where $A(\cdot) = B(\cdot)$, a standard Brownian motion (with zero drift). It appears that an explicit formula can be derived for $\mathbb{P}(Q_2 > u)$. Let $\Psi(\cdot)$ denote the tail distribution of a standard Normal random variable $N$: $\Psi(x) = \mathbb{P}(N > x)$.

THEOREM 4.3: *Let $A(\cdot)$ be a standard Brownian motion. Then, for each $u \ge 0$,*

$$\mathbb{P}(Q_2 > u) = \frac{c_1 - 2c_2}{c_1 - c_2} e^{-2c_2 u} \left(1 - \Psi\left(\frac{c_1 - 2c_2}{\sqrt{c_1 - c_2}}\sqrt{u}\right)\right) + \frac{c_1}{c_1 - c_2} \Psi\left(\frac{c_1}{\sqrt{c_1 - c_2}}\sqrt{u}\right).$$

PROOF: Following Theorem 2.4 we have

$$\mathbb{P}(Q_2 > u) = \mathbb{P}\left(\sup_{t\in[0,\infty)} \{B_1(t) - c_2 t\} > \sup_{t\in(0,t_u]} \{B_2(t) + c_1 t\}\right),$$

where $B_1(\cdot)$ and $B_2(\cdot)$ are independent standard Brownian motions. Using the well-known fact that, for each $x \ge 0$,

$$\mathbb{P}\left(\sup_{t\in[0,\infty)} \{B_1(t) - c_2 t\} > x\right) = e^{-2c_2 x}$$

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

and (see, e.g., Baxter and Donsker [6])

$$\rho(x) := \frac{d}{dx}\mathbb{P}\left(\sup_{t\in(0,t_u]}\{B_2(t)+c_1 t\}\le x\right)$$

$$= \sqrt{\frac{2}{\pi t_u}}\exp\left(-\frac{(x-c_1 t_u)^2}{2t_u}\right) - 2c_1 e^{2c_1 x}\Psi\left(\frac{x+c_1 t_u}{\sqrt{t_u}}\right),$$

we get two integrals, which we will calculate separately:

$$\mathbb{P}(Q_2 > u) = \int_0^\infty e^{-2c_2 x}\rho(x)\,dx = I_1(u) - I_2(u),$$

where

$$I_1(u) := \int_0^\infty e^{-2c_2 x}\sqrt{\frac{2}{\pi t_u}}\exp\left(-\frac{(x-c_1 t_u)^2}{2t_u}\right)dx,$$

$$I_2(u) := \int_0^\infty e^{-2c_2 x}2c_1 e^{2c_1 x}\Psi\left(\frac{x+c_1 t_u}{\sqrt{t_u}}\right)dx.$$

Integral $I_1(u)$ is evaluated as follows:

$$I_1(u) = 2\int_0^\infty \frac{1}{\sqrt{2\pi t_u}}\exp\left(-\frac{x^2-2c_1 t_u x + c_1^2 t_u^2 + 4c_2 x t_u}{2t_u}\right)dx$$

$$= 2e^{-c_2(c_1-c_2)t_u}\int_0^\infty \frac{1}{\sqrt{2\pi t_u}}\exp\left(-\frac{(x-(c_1-2c_2)t_u)^2}{2t_u}\right)dx$$

$$= 2e^{-2c_2 u}\left(1-\Psi\left(\frac{c_1-2c_2}{\sqrt{c_1-c_2}}\sqrt{u}\right)\right). \qquad \textbf{(16)}$$

For $I_2(u)$, we find

$$I_2(u) = \int_0^\infty 2c_1 e^{2x(c_1-c_2)}\Psi\left(\frac{x+c_1 t_u}{\sqrt{t_u}}\right)dx$$

$$= \frac{c_1}{c_1-c_2}\int_0^\infty 2(c_1-c_2)e^{2x(c_1-c_2)}\Psi\left(\frac{x+c_1 t_u}{\sqrt{t_u}}\right)dx.$$

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

Now integration by parts gives that $I_2(u) = c_1\tilde{I}_2(u)/(c_1 - c_2)$, with

$$\begin{aligned}
\tilde{I}_2(u) &= -\Psi(c_1\sqrt{t_u}) + \int_0^\infty e^{2x(c_1-c_2)} \frac{1}{\sqrt{2\pi t_u}} \exp\left(-\frac{(x + c_1 t_u)^2}{2t_u}\right) dx \\
&= -\Psi(c_1\sqrt{t_u}) + \exp\left(-\frac{c_1^2 t_u}{2} + \frac{(c_1 - 2c_2)^2 t_u}{2}\right) \\
&\quad \times \int_0^\infty \frac{1}{\sqrt{2\pi t_u}} \exp\left(-\frac{x^2 - 2xt_u(c_1 - 2c_2) + (c_1 - 2c_2)^2 t_u^2}{2t_u}\right) dx \\
&= -\Psi(c_1\sqrt{t_u}) + e^{-c_2(2c_1-2c_2)t_u} \int_0^\infty \frac{1}{\sqrt{2\pi t_u}} \exp\left(-\frac{(x - (c_1 - 2c_2)t_u)^2}{2t_u}\right) dx \\
&= -\Psi(c_1\sqrt{t_u}) + e^{-2c_2 u}\left(1 - \Psi\left(\frac{c_1 - 2c_2}{\sqrt{c_1 - c_2}}\sqrt{u}\right)\right). \qquad \textbf{(17)}
\end{aligned}$$

Combining (16) and (17) yields the desired result. ■

The next corollary finds the exact asymptotics of the buffer content distribution in the second node, by applying the general result of Theorem 4.3.

