



## UvA-DARE (Digital Academic Repository)

### Tail asymptotics of the M/G/ $\infty$ model

Mandjes, M.; Żuraniewski, P.

**DOI**

[10.1080/15326349.2011.542730](https://doi.org/10.1080/15326349.2011.542730)

**Publication date**

2011

**Document Version**

Final published version

**Published in**

Stochastic Models

[Link to publication](#)

**Citation for published version (APA):**

Mandjes, M., & Żuraniewski, P. (2011). Tail asymptotics of the M/G/ $\infty$  model. *Stochastic Models*, 27(1), 77-93. <https://doi.org/10.1080/15326349.2011.542730>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

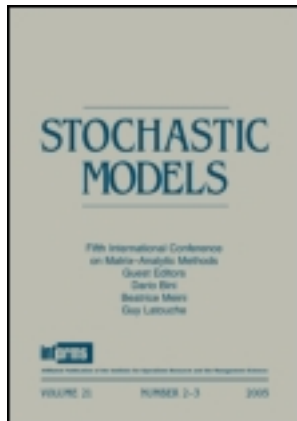
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

This article was downloaded by: [UVA Universiteitsbibliotheek SZ]

On: 17 June 2012, At: 23:49

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Stochastic Models

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/Istm20>

### Tail Asymptotics of the M/G/ $\infty$ Model

M. Mandjes<sup>a</sup> & P. Żuraniewski<sup>a b c</sup>

<sup>a</sup> Korteweg-de-Vries Institute for Mathematics, Amsterdam, The Netherlands

<sup>b</sup> Department of Applied Mathematics, AGH University of Science and Technology, Krakow, Poland

<sup>c</sup> TNO, Delft, The Netherlands

Available online: 24 Jan 2011

To cite this article: M. Mandjes & P. Żuraniewski (2011): Tail Asymptotics of the M/G/ $\infty$  Model, Stochastic Models, 27:1, 77-93

To link to this article: <http://dx.doi.org/10.1080/15326349.2011.542730>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## TAIL ASYMPTOTICS OF THE M/G/∞ MODEL

M. Mandjes<sup>1</sup> and P. Żuraniewski<sup>1,2,3</sup>

<sup>1</sup>Korteweg-de-Vries Institute for Mathematics, Amsterdam, The Netherlands

<sup>2</sup>Department of Applied Mathematics, AGH University of Science and Technology, Krakow, Poland

<sup>3</sup>TNO, Delft, The Netherlands

□ This paper considers the so-called M/G/∞ model: jobs arrive according to a Poisson process with rate  $\lambda$ , and each of them stays in the system during a random amount of time, distributed as a non-negative random variable  $B$ ; throughout it is assumed that  $B$  is light-tailed. With  $N(t)$  denoting the number of jobs in the system, the random process  $A(t)$  records the load imposed on the system in  $[0, t]$ , i.e.,  $A(t) := \int_0^t N(s)ds$ . The main result concerns the tail asymptotics of  $A(t)/t$ : we find an explicit function  $f(\cdot)$  such that

$$f(t) \sim \mathbb{P}\left(\frac{A(t)}{t} > \varrho(1 + \varepsilon)\right),$$

for  $t$  large; here  $\varrho := \lambda EB$ . A crucial issue is that  $A(t)$  does not have i.i.d. increments, which makes direct application of the classical Bahadur–Rao result impossible; instead an adaptation of this result is required. We compare the asymptotics found with the (known) asymptotics for  $\varrho \rightarrow \infty$  (and  $t$  fixed).

**Keywords** Infinite-servers queues; Large deviations; Tail asymptotics.

**Mathematics Subject Classification** 60K25.

### 1. INTRODUCTION

Consider the so-called M/G/∞ model: jobs arrive according to a Poisson process with rate  $\lambda$ , and their stay in the system constitutes an i.i.d. sequence of random variables  $(B_i)_{i \in \mathbb{N}}$ , distributed as some generic non-negative random variable  $B$ . With  $N(t)$  denoting the number of jobs in the system, the random process  $A(t)$  records the load imposed on the system in  $[0, t]$ , i.e.,  $A(t) := \int_0^t N(s)ds$ .

Received January 2010; Accepted August 2010

Address correspondence to Michel Mandjes, Korteweg-de-Vries Institute for Mathematics, Plantage Muidergracht 24, Amsterdam, The Netherlands; E-mail: mmandjes@science.uva.nl

It is known that, in stationarity, the number of jobs in the system obeys a Poisson distribution with mean  $\varrho := \lambda \mathbb{E}B$ ; importantly, this distribution is insensitive in the higher moments of the job duration  $B$ . The distribution of  $A(t)$ , however, can be expressed only implicitly. Recognizing that  $A(t)$  can be decomposed in terms of the sum of the contributions of (i) the jobs that were already present at time 0 (the number of which being Poisson distributed with mean  $\varrho$ ) and (ii) the jobs arriving in  $(0, t)$  (the number of which being Poisson distributed with mean  $\lambda t$ ), the Laplace transform of  $A(t)$  can be given explicitly. This enables the computation of all moments of  $A(t)$ . The computation of the distribution of  $A(t)$  is less straightforward though.

### Contribution

It is known that  $\mathbb{E}A(t) = \varrho t$ ; in the sequel we sometimes write  $A_\varrho(t)$  rather than  $A(t)$  if we wish to emphasize the dependence on the system load  $\varrho$ . We thus have that  $A(t)/t$  converges to  $\varrho$ , and it is not hard to find the corresponding central limit theorem. It is less understood, however, how to find the exact asymptotics of the tail distribution of  $A(t)/t$ , i.e., how to find an explicit function  $f(\cdot)$  such that

$$f(t) \sim p_\varrho(t) := \mathbb{P}(A_\varrho(t) > \varrho t(1 + \varepsilon)),$$

for  $t$  large and  $\varepsilon > 0$ ; identifying the tail asymptotics of  $A(t)/t$  is the main contribution of this paper. As we are dealing with a rare-event probability here, the suitable theory to be used is *large-deviations theory*. The complication, however, lies in the fact that  $A(t)$  does *not* have i.i.d. increments (i.e., it is not a Lévy process). If that would have been the case, then the classical Bahadur–Rao asymptotics would have been directly applicable (see Theorem 3.7.4 in Ref.<sup>[21]</sup>). We will argue, though, that it is possible to write  $A(t)$  as the sum of i.i.d. random variables, but these random variables *depend on  $t$* . The main contribution of our work is to show that it is possible to adapt the proof of the Bahadur–Rao result to this setting, thus identifying the exact asymptotics  $f(\cdot)$ . The proof relies on a change-of-measure argument, in conjunction with delicate Berry–Esseen-type estimates.

To the best of our knowledge, our result is among the first results on *exact* asymptotics for sample means of  $t$  i.i.d. random variables, each of them (mildly) depending on  $t$ . Here we note that there is already a body of powerful results on rough, *logarithmic* asymptotics that apply here, most notably the celebrated Gärtner–Ellis result;<sup>[6,9,13]</sup> these can be used to find the limiting value of  $t^{-1} \cdot \log p_\varrho(t)$ . We remark that throughout this paper is assumed that the job durations are ‘light tailed’, i.e., their moment

generating function is finite in some open neighborhood of 0. In our paper, we also briefly discuss what happens in heavy-tailed cases.

Renormalizing time such that  $\mathbb{E}B = 1$ , we have that  $\varrho = \lambda$ . An alternative asymptotic regime is that of  $\varrho$  tending to  $\infty$ , keeping  $t$  fixed: how does  $p_\varrho(t)$  behave for  $\varrho$  large? In this case, it is possible to write  $A_\varrho(t)$  as the sum of i.i.d. increments (that do not depend on  $\varrho$ ), and as a result the Bahadur–Rao result applies here. We determine what these asymptotics look like in this regime. Through numerical experiments we assess the differences between the asymptotic regimes, for a set of representative job duration distributions.

## Literature

Our approach heavily relies on the seminal work by Bahadur and Rao,<sup>[2]</sup> and in particular the proof of this result as presented in Ref.<sup>[5]</sup>; a substantially more general result is due to Höglund.<sup>[10]</sup> There are close connections to the vast body of literature that is devoted to asymptotic expansions of Poisson sums of i.i.d. random variables; in this respect we mention the results in Refs.<sup>[4,7]</sup>

## Organization

This paper is organized as follows. In Section 2 the model is introduced in detail. Also, a set of preliminaries is given (with emphasis on large deviations results). Section 3 contains the main results, most notably the exact asymptotics of  $p_\varrho(t)$  for  $t$  large. Section 4 indicates what happens when  $\varrho$  is large. Numerics are provided that give insight into the accuracy of the resulting approximations. Section 4 also contains a brief discussion of the case that the service times are heavy-tailed.

## 2. MODEL, PRELIMINARIES AND GOAL

### Model

We consider a model in which jobs arrive according to a Poisson process with rate  $\lambda > 0$ . While in the system, they generate traffic at a (normalized) rate 1, for a duration that is distributed as a generic random variable  $B$ . The durations are independent of each other, and independent of the arrival process. It is assumed that  $B$  has a finite moment generating function in some open neighborhood of 0, so that *a fortiori* also the mean  $\mathbb{E}B$  is finite. Denote the instantaneous load imposed on the system, i.e.,  $\lambda\mathbb{E}B$ , by  $\varrho$ . This model is commonly referred to as the M/G/∞ input model. Let  $N(t)$  be the number of jobs in the system at time  $t$ ; it is a classical result that its equilibrium distribution is Poisson with mean  $\varrho$

(see p. 181 Ref.<sup>[14]</sup>). Denote the load imposed on the system in the interval  $[0, t]$  by

$$A(t) := \int_0^t N(s) ds.$$

In this paper we study  $A(t)$ , where it is assumed that the M/G/ $\infty$  system is in equilibrium at time 0.