COROLLARY 4.4: *Let $A(\cdot)$ be a standard Brownian motion.*

*(i) If $c_1 > 2c_2$, then*

$$\mathbb{P}(Q_2 > u) = \frac{c_1 - 2c_2}{c_1 - c_2} e^{-2c_2 u}(1 + o(1)) \quad \text{as } u \to \infty.$$

*(ii) If $c_1 = 2c_2$, then*

$$\mathbb{P}(Q_2 > u) = \frac{1}{\sqrt{2\pi c_2}} \frac{1}{\sqrt{u}} e^{-2c_2 u}(1 + o(1)) \quad \text{as } u \to \infty.$$

*(iii) If $c_1 < 2c_2$, then*

$$\begin{aligned}
\mathbb{P}(Q_2 > u) = \frac{1}{\sqrt{2\pi}} \left(\frac{c_1 - c_2}{u}\right)^{3/2} \frac{4c_2}{c_1^2(c_1 - 2c_2)^2} \\
\times \exp\left(-\frac{c_1^2}{2(c_1 - c_2)} u\right)(1 + o(1)) \quad \text{as } u \to \infty.
\end{aligned}$$

PROOF: Let $J_1(u)$ and $J_2(u)$ be defined as

$$J_1(u) := \frac{c_1 - 2c_2}{c_1 - c_2} e^{-2c_2 u}\left(1 - \Psi\left(\frac{c_1 - 2c_2}{\sqrt{c_1 - c_2}}\sqrt{u}\right)\right)$$

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

and

$$J_2(u) := \frac{c_1}{c_1 - c_2}\,\Psi\left(\frac{c_1}{\sqrt{c_1 - c_2}}\sqrt{u}\right);$$

recall from Theorem 4.3 that $\mathbb{P}(Q_2 > u) = J_1(u) + J_2(u)$. We also recall the standard "first-order" asymptotic result,

$$\Psi(u) = \frac{1}{\sqrt{2\pi}}\,\frac{1}{u}\,e^{-u^2/2}(1 + o(1)), \tag{18}$$

and "second-order" result,

$$\frac{1}{\sqrt{2\pi}}\,\frac{1}{u}\,e^{-u^2/2} - \Psi(u) = \frac{1}{\sqrt{2\pi}}\,\frac{1}{u^3}\,e^{-u^2/2}(1 + o(1)). \tag{19}$$

We now consider the three cases separately.

(i) $c_1 > 2c_2$. Then

$$\lim_{u\to\infty}\frac{\log J_1(u)}{\log J_2(u)} = \frac{c_1^2}{4c_2(c_1 - c_2)} = \frac{(c_1/c_2)^2}{4(c_1/c_2 - 1)} > 1,$$

due to $x^2 > 4(x-1) > 0$ for $x > 2$. The proof follows after applying (18).

(ii) $c_1 = 2c_2$. Then $J_1(u) = 0$, and the proof follows from (18).

(iii) $c_1 < 2c_2$. In this case, remarkably, both $-J_1(u)$ and $J_2(u)$ equal, as $u \to \infty$,

$$\sqrt{\frac{1}{c_1 - c_2}}\,\frac{1}{\sqrt{2\pi}}\,\frac{1}{\sqrt{u}}\exp\left(-\frac{c_1}{2(c_1 - c_2)}\,u\right)(1 + o(1)),$$

due to (18). As a consequence, we have to rely on the more precise asymptotics (19). After tedious computations we derive the corollary. ■

We see that there are two regimes: With $c_1^\star := 2c_2$, there is essentially different behavior for $c_1 > c_1^\star$ and $c_1 \le c_1^\star$; note that, in the terminology of Section 4.1, we indeed have that $c_1^\star = -\kappa'(x_2)$ (with $\kappa(s) = \frac{1}{2}s^2$ and $x_2 = -2c_2$). In the first case, upon overflow in the second queue, the first queue remains essentially empty and, consequently, the large-buffer asymptotics resemble those of a single queue with service rate $c_2$. In the second case, the first queue will be exactly "balanced," as traffic is generated at about a rate of $c_1$ during $t_u$ units of time.

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

### 4.3. Tandem with Poisson Input

Now we assume that $A_i(t)$, with $i = 1,2$, is Poisson distributed with mean $\lambda t$ (for some $\lambda \in (0, c_2)$). We first determine

$$\mathbb{P}\Big(\sup_{t\in[0,\infty)}\{A_1(t) - c_2 t\} > x\Big) \quad \text{and} \quad \mathbb{P}\Big(\sup_{t\in[0,\,t_u]}\{-A_2(t) + c_1 t\} \geq x\Big).$$

Consider the first of these probabilities. For instance, using the Beneš approach (see, e.g., [31, Eq. (15.1.3)]) with $\rho := \lambda/c_2$,

$$\zeta(x) := \mathbb{P}\Big(\sup_{t\in[0,\infty)}\{A_1(t) - c_2 t\} > x\Big) = (1-\rho)\sum_{n=\lceil x\rceil}^{\infty}\frac{(\rho(n-x))^n}{n!}\,e^{-\rho(n-x)}. \quad \textbf{(20)}$$

Now consider the second probability. Evidently,

$$\eta(x) := \mathbb{P}\Big(\sup_{t\in[0,\,t_u]}\{-A_2(t) + c_1 t\} \geq x\Big) = \mathbb{P}(\exists t \in [0,t_u]: A_2(t) \leq c_1 t - x).$$

This expression can be computed by conditioning on the value of $A_2(t_u)$, as follows:

$$\begin{aligned}&\mathbb{P}(\exists t \in [0,t_u]: A_2(t) \leq c_1 t - x)\\ &\quad = \mathbb{P}(A_2(t_u) \leq c_1 t_u - x)\\ &\qquad + \sum_{i=\lceil c_1 t_u - x\rceil}^{\infty}\mathbb{P}(\exists t \in [0,t_u]: A_2(t) \leq c_1 t - x \,|\, A_2(t_u) = i)\,\mathbb{P}(A_2(t_u) = i).\end{aligned}$$

Obviously,

$$\mathbb{P}(A_2(t_u) \leq c_1 t_u - x) = \sum_{i=0}^{\lfloor c_1 t_u - x\rfloor}\frac{(\lambda t_u)^i}{i!}\,e^{-\lambda t_u}; \qquad \mathbb{P}(A_2(t_u) = i) = \frac{(\lambda t_u)^i}{i!}\,e^{-\lambda t_u}.$$

Finally, focus on $\mathbb{P}(\exists t \in [0,t_u]: A_2(t) \leq c_1 t - x \,|\, A_2(t_u) = i)$. Reversing time, it is easily seen that this probability coincides with

$$\mathbb{P}(\exists t \in [0,t_u]: A_2(t) \geq i + x - c_1 t_u + c_1 t \,|\, A_2(t_u) = i). \quad \textbf{(21)}$$

Given $A_2(t_u) = i$, these $i$ arrivals are uniformly distributed on $[0,t_u]$. Hence, probability (21) equals $p(i + x - c_1 t_u, c_1 | i, t_u)$, where $p(x,c|N,D)$ is the overflow probability over level $x$ in a queue with $N$ periodic inputs (with period $D$) and service rate $c$ (which is sometimes referred to as the $N{\cdot}D/D/1$ queue). An explicit expression for $p(x,c|N,D)$ is given in the following lemma.

LEMMA 4.5: *For $x \in [0,N]$ and $N < x + c{\cdot}D$, it holds that*

$$p(x,c|N,D) = \sum_{n\in\mathbb{N}:\,n\in(x,N]}\mathbb{B}\mathrm{in}\left(n\,\middle|\,N,\frac{n-x}{c{\cdot}D}\right)\frac{cD - N + x}{cD - n + x},$$

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

*with*

$$\mathbb{B}\mathrm{in}(n|N,p) := \binom{N}{n} p^n (1-p)^{N-n}.$$

*If* $N \geq x + c\cdot D$, *then* $p(x,c|N,D) = 0$.

PROOF: See Humblet, Bhargava, and Hluchyj [21, Sect. III.C], and Roberts and Virtamo [32]. ■

This leads to the following expression:

$$\eta(x) = \mathbb{P}\left(\sup_{t\in[0,\,t_u]} \{-A_2(t) + c_1 t\} > x\right)$$

$$= \sum_{i=0}^{\lfloor c_1 t_u - x\rfloor} \frac{(\lambda t_u)^i}{i!}\, e^{-\lambda t_u} + \sum_{i=\lceil c_1 t_u - x\rceil}^{\infty} p(i + x - c_1 t_u, c_1 | i, t_u)\, \frac{(\lambda t_u)^i}{i!}\, e^{-\lambda t_u}. \quad \textbf{(22)}$$

THEOREM 4.6: *Let* $A(\cdot)$ *be a Poisson process with rate* $\lambda \in (0, c_2)$. *Then, for each* $u \geq 0$,

$$\mathbb{P}(Q_2 > u) = -\int_{[0,\infty)} \zeta(x)\, d\eta(x), \quad \textbf{(23)}$$

*where* $\zeta(\cdot)$ *and* $\eta(\cdot)$ *are given by (20) and (22), respectively.*

*Remark 4.7:* As in Corollary 4.4, one could pursue the exact distribution of $Q_2$, as in Theorem 4.6, to find its large-buffer asymptotics. However, this turns out to be a nontrivial task, due to the complicated structure of (23).