### Computation of the Moment Generating Function

In principle,  $A(t)$  is uniquely characterized by its moment generating function (mgf). We now show how to compute this mgf. We decompose  $A(t)$  into  $A_-(t)$  and  $A_+(t)$ . Here  $A_-(t)$  is the contribution due to the jobs arriving in  $(0, t]$ , and  $A_+(t)$  due to the jobs that were already present at time 0. Due to the stationarity of the arrival process,  $A_-(t)$  and  $A_+(t)$  are independent. The mgfs of both  $A_-(t)$  and  $A_+(t)$  can be computed. Regarding  $A_+(t)$ , recall that the number of jobs present at time 0 has a Poisson distribution with mean  $\varrho$ . The distribution of its residual life time  $B^r$  is given by Theorem 3.10 in Ref.<sup>[11]</sup>

$$\mathbb{P}(B^r > t) = \frac{1}{\mathbb{E}B} \int_t^\infty \mathbb{P}(B > s) ds.$$

Distinguishing between the situation that the job leaves before time  $t$  and is still present at time  $t$ , we readily obtain

$$\begin{aligned} \log \mathbb{E} e^{\vartheta A_+(t)} &= \log \sum_{k=0}^{\infty} e^{-\varrho} \frac{\varrho^k}{k!} \left[ \int_0^t e^{\vartheta s} f_{B^r}(s) ds + e^{\vartheta t} \mathbb{P}(B^r > t) \right]^k \\ &= \varrho V_t(\vartheta), \quad \text{where } V_t(\vartheta) := \int_0^t e^{\vartheta s} f_{B^r}(s) ds + e^{\vartheta t} \mathbb{P}(B^r > t) - 1. \end{aligned}$$

This means that if  $e^{\vartheta t} \mathbb{P}(B^r > t) \rightarrow 0$  as  $t \rightarrow \infty$ , then

$$\log \mathbb{E} e^{\vartheta A_+(t)} \rightarrow \varrho (\mathbb{E} e^{\vartheta B^r} - 1) = \varrho \left( \frac{M_B(\vartheta) - 1}{\vartheta \cdot \mathbb{E}B} - 1 \right), \quad (1)$$

where  $M_B(\vartheta) := \mathbb{E} e^{\vartheta B}$ .

We now focus on  $A_-(t)$ . First recall that the number of arriving jobs has a Poisson distribution with mean  $\lambda t$ . Observe that any job that has entered the system in the time interval  $(0, t]$ , it has done so at a time epoch that is uniformly distributed on this interval (and, in addition, the arrival epochs of the individual jobs, conditional on the number of jobs,

are independent). Again distinguishing between the jobs that have left by time  $t$  and the jobs that are still present, we obtain

$$\begin{aligned} & \log \mathbb{E} e^{\vartheta A_-(t)} \\ &= \log \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \left[ \int_0^t \frac{1}{t} \int_0^{t-s} f_B(r) e^{\vartheta r} dr ds + \int_0^t \frac{1}{t} \int_{t-s}^{\infty} f_B(r) e^{\vartheta(t-s)} dr ds \right]^k. \end{aligned}$$

Routine calculations yield:

$$\log \mathbb{E} e^{\vartheta A_-(t)} = \lambda W_t(\vartheta), \quad (2)$$

where

$$W_t(\vartheta) := -t + \int_0^t (t-s) f_B(s) e^{\vartheta s} ds + \int_0^t e^{\vartheta s} \mathbb{P}(B > s) ds.$$

### Instantaneous Input

We now consider a related model, namely the model in which the arriving jobs feed their work immediately into the system. Let  $\bar{A}(t)$  be the input process for the corresponding model with instantaneous input (i.e., compound Poisson). Then it requires a trivial computation to show that, for any  $t \geq 0$ ,

$$\log \mathbb{E} e^{\vartheta \bar{A}(t)} = t \Lambda(\vartheta), \quad \text{where } \Lambda(\vartheta) := \lambda M_B(\vartheta) - \lambda;$$

observe that  $\bar{A}(t)$  has (being a compound Poisson process) independent and identically distributed increments (unlike  $A(t)$ ).

### Law of Large Numbers, Central-Limit Theorem

It is readily verified that, as  $t \rightarrow \infty$ ,

$$\mathbb{E} \exp\left(\vartheta \frac{\bar{A}(t) - \varrho t}{\sqrt{t}}\right) \rightarrow \exp\left(\frac{\sigma^2}{2} \vartheta^2\right), \quad \text{with } \sigma^2 := \lambda \mathbb{E} B^2.$$

This implies, apart from  $\bar{A}(t)/t$  converging to  $\varrho$  almost surely, that

$$\frac{\bar{A}(t) - \varrho t}{\sigma \sqrt{t}} \rightarrow_d N(0, 1),$$

with  $N(0,1)$  being a standard-Normal random variable, and ‘ $\rightarrow_d$ ’ denoting convergence in distribution.

The next question is: does  $A(t)$  obey *the same* law of large numbers and central-limit theorem as its instantaneous-input counterpart  $\bar{A}(t)$ ? To this end, we determine

$$\lim_{t \rightarrow \infty} \mathbb{E} \exp\left(\vartheta \frac{A(t) - \varrho t}{\sqrt{t}}\right).$$

It is first observed from Eq. (1) that

$$\lim_{t \rightarrow \infty} \mathbb{E} \exp\left(\vartheta \frac{A_+(t)}{\sqrt{t}}\right) = 0.$$

Bearing in mind Eq. (2), and using

$$\lim_{t \rightarrow \infty} \lambda \int_0^t e^{\vartheta s/\sqrt{t}} \mathbb{P}(B > s) ds = \varrho,$$

and

$$\begin{aligned} & \lim_{t \rightarrow \infty} \left( -\lambda t + \lambda \int_0^t (t-s) f_B(s) e^{\vartheta s/\sqrt{t}} ds - \vartheta \varrho \sqrt{t} \right) \\ &= \lim_{t \rightarrow \infty} \left( -\lambda t + \lambda \int_0^t (t-s) f_B(s) \left( 1 + \frac{\vartheta s}{\sqrt{t}} + \frac{\vartheta^2 s^2}{2t} \right) ds - \vartheta \varrho \sqrt{t} \right) \\ &= -\varrho + \frac{\sigma^2}{2} \vartheta^2, \end{aligned}$$

we conclude that  $A(t)$  has the same law of large numbers and central-limit theorem as  $\bar{A}(t)$ .