An alternative is to use a procedure that is similar to the one pursued in [1] for the priority queue. First, realize that the output process of the first queue consists of independent on- and off-periods. The off-periods are obviously exponential (with mean $\lambda^{-1}$), whereas the on-periods correspond to the duration of the busy period, say $B$, in the first queue. The asymptotics of the busy period are known (Cox and Smith [13, Sect. 5.6]), and are of the type

$$\mathbb{P}(B > u) = \alpha u^{-3/2} e^{-\beta u}(1 + o(1))$$

for positive numbers $\alpha, \beta$, and $u \to \infty$. Standard results for the large-buffer asymptotics of queues fed by a single on–off source now indicate that, as in the Brownian case, there are two regimes, separated by a critical service rate.

The critical service rate can be found as follows. Let $\theta^\star > 0$ solve $\log \mathbb{E}e^{\theta A(t)} = c_2\theta t$, or, equivalently, $\lambda(e^\theta - 1) = c_2\theta$, and define $\lambda^\star := \lambda e^{\theta^\star}$. If $c_1 > c_1^\star := \lambda^\star$, then the first queue is essentially transparent and the large-buffer asymptotics are purely exponential: $\alpha_+ e^{-\beta_+ u}$, with $\beta_+ := \theta^\star$. For $c_1 < c_1^\star$, the large-buffer asymptotics are of the form $\alpha_- u^{-3/2} e^{-\beta_- u}$; in this case,

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

$$\beta_- = \left(\frac{1}{c_1 - c_2}\right)\left(c_1 \log \frac{c_1}{\lambda} - c_1 + \lambda\right). \tag{24}$$

We recognize in (24) the Kullback–Leibler distance of a Poisson$(c_1)$ random variable with respect to a Poisson$(\lambda)$ random variable. Indeed, (24) can be interpreted as the decay rate of the Poisson source generating traffic at rate $c_1$ during $t_u$ units of time, thus building up an amount of traffic $u$ in the downstream queue.

*Remark 4.8:* The approach presented in this subsection is not restricted to the Poisson input of batches of deterministic (unit) size. Let us, for instance, consider the situation of a Poisson$(\lambda)$ stream of nonnegative i.i.d. batches $(A_i)_{i\in\mathbb{N}}$. It turns out that Theorem 1 of Takács [39, p. 24] is quite useful here. To make this theorem directly applicable, assume for the moment that time is slotted (i.e., the number of batches arriving per slot has a Poisson distribution with mean $\lambda$), $c_1 = 1$, and that the batch sizes are i.i.d. and take on nonnegative integral values.

We can repeat the above argument. Now $\zeta(\cdot)$ is the overflow probability in the corresponding $M/G/1$ queue, which is explicitly given by the Pollaczek–Khintchine formula; see, for instance, [4, Thm. 5.7, Eq. (5.5)]. Also, $\eta(\cdot)$ follows easily: We can use exactly the same procedure as earlier, by first conditioning on the value of $A_2(t_u)$; note that the per-slot arrivals are interchangeable random variables, such that the counterpart of probability (21) follows immediately from [39, Thm. 1, p. 24].

## 5. EXAMPLES: HEAVY-TAILED CASE

In this section we analyze the asymptotics of the tail distribution of $Q_2$ in a case when $A(\cdot)$ possesses a *heavy-tailed structure*.

Before we focus on particular examples, we need the following proposition, which relates the asymptotic behavior of $\mathbb{P}(Q_2 > u)$ with $\mathbb{P}(Q > u)$; recall that $Q$ represents the total workload.

PROPOSITION 5.1: *Let* $v \in (0,1)$ *and*

$$\mathbb{P}(Q > u) = u^{-v}\ell(u)(1 + o(1))$$

*as* $u \to \infty$, *where* $\ell(\cdot)$ *is slowly varying at* $\infty$. *Then*

$$\mathbb{P}(Q_2 > u) = \varrho^v u^{-v}\ell(u)(1 + o(1))$$

*as* $u \to \infty$.

PROOF: Assume that as $u \to \infty$, $\mathbb{P}(Q > u) = u^{-v}\ell(u)(1 + o(1))$ for $v \in (0,1)$. Then, following Lemma 2.2 in [9], we have

$$\mathbb{E}e^{-xQ} - 1 = \Gamma(1 - v)x^v\ell\left(\frac{1}{x}\right)(1 + o(1))$$