**Remark 1.** Later we will use that  $A_-(t)$  also obeys a central-limit theorem. It is a matter of straightforward calculus to verify that  $\varrho_-(t) := \mathbb{E}A_-(t)$  satisfies

$$\lim_{t \rightarrow \infty} \varrho_-(t) - \varrho t = -\frac{\sigma^2}{2},$$

whereas  $v_-(t) := \text{Var} A_-(t)$  is such that

$$\lim_{t \rightarrow \infty} v_-(t) - \sigma^2 t = -\frac{2}{3} \lambda \mathbb{E}B^3.$$

An argument similar to the one used above then yields

$$\lim_{t \rightarrow \infty} \mathbb{E} \exp\left(\vartheta \frac{A_-(t) - \varrho_- t}{\sqrt{v_-(t)}}\right) = \exp\left(\frac{\vartheta^2}{2}\right),$$



resulting in asymptotic Normality. For later reference, we include here the corresponding Berry–Esseen-type estimate. Define by  $\Phi(\cdot)$  the distribution function of a standard normal random variable, and by  $\phi(\cdot)$  the corresponding density, and let  $\zeta(x) := (1 - x^2)\phi(x)$ . Then for some finite, positive  $m$ ,

$$\sqrt{t} \left( \sup_x \left( \mathbb{P} \left( \frac{A_-(t) - \varrho_- t}{\sqrt{v_-(t)}} \leq x \right) - \Phi(x) - \frac{m}{6\sqrt{t}} \zeta(x) \right) \right) \rightarrow 0, \quad (3)$$

as  $t \rightarrow \infty$  (see Section III.11 in Ref.<sup>[15]</sup>; cf. Eq. XVI.4.1 in Ref.<sup>[8]</sup>).

### Large Deviations

Now that we have seen that  $\bar{A}(t)$  and  $A(t)$  have the same law of large numbers and central-limit theorem, we may wonder whether they have the same large deviations. To this end, let us first compute the exponential decay rate of the probability that  $A(t)/t$  exceeds an extreme value, i.e., some value larger than  $\varrho$ . First observe that Eq. (2) implies that

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E} e^{\vartheta A_-(t)} = \lim_{t \rightarrow \infty} \left( -\lambda + \lambda \int_0^\infty \left( 1 - \frac{r}{t} \right) \mathbb{1}_{[0,t]}(r) f_B(r) e^{\vartheta r} dr \right),$$

which equals, by virtue of ‘monotone convergence’,

$$-\lambda + \lambda \int_0^\infty f_B(r) e^{\vartheta r} dr = \Lambda(\vartheta).$$

From Eq. (1), we conclude that the contribution of  $A_+(t)$  can be neglected in the sense that also

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E} e^{\vartheta A(t)} = \Lambda(\vartheta).$$

It now follows from the Gärtner–Ellis result (see Theorem 2.3.6 in Ref.<sup>[5]</sup>) that  $A(t)/t$  obeys the following logarithmic asymptotics:

**Lemma 2.1.** *The logarithmic asymptotics of  $A(t)/t$  are, for  $a > \varrho$ , given by*

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{P} \left( \frac{A(t)}{t} > a \right) = -I(a) := -\sup_{\vartheta} (\vartheta a - \Lambda(\vartheta)).$$

It is immediate from the Gärtner–Ellis theorem that  $\bar{A}(t)/t$  satisfies the same large deviations as  $A(t)/t$ , i.e., there is exponential decay with rate

function  $I(\cdot)$ . Unlike  $A(t)$ , however,  $\bar{A}(t)$  has independent and identically distributed increments; assuming for ease that  $t$  is integer, we can write

$$\bar{A}(t) = \sum_{i=1}^t X_i, \quad \text{with } X_i := \bar{A}(i-1, i),$$

with  $\bar{A}(\tau, \sigma)$  the traffic generated in this instantaneous-input system between time epochs  $\tau$  and  $\sigma$ . Realizing that their distribution coincides with that of a  $\text{Poisson}(\lambda)$  number of independent jobs (each of them distributed as the random variable  $B$ ), it follows that  $\log \mathbb{E} \exp(\vartheta X_i) = \Lambda(\vartheta)$ . Now the so-called Bahadur–Rao result (see Theorem 3.7.4 Ref.<sup>[2]</sup>) can be applied to find the exact asymptotics of  $\bar{A}(t)/t$ :

$$\lim_{t \rightarrow \infty} \sqrt{t} e^{U(a)} \cdot \mathbb{P} \left( \frac{\bar{A}(t)}{t} > a \right) = C_a,$$

where

$$C_a := \left( \eta \sqrt{2\pi \Lambda''(\eta)} \right)^{-1} = \left( \eta \sqrt{2\pi \lambda M_B''(\eta)} \right)^{-1},$$

where  $\eta \equiv \eta_a$  solves  $\Lambda'(\eta) = a$ . The goal of this note is to find the exact asymptotics of  $A(t)/t$ , thus answering the question whether  $A(t)/t$  has the same exact asymptotics as  $\bar{A}(t)/t$ . The next section will reveal that the respective exact asymptotics differ by a constant.

### 3. EXACT ASYMPTOTICS

We mentioned above that  $A(t)$  does not have i.i.d. increments, being the reason why we cannot directly use Bahadur–Rao to compute the exact asymptotics of  $A(t)/t$ . More precisely: the  $A(i-1, i)$  (with  $A(\tau, \sigma)$  the amount of traffic entering between  $\tau$  and  $\sigma$  in our gradual-input system) are identically distributed but not independent. Interestingly, however, we can still write  $A(t)$  as the sum of  $t$  i.i.d. increments (assuming for ease  $t$  to be integer). The price to be paid is that these increments depend on  $t$ . This is done as follows.

Let  $X_i$  be constructed as follows. First we draw  $N_i$  from a Poisson distribution with mean  $\lambda$ . Then each of these  $N_i$  corresponds to an arrival at a uniformly distributed epoch in  $(0, t]$ ; say the  $j$ -th arrives at time  $T_{ij}$ . Then  $X_i$  is the amount of traffic generated by these  $N_i$  jobs in  $(0, t]$ . This can be alternatively written as

$$X_i := \sum_{j=1}^{N_i} \min\{t - T_{ij}, B_{ij}\},$$

with the  $B_{ij}$  i.i.d., distributed as the random variable  $B$ . Then the  $X_i$  are i.i.d., and  $A_-(t)$  equals  $\sum_{i=1}^t X_i$ ; this is essentially due to the fact that the sum of  $t$  independent Poisson processes with rate  $\lambda$  is a Poisson process with rate  $t\lambda$ .

The fact that  $A(t)$  is now written as the sum of  $t$  i.i.d. random variables does, however, not mean that we can use the Bahadur–Rao result now, as in our setting the i.i.d. increments  $X_i$  depend on  $t$  (and this is why we write in the sequel  $X_i(t)$  rather than just  $X_i$ ). Our goal is now to enhance the Bahadur–Rao result so that it also covers our setting.

Our proof essentially mimics the line of reasoning of the Bahadur–Rao result in Dembo and Zeitouni (see Theorem 3.7.4 in Ref.<sup>[51]</sup>). We start, however, with a number of results on the mgf of  $X(t)$ , where  $X(t)$  is a generic random variable distributed as the  $X_i(t)$ 's. Relying on computations similar to those leading to Eq. (2), we get that

$$\Lambda_t(\vartheta) := \log \mathbb{E} e^{\vartheta X(t)} = -\lambda + \lambda \int_0^t \left(1 - \frac{s}{t}\right) f_B(s) e^{\vartheta s} ds + \lambda \cdot \frac{1}{t} \int_0^t e^{\vartheta s} \mathbb{P}(B > s) ds.$$

We observe that for any  $\vartheta \geq 0$ , we have that  $\Lambda_t(\vartheta) \uparrow \Lambda(\vartheta)$ . It is an easy verification that

$$\Lambda(\vartheta) - \Lambda_t(\vartheta) \sim \frac{\lambda}{t} \left( \int_0^\infty s f_B(s) e^{\vartheta s} ds - \int_0^\infty e^{\vartheta s} \mathbb{P}(B > s) ds \right) = \frac{\lambda}{t} \cdot \varphi(\vartheta); \quad (4)$$

here

$$\varphi(\vartheta) := M'_B(\vartheta) + \frac{1}{\vartheta} - \frac{1}{\vartheta} \cdot M_B(\vartheta) \geq 0,$$

where the last inequality (being valid for  $\vartheta > 0$ ) follows from  $M_B(\vartheta) - \vartheta M'_B(\vartheta) \leq M_B(0) = 1$ , due to the convexity of  $M_B(\vartheta)$ . Likewise,

$$\Lambda'(\vartheta) - \Lambda'_t(\vartheta) \sim \frac{\lambda}{t} \cdot \psi(\vartheta), \quad \text{with}$$

$$\psi(\vartheta) := M''_B(\vartheta) - \frac{1}{\vartheta^2} + \frac{1}{\vartheta^2} \cdot M_B(\vartheta) - \frac{1}{\vartheta} \cdot M'_B(\vartheta).$$

Now sample the  $X_i(t)$  under a new measure under which the density  $f(\cdot)$  of the  $X_i(t)$  is replaced by the density  $g(\cdot)$ , given through  $g(x) = f(x) \cdot \exp(\eta x - \Lambda_t(\eta))$ . The standard change-of-measure identity yields, in self-evident notation:

$$\begin{aligned} \mathbb{P}\left(\frac{A_-(t)}{t} \geq a\right) &= \mathbb{E}_g\left(e^{-\eta A_-(t) + t \Lambda_t(\eta)} \cdot \mathbf{1}_{\{A_-(t) \geq at\}}\right) \\ &= e^{t(\Lambda_t(\eta) - \eta a)} \cdot \mathbb{E}_g\left(e^{-\eta(A_-(t) - at)} \cdot \mathbf{1}_{\{A_-(t) \geq at\}}\right); \end{aligned}$$

here we use the notation  $a_t := \Lambda'_t(\eta)$ ; recall that  $a = \Lambda'(\eta)$ . Now note that Eq. (4) implies that

$$\lim_{t \rightarrow \infty} \frac{\exp(t(\Lambda_t(\eta) - \eta a_t))}{\exp(t(\Lambda(\eta) - \eta a))} = e^{-\lambda\varphi(\eta)} \cdot e^{\eta\lambda\psi(\eta)}.$$