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

as $x \to 0$. Thus, the use of (12) leads to

$$\begin{aligned}\mathbb{E}e^{-xQ_2} - 1 &= \mathbb{E}e^{-\theta_1^{-1}(x(c_1-c_2))Q} - 1\\ &= \Gamma(1-v)\theta_1^{-1}(x(c_1-c_2))^v \ell\left(\frac{1}{\theta_1^{-1}(x(c_1-c_2))}\right)(1+o(1))\\ &= \Gamma(1-v)\left(\frac{\theta_1^{-1}(x(c_1-c_2))}{x}\right)^v x^v \ell\left(\frac{x}{\theta_1^{-1}(x(c_1-c_2))}\frac{1}{x}\right)(1+o(1))\\ &= \Gamma(1-v)\varrho^v x^v \ell\left(\frac{1}{x}\right)(1+o(1))\end{aligned}$$

as $x \to 0$, since $\lim_{x\downarrow 0} \theta_1^{-1}(x(c_1-c_2))/x = \varrho$. Thus, again, the use of [9, Lemma 2.2] completes the proof. ■

In the following proposition we give the exact asymptotics of the tail distribution of $Q_2$ in a case when $H(\cdot)$ belongs to the class of subexponential distributions.

PROPOSITION 5.2: *Let $\{A(t), t \in \mathbb{R}\}$ be a spectrally positive Lévy process with $\mu := \mathbb{E}A(1) < c_2$. If $H(\cdot)$ defined in (14) is subexponential, then*

$$\mathbb{P}(Q_2 > u) = \frac{c_1 - \mu}{c_2 - \mu}\bar{H}(u)(1+o(1))$$

*as $u \to \infty$.*

PROOF: From Theorem 3.6, we have that for each $u \geq 0$,

$$\begin{aligned}\mathbb{P}(Q_2 \leq u) &= (1-\varrho)\sum_{i=1}^{\infty}\varrho^{i-1}H^{\star i}(u)\\ &= (1-\varrho)\sum_{i=0}^{\infty}\varrho^i(H^{\star i}\star H)(u) = (K\star H)(u), \qquad \textbf{(25)}\end{aligned}$$

where $K(u) = (1-\varrho)\sum_{i=0}^{\infty}\varrho^i H^{\star i}(u)$ and $\varrho$, as earlier, defined as $(c_1-c_2)/(c_1-\mu)$. Now, following Embrechts, Goldie, and Veraverbeke [17],

$$\bar{K}(u) = \frac{\varrho}{1-\varrho}\bar{H}(u)(1+o(1))$$

as $u \to \infty$; the overbars indicate *complementary* distribution functions. Thus, using the convolution closure property of the class of subexponential random variables (see, e.g., Embrechts and Goldie [16] or Cline [12]), we conclude that

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

$$\begin{aligned}\mathbb{P}(Q_2 > u) &= \overline{K \star H}(u) \\ &= \left(\frac{\varrho}{1-\varrho} + 1\right)\bar{H}(u)(1+o(1)) \\ &= \frac{1}{1-\varrho}\,\bar{H}(u)(1+o(1)) \\ &= \frac{c_1-\mu}{c_2-\mu}\,\bar{H}(u)(1+o(1)).\end{aligned}$$

This completes the proof. ■

### 5.1. Tandem with $\alpha$-Stable Lévy Input

In this subsection we focus on the case of $\alpha$-stable Lévy input. Let $X_{\alpha,\sigma,\beta}(\cdot)$ denote an $\alpha$-stable Lévy motion with scale parameter $\sigma$ and skewness parameter $\beta$ (see, e.g., [34]). We analyze $\mathbb{P}(Q_2 > u)$ for $A(\cdot)$ an $\alpha$-stable Lévy motion with $\alpha \in (1,2)$, $\sigma = 1$, and $\beta \in (-1,1]$. For the general case of $\beta \in (-1,1]$, we present the generic asymptotic upper and lower bounds for $\mathbb{P}(Q_2 > u)$ in the large-buffer regime (i.e., $u \to \infty$). Moreover, if $\beta = 1$, then the application of Proposition 5.1 provides us with the exact asymptotics of $\mathbb{P}(Q_2 > u)$.

In the rest of this section we use the notation

$$B_1(\alpha,\beta) := \frac{\Gamma(1+\alpha)}{\pi}\sqrt{1+\beta^2\tan^2(\pi\alpha/2)}\,\sin\left(\frac{\pi\alpha}{2} + \arctan\left(\beta\tan\left(\frac{\pi\alpha}{2}\right)\right)\right),$$

$$B_2(\alpha,\beta) := \frac{1+\beta}{2}\,\frac{\alpha-1}{\Gamma(2-\alpha)\cos(\pi(\alpha-2)/2)}.$$

THEOREM 5.3: *Let $A(t) = X_{\alpha,1,\beta}(t)$ with $\alpha \in (1,2)$ and $\beta \in (-1,1]$. Then the following hold:*

*(i) As $u \to \infty$,*

$$\mathbb{P}(Q_2 > u) \le \left(\frac{B_1(\alpha,\beta)}{c_2\alpha(\alpha-1)} + \frac{B_2(\alpha,\beta)}{c_1}\right)\left(\frac{c_1}{c_1-c_2}\right)^{1-\alpha} u^{1-\alpha}(1+o(1)).$$

*(ii) As $u \to \infty$,*

$$\mathbb{P}(Q_2 > u) \ge \frac{B_1(\alpha,\beta)}{c_2\alpha(\alpha-1)}\left(\frac{c_1}{c_1-c_2}\right)^{1-\alpha} u^{1-\alpha}(1+o(1)).$$

The following lemma from [29] (see also [20, Prop. 3.7], where the substitution $\beta := -\beta$ matches our notation) is needed in our analysis.