In order to get into a central-limit scaling, we define

$$Y_i(t) := \frac{X_i(t) - a_t}{\sqrt{\Lambda'_t(\eta)}}, \quad W(t) := \frac{1}{\sqrt{t}} \cdot \sum_{i=1}^t Y_i(t).$$

Hence, with  $F_t(\cdot)$  the distribution function of  $W(t)$  under  $g$ ,

$$\begin{aligned} \mathbb{E}_g \left( e^{-\eta(A_-(t) - ta_t)} \cdot \mathbf{1}_{\{A_-(t) \geq at\}} \right) &= \mathbb{E}_g \left( e^{-\eta\sqrt{\Lambda'_t(\eta)}\sqrt{t}W(t)} \cdot \mathbf{1}_{\{W(t) \geq (a-a_t)\sqrt{t}/\sqrt{\Lambda'_t(\eta)}\}} \right) \\ &= \int_{(a-a_t)\sqrt{t}/\sqrt{\Lambda'_t(\eta)}}^{\infty} e^{-\eta\sqrt{\Lambda'_t(\eta)}\sqrt{t}x} dF_t(x). \end{aligned}$$

Now we can finish the proof as in Ref.<sup>[5]</sup> (Thm. 3.7.4), using Eq. (3) as the counterpart of Eq. (3.7.8) in Ref.<sup>[5]</sup>. As a result, we can essentially replace the distribution function  $F_t(\cdot)$  by that of a standard Normal random variable (i.e., one first replaces  $F_t(\cdot)$  by the standard Normal distribution function adjusted by some error term, and then one uses dominated convergence to prove that the error term has no impact, similarly to the proof of Theorem 3.7.4 in Ref.<sup>[5]</sup> It follows that, for  $t \rightarrow \infty$ , with  $N(\mu, \sigma^2)$  denoting, as usual a Normally distributed random variable with mean  $\mu$  and variance  $\sigma^2$ ,

$$\begin{aligned} &\int_{(a-a_t)\sqrt{t}/\sqrt{\Lambda'_t(\eta)}}^{\infty} e^{-\eta\sqrt{\Lambda'_t(\eta)}\sqrt{t}x} dF_t(x) \\ &\sim \int_{(a-a_t)\sqrt{t}/\sqrt{\Lambda'_t(\eta)}}^{\infty} e^{-\eta\sqrt{\Lambda'_t(\eta)}\sqrt{t}x} d\Phi(x) \\ &= \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}\eta^2\Lambda'_t(\eta)t} \int_{(a-a_t)\sqrt{t}/\sqrt{\Lambda'_t(\eta)}}^{\infty} e^{-\frac{1}{2}(x+\eta\sqrt{\Lambda'_t(\eta)}\sqrt{t})^2} dx \\ &= e^{\frac{1}{2}\eta^2\Lambda'_t(\eta)t} \mathbb{P} \left( N \left( -\eta\sqrt{\Lambda'_t(\eta)}\sqrt{t}, 1 \right) > \frac{(a-a_t)\sqrt{t}}{\sqrt{\Lambda'_t(\eta)}} \right) \\ &\sim \left( 1 - \Phi \left( \eta\sqrt{\Lambda'_t(\eta)}\sqrt{t} + \frac{\lambda\psi(\eta)}{\sqrt{\Lambda'_t(\eta)}\sqrt{t}} \right) \right) e^{\frac{1}{2}\eta^2\Lambda'_t(\eta)t} \sim \frac{e^{-\eta\lambda\psi(\eta)}}{\eta\sqrt{2\pi\Lambda''(\eta)}} \cdot \frac{1}{\sqrt{t}}; \end{aligned}$$

in the last step the standard estimate  $1 - \Phi(x) \sim (2\pi x^2)^{-1/2} e^{-x^2/2}$  (as  $x \rightarrow \infty$ ) was applied.<sup>[1]</sup> Combining the above results, we arrive at the following proposition.

**Proposition 3.1.** *The exact asymptotics of  $A_-(t)/t$  are, for  $a > \varrho$ , given by*

$$\lim_{t \rightarrow \infty} \sqrt{t} e^{u(a)} \cdot \mathbb{P}\left(\frac{A_-(t)}{t} > a\right) = C_a \cdot e^{-\lambda\varphi(\eta)}.$$

**Remark 2.** It is an easy exercise in calculus to show that we can rewrite

$$\lambda\varphi(\eta) = a - \frac{\lambda}{\eta}(M_B(\eta) - 1).$$

As remarked earlier,  $\lambda\varphi(\eta) \geq 0$ ; this makes sense, as the exceedance probabilities should be lower than in the instantaneous-input model (to see this, recall that  $A(t) \leq \bar{A}(t)$ ).