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

LEMMA 5.4: *Let* $\alpha \in (1,2)$, $\beta \in (-1,1]$, *and* $c > 0$. *Then*

$$\mathbb{P}\Big(\sup_{t\in[0,\infty)}\{X_{\alpha,1,\beta}(t) - ct\} > u\Big) = \frac{B_1(\alpha,\beta)}{c\alpha(\alpha-1)}\, u^{1-\alpha}(1+o(1))$$

*as* $u \to \infty$.

PROOF OF THEOREM 5.3: Let $X_{\alpha,1,\beta}(\cdot)$ and $Y_{\alpha,1,\beta}(\cdot)$ be mutually independent $\alpha$-stable Lévy motions and recall that, as earlier, $t_u = u/(c_1 - c_2)$.

(i): Following Theorem 2.4, we have that

$$\mathbb{P}(Q_2 > u) \le \mathbb{P}\Big(\sup_{t\in[0,\infty)}\{X_{\alpha,1,\beta}(t) - c_2 t\} + \{Y_{\alpha,1,\beta}(t_u) > c_1 t_u\}\Big).$$

Moreover, using self-similarity of $Y_{\alpha,1,\beta}(\cdot)$ and following [34, Property 1.2.15],

$$\begin{aligned}\mathbb{P}(Y_{\alpha,1,\beta}(t_u) > c_1 t_u) &= \mathbb{P}(Y_{\alpha,1,\beta}(1) > c_1 t_u^{(\alpha-1)/\alpha}) \\ &= B_2(\alpha,\beta)c_1^{-\alpha} t_u^{1-\alpha}(1+o(1)) \qquad \textbf{(26)}\end{aligned}$$

as $u \to \infty$.

Combining (26) with Lemma 5.4 and using that both asymptotics are regularly varying at infinity, we immediately obtain

$$\begin{aligned}\mathbb{P}(Q_2 > u) &\le \left(\frac{B_1(\alpha,\beta)}{c_2\alpha(\alpha-1)}\, c_1^{1-\alpha} + B_2(\alpha,\beta)c_1^{-\alpha}\right) t_u^{1-\alpha}(1+o(1)) \\ &= \left(\frac{B_1(\alpha,\beta)}{c_2\alpha(\alpha-1)} + \frac{B_2(\alpha,\beta)}{c_1}\right) \\ &\quad \times \left(\frac{c_1}{c_1-c_2}\right)^{1-\alpha} u^{1-\alpha}(1+o(1))\end{aligned}$$

as $u \to \infty$, which completes the proof of (i).

(ii): Following Theorem 2.4, we have that

$$\begin{aligned}&\mathbb{P}(Q_2 > u) \\ &\quad = \mathbb{P}\Big(\sup_{t\in[0,\infty)}\{X_{\alpha,1,\beta}(t) - c_2 t\} > \sup_{t\in[0,t_u]}\{-Y_{\alpha,1,\beta}(t) + c_1 t\}\Big).\end{aligned}$$

Hence,

$$\begin{aligned}&\mathbb{P}(Q_2 > u) \\ &\quad \ge \mathbb{P}\Big(\sup_{t\in[0,\infty)}\{X_{\alpha,1,\beta}(t) - c_2 t\} - \sup_{t\in[0,t_u]}\{-Y_{\alpha,1,\beta}(t)\} > c_1 t_u\Big) \\ &\quad = \mathbb{P}\Big(\sup_{t\in[0,\infty)}\{X_{\alpha,1,\beta}(t) - c_2 t\} > c_1 t_u\Big)(1+o(1)) \qquad \textbf{(27)} \\ &\quad = \frac{B_1(\alpha,\beta)}{c_2\alpha(\alpha-1)}\left(\frac{c_1}{c_1-c_2}\right)^{1-\alpha} u^{1-\alpha}(1+o(1)) \qquad \textbf{(28)}\end{aligned}$$

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

as $u \to \infty$, where (27) follows from [28, Lemma 6.3] combined with the fact that, due to Lemma 5.4, $\sup_{t\in[0,\infty)}\{X_{\alpha,1,\beta}(t) - c_2 t\}$ belongs to the class of subexponential random variables. Equation (28) straightforwardly follows from Lemma 5.4. This completes the proof. ■

In the following theorem we analyze an important special case of the $\alpha$-stable Lévy input process where $\beta = 1$. The importance of this class of input processes stems, for example, from [27,41], where it was proved that $X_{\alpha,1,1}(\cdot)$ $(\alpha \in (1,2))$ can be used as an approximation for an appropriately scaled superposition of many on–off input processes.

THEOREM 5.5: *Let* $A(\cdot) = X_{\alpha,1,1}(\cdot)$ *with* $\alpha \in (1,2)$. *Then, as* $u \to \infty$,

$$\mathbb{P}(Q_2 > u) = \frac{1}{\Gamma(2-\alpha)\cos(\pi(\alpha-2)/2)} \frac{1}{c_2}\left(\frac{c_1}{c_1 - c_2}\right)^{1-\alpha} u^{1-\alpha}(1 + o(1)).$$

PROOF: The combination of Proposition 5.1 with Lemma 5.4 gives

$$\mathbb{P}(Q_2 > u) = \frac{B_1(\alpha,\beta)}{c_2\alpha(\alpha-1)}\left(\frac{c_1}{c_1 - c_2}\right)^{1-\alpha} u^{1-\alpha}(1 + o(1)).$$

Routine calculation, as in [20, Chap. 3], yields that

$$B_1(\alpha,1) = \frac{\alpha(\alpha-1)}{\Gamma(2-\alpha)\cos(\pi(\alpha-2)/2)},$$

which completes the proof. ■

*Remark 5.6:* Observe that the exact asymptotic presented in Theorem 5.5 matches the lower bound given in Theorem 5.3. We conjecture that this observation extends to the whole range of parameters $\beta \in (-1,1]$.

### 5.2. Tandem with Compound Poisson Input, with Regularly Varying Jumps

Consider again the case of compound Poisson$(\lambda, F)$ input. In the following theorem, we present the exact asymptotics of the tail distribution of $Q_2$ under the assumption that $\bar{F}(\cdot)$ is regularly varying at $\infty$.

THEOREM 5.7: *Let* $A(\cdot)$ *correspond to a compound Poisson process* $(\lambda, F)$. *Suppose that*

$$1 - F(x) = x^{-\alpha}\ell(x),$$

*where* $\alpha \in (1,2)$, $\ell(\cdot)$ *is slowly varying at* $\infty$, *and* $c_2 > \lambda\mathbb{E}X$. *Then*

$$\mathbb{P}(Q_2 > u) = \frac{\lambda}{c_2 - \lambda\mathbb{E}X}\left(\frac{c_1 - \lambda\mathbb{E}X}{c_1 - c_2}\right)^{1-\alpha} \frac{1}{\alpha - 1} u^{1-\alpha}\ell(u)(1 + o(1))$$

*as* $u \to \infty$.

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

PROOF: Observe that $\mathbb{E}A(t) = \lambda \mathbb{E}Xt$. Following [3, Thm. 2.1], we have that

$$\mathbb{P}(Q > u) = \frac{\lambda \mathbb{E}X}{c_2 - \lambda \mathbb{E}X} \frac{1}{\mathbb{E}X} \left( \int_u^\infty (1 - F(x))\, dx \right) (1 + o(1))$$

as $u \to \infty$. Thus, using Karamata's theorem, we have

$$\mathbb{P}(Q > u) = \frac{\lambda}{c_2 - \lambda \mathbb{E}X} \frac{1}{\alpha - 1} u^{1-\alpha} \ell(u)(1 + o(1))$$

as $u \to \infty$. Hence, the theorem follows by application of Proposition 5.1. ■

*Acknowledgments*

The first and second authors would like to thank the Mittag-Leffler Institute, Djursholm, Sweden for hospitality. Part of the work was done while the first and third authors were also at CWI. The first author was partially supported by University of Wrocław under grant 2529/W/IM/2004 and by KBN grant 1 P03A 031 28 (2005–2007). Pascal Lieshout (CWI) is thanked for providing useful remarks concerning Corollary 4.4.

*References*

1. Abate, J. & Whitt, W. (1997). Asymptotics for M/G/1 low-priority waiting-time tail probabilities. *Queueing Systems* 25: 173–233.
2. Adan, I. & Resing, J. (1996). A note on a fluid queue driven by an M/M/1 queue. *Queueing Systems* 22: 171–174.
3. Asmussen, S. (2000). *Ruin probabilities*. London: World Scientific.
4. Asmussen, S. (2003). *Applied probability and queues*. New York: Springer-Verlag.
5. Avi-Itzhak, B. (1965). A sequence of service stations with arbitrary input and regular service times. *Management Science* 11: 565–571.
6. Baxter, G. & Donsker, M.D. (1957). On the distribution of the supremum functional for the processes with stationary independent increments. *Transactions of the American Mathematical Society* 85: 73–87.
7. Bertoin, J. (1996). *Lévy processes*. Cambridge: Cambridge University Press.
8. Bingham, N. (1975). Fluctuation theory in continuous time. *Advances in Applied Probability* 7: 705–766.
9. Boxma, O. & Dumas, V. (1998). The busy period in the fluid queue. *Performance Evaluation Review* 26: 100–110.
10. Cao, J. & Ramanan, K. (2002). A Poisson limit for buffer overflow probabilities. In *Proceedings IEEE INFOCOM*, New York, pp. 994–1003.
11. Chang, C.-S., Heidelberger, P., Juneja, S. & Shahabuddin, P. (1994). Effective bandwidth and fast simulation of ATM intree networks. *Performance Evaluation* 20: 45–65.
12. Cline, D.B.H. (1986). Convolution tails, product tails and domains of attraction. *Probability Theory and Related Fields* 72: 529–557.
13. Cox, D. & Smith, W. (1961). *Queues*. London: Methuen.
14. Dębicki, K. & Mandjes, M. (2004). Traffic with an fBm limit: Convergence of the stationary workload process. *Queueing Systems* 46: 113–127.
15. Dębicki, K. & Palmowski, Z. (1999). On–off fluid models in heavy traffic environment. *Queueing Systems* 33: 327–338.
16. Embrechts, P. & Goldie, C.M. (1982). On convolution tails. *Stochastic Processes and Their Applications* 13: 263–268.