We conclude by considering the exact asymptotics of  $A(t)/t$ . These can be determined in essentially the same way. Now  $A(t)$  can be written as the sum of  $\check{X}_i$  up to  $\check{X}_t$ , where

$$\check{X}_i := \sum_{j=1}^{N_i} \min\{t - T_{ij}, B_{ij}\} + \sum_{j=1}^{M_i} \min\{t, B_{ij}^r\}, \quad (5)$$

with  $N_i$  and the  $B_{ij}$  as before, and  $M_i$  a Poisson random variable with mean  $\varrho/t$  and the  $B_{ij}^r$  i.i.d., distributed as the random variable  $B^r$ ; in addition, both sums in (5) are independent. Let  $\check{X}(t)$  denote the generic random variable distributed as the  $\check{X}_i(t)$ 's. In the proof, we again get that

$$(\Lambda(\vartheta) - \log \mathbb{E} e^{\vartheta \check{X}(t)}) \cdot t$$

goes to a constant, say  $\check{\phi}(\vartheta)$ , when  $t$  grows large; it is easily verified that

$$\check{\phi}(\vartheta) = \varphi(\vartheta) - \frac{1}{\lambda} \log \mathbb{E} e^{\vartheta A_+(\infty)} = \varphi(\vartheta) - \left( \frac{M_B(\vartheta) - 1}{\vartheta} - \mathbb{E} B \right).$$

**Theorem 3.2.** *The exact asymptotics of  $A(t)/t$  are, for  $a > \varrho$ , given by*

$$\lim_{t \rightarrow \infty} \sqrt{t} e^{u(a)} \cdot \mathbb{P}\left(\frac{A(t)}{t} > a\right) = C_a \cdot e^{-\lambda\varphi(\eta)} \cdot \mathbb{E} e^{\eta A_+(\infty)}.$$

A next question is whether the asymptotics of the instantaneous-input model  $\bar{A}(t)$  and those of the gradual-input model  $A(t)$  are 'ordered',

as we have that  $e^{-\lambda\varphi(\eta)} \cdot \mathbb{E}e^{\eta A_+(\infty)}$  is always smaller or larger than 1. This makes sense as in the gradual-input model traffic arrives more smoothly than in the instantaneous-input model. It is not a property that is *a priori* clear, though:  $A(t)$  equals  $\bar{A}(t)$ , increased by the contribution of the flows present at time 0, and decreased by the part of the flows that enters the system after time  $t$ . The next proposition formalizes the ordering.

**Proposition 3.3.** *As  $t \rightarrow \infty$ ,*

$$\mathbb{P}\left(\frac{A(t)}{t} > a\right) / \mathbb{P}\left(\frac{\bar{A}(t)}{t} > a\right) \rightarrow e^{-\lambda\varphi(\eta)} \cdot \mathbb{E}e^{\eta A_+(\infty)} < 1.$$

*Proof.* First check that the criterion  $e^{-\lambda\varphi(\eta)} \cdot \mathbb{E}e^{\eta A_+(\infty)} < 1$  can be rewritten as

$$\frac{M_B(\eta) - 1}{\eta} > \frac{M'_B(\eta) + M'_B(0)}{2} = \frac{M'_B(\eta) + \mathbb{E}B}{2},$$

which can in turn be written as

$$H_1(\eta) := 2(M_B(\eta) - 1) < \eta(M'_B(\eta) + \mathbb{E}B) =: H_2(\eta). \quad (6)$$

The rest of the proof is devoted to establishing this inequality for  $\eta > 0$ .

To this end, first observe that  $M'_B(\cdot)$  is convex. Hence  $M''_B(\eta) > (M'_B(\eta) - M'_B(0))/\eta$ , which is equivalent to

$$M'_B(\eta) - \eta M''_B(\eta) < \mathbb{E}B.$$

This implies (6), as  $H_1(0) = H_2(0) = 0$  and

$$H'_1(\eta) = 2M'_B(\eta) < M'_B(\eta) + \eta M''_B(\eta) + \mathbb{E}B = H'_2(\eta).$$

This proves the stated. □

## 4. DISCUSSION AND NUMERICS

In this section we consider a second asymptotic regime: we show that the exact asymptotics of  $p_\varrho(t)$  for  $\varrho$  follow immediately from the Bahadur–Rao result. We also provide a numerical study to assess the quality of the various approximations.

### Large-Load Estimates

Renormalize time such that  $\mathbb{E}B = 1$ , implying that  $\varrho = \lambda$ . We now derive the asymptotics of  $p_\varrho(t)$  for fixed  $t$  and  $\varrho$  large. Assume for ease for

the moment that  $\varrho \in \mathbb{N}$ . It is trivial to see that  $A_\varrho(t)$  can be equivalently written as  $\sum_{i=1}^{\varrho} Z_i(t)$ , where the  $Z_i(t)$  are i.i.d. with log mgf equal to

$$U_i(\vartheta) := V_i(\vartheta) + W_i(\vartheta);$$

recall that  $V_i(\vartheta)$  and  $W_i(\vartheta)$  were defined in Section 2. The Bahadur–Rao result then entails that, for  $s > t$ ,

$$\lim_{\varrho \rightarrow \infty} \sqrt{\varrho} e^{\varrho J(s)} \cdot \mathbb{P}\left(\frac{A_\varrho(t)}{\varrho} \geq s\right) = K_s, \quad (7)$$

where  $J(s) := \sup_{\vartheta} (\vartheta s - U_i(\vartheta))$ , and

$$K_s := \left(\zeta \sqrt{2\pi U_i''(\zeta)}\right)^{-1},$$

where  $\zeta \equiv \zeta_s$  solves  $U_i'(\vartheta) = s$ .