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press

17. Embrechts, P., Goldie, C.M., & Veraverbeke, N. (1979). Subexponentiality and infinite divisibility. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwunde Gebiete* 49: 335–347.
18. Friedman, H. (1965). Reduction methods for tandem queueing systems. *Operations Research* 13: 121–131.
19. Fristedt, B. (1974). Sample functions of stochastic processes with stationary independent increments. *Advances in Probability* 3: 241–396.
20. Furrer, H. (1997). Risk theory and heavy-tailed Lévy processes. Ph.D. dissertation, ETH, Zurich. Available at http://www.math.ethz.ch/~hjfurrer/.
21. Humblet, P., Bhargava, A., & Hluchyj, M. (1993). Ballot theorems applied to the transient analysis of nD/D/1 queues. *IEEE/ACM Transactions on Networking* 1: 81–95.
22. Kaj, I. & Taqqu, M. (2004). Convergence to fractional Brownian motion and to the Telecom process: The integral representation approach. Available at http://www.math.uu.se/~ikaj/preprints/kajtaqqu040712.pdf.
23. Kella, O. (1993). Parallel and tandem fluid networks with dependent Lévy inputs. *Annals of Applied Probability* 3: 682–695.
24. Kella, O. (2001). Markov-modulated feedforward fluid networks. *Queueing Systems* 37: 141–161.
25. Kella, O. & Whitt, W. (1992). A tandem fluid network with Lévy input. In U. Bhat & I. Basawa (eds.), *Queueing and related models*. Oxford: Clarendon Press.
26. Mandjes, M. & van Uitert, M. (2005). Sample-path large deviations for tandem and priority queues with Gaussian inputs. *Annals of Applied Probability* 15: 1193–1226.
27. Mikosch, T., Resnick, S., Rootzén, H. & Stegeman, A. (2002). Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Annals of Applied Probability* 12: 23–68.
28. Pitman, E.J.G. (1980). Subexponential distribution functions. *Journal of the Australian Mathematical Society A* 29: 337–347.
29. Port, S. (1989). Stable processes with drift on the line. *Transactions of the American Mathematical Society* 313: 805–841.
30. Reich, E. (1958). On the integrodifferential equation of Takács I. *Annals of Mathematical Statistics* 29: 563–570.
31. Roberts, J., Mocci, U., & Virtamo, J. (1996). *Broadband network teletraffic. Final report of action COST 242*. Berlin: Springer-Verlag.
32. Roberts, J. & Virtamo, J. (1991). The superposition of periodic cell arrival streams in an ATM multiplexer. *IEEE Transactions on Communications* 39: 298–303.
33. Rubin, I. (1974). Path delays in communication networks. *Applied Mathematics and Optimization* 1: 193–221.
34. Samorodnitsky, G. & Taqqu, M. (1994). *Stable non-Gaussian random processes*. London: Chapman & Hall.
35. Sato, K. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge: Cambridge University Press.
36. Scheinhardt, W. & Zwart, B. (2002). A tandem fluid queue with gradual input. *Probability in the Engineering and Informational Sciences* 16: 29–45.
37. Shalmon, M. & Kaplan, M. (1984). A tandem network of queues with deterministic service and intermediate arrivals. *Operations Research* 32: 753–773.
38. Takács, L. (1965). On the distribution of the supremum of stochastic processes with exchangeable increments. *Transactions of the American Mathematical Society* 119: 367–379.
39. Takács, L. (1967). *Combinatorial methods in the theory of stochastic processes*. New York: Wiley.
40. Takagi, H. (1991). *Queueing analysis. A Foundation of Performance Evaluation*. Vol. I: *Vacation and Priority Systems, Part 1*. Amsterdam: North-Holland.
41. Taqqu, M., Willinger, W., & Sherman, R. (1997). Proof of a fundamental result in self-similar traffic modeling. *Computer Communications Review* 27: 5–23.
42. Zolotarev, V.M. (1964). The first passage time of a level and the behaviour at infinity for a class of processes with independent increments. *Theory of Probability and Its Applications* 9: 653–661.

https://doi.org/10.1017/S0269964807070064 Published online by Cambridge University Press