### Numerical Example: M/M/∞

In this example we consider the case of exponentially distributed jobs (with mean  $1/\mu$ ). The goal is to compare the long-timescale asymptotics of Theorem 3.2, the large-load asymptotics presented above, a Normal approximation, and simulation-based estimates.

- We first consider the long-timescale regime. It is readily verified that

$$I(a) := \sup_{\vartheta} \left( \vartheta a - \frac{\lambda \vartheta}{\mu - \vartheta} \right) = (\sqrt{\mu a} - \sqrt{\lambda})^2,$$

such that  $\eta = \mu - \sqrt{\lambda \mu / a}$ . Elementary calculus yields

$$\Lambda''(\eta) = \lambda M_B''(\eta) = 2 \frac{a^{3/2}}{\sqrt{\lambda \mu}},$$

and

$$C_a = \left(\frac{\lambda \mu}{a}\right)^{1/4} \frac{1}{2\sqrt{\pi}} \cdot \sqrt{\frac{1}{\mu}} \frac{1}{\sqrt{\mu a} - \sqrt{\lambda}}; \quad \varphi(\eta) = \frac{a}{\lambda} - \sqrt{\frac{a}{\lambda \mu}}.$$

Noting that

$$\log \mathbb{E} e^{\eta A_+(\infty)} = \frac{\sqrt{\lambda}}{\mu} (\sqrt{\mu a} - \sqrt{\lambda}),$$

we eventually obtain that, as  $t \rightarrow \infty$ ,

$$\begin{aligned} \mathbb{P}\left(\frac{A_\varrho(t)}{t} \geq a\right) &\sim \frac{C_a}{\sqrt{t}} e^{-(t+1/\mu)(\sqrt{\mu a} - \sqrt{\lambda})^2} \\ &\sim \mathbb{P}\left(\frac{\bar{A}_\varrho(t)}{t} \geq a\right) \cdot e^{-(1/\mu)(\sqrt{\mu a} - \sqrt{\lambda})^2}. \end{aligned} \quad (8)$$

- Let us now consider the large-load estimates. To this end, we first compute

$$\begin{aligned} V_i(\vartheta) &= \frac{\vartheta}{\mu - \vartheta} (1 - e^{-(\mu - \vartheta)t}), \\ W_i(\vartheta) &= -\frac{\vartheta}{(\mu - \vartheta)^2} (1 - e^{-(\mu - \vartheta)t}) + t \frac{\vartheta}{\mu - \vartheta}. \end{aligned}$$

Recalling that we took  $\mu = 1$ , it follows that

$$U_i(\vartheta) = t \frac{\vartheta}{1 - \vartheta} - \left(\frac{\vartheta}{1 - \vartheta}\right)^2 (1 - e^{-(1 - \vartheta)t}).$$

Straightforward algebra leads to expressions for  $U'_i(\vartheta)$  and  $U''_i(\vartheta)$ . Now Eq. (7) enables us to compute the large-load asymptotics.

- A third approximation that we consider here is the Normal approximation

$$\mathbb{P}(A_\varrho(t) \geq \varrho t(1 + \varepsilon)) \approx \Phi\left(-\frac{\varrho t \varepsilon}{\sqrt{\text{Var } A_\varrho(t)}}\right),$$

with  $\Phi(\cdot)$  being the standard Normal distribution function. From the above formulae we also obtain that  $\text{Var } A_\varrho(t) = \Lambda''_i(0) = 2\varrho(t - 1 + e^{-t})$ , cf.<sup>[3,12]</sup>.

## Tables

In Table 1 we present numerical results for  $\varepsilon = 0.01$ . In the simulation results, the experiment was terminated at the moment that a relative efficiency (defined as the ratio of the confidence interval's half length and the estimate) is below 10%; we chose a confidence of 95%. It is observed that for (relatively) large  $\varrho$  and  $t$ , and  $\varepsilon$  small, all three approximations are more or less equal. This can be understood as follows.



**TABLE 1** Numerical results,  $\varepsilon = 0.01$ 

		$\varrho = 5$	$\varrho = 50$	$\varrho = 100$	$\varrho = 150$
$t = 100$	Sim	$4.57 \cdot 10^{-1}$	$2.90 \cdot 10^{-1}$	$2.53 \cdot 10^{-1}$	$1.91 \cdot 10^{-1}$
	large $t$	$2.49 \cdot 10^0$	$7.04 \cdot 10^{-1}$	$4.39 \cdot 10^{-1}$	$3.16 \cdot 10^{-1}$
	large $\varrho$	$2.48 \cdot 10^0$	$7.00 \cdot 10^{-1}$	$4.37 \cdot 10^{-1}$	$3.14 \cdot 10^{-1}$
	Norm	$4.37 \cdot 10^{-1}$	$3.08 \cdot 10^{-1}$	$2.39 \cdot 10^{-1}$	$1.92 \cdot 10^{-1}$
$t = 1000$	Sim	$2.76 \cdot 10^{-1}$	$5.57 \cdot 10^{-2}$	$1.26 \cdot 10^{-2}$	$3.25 \cdot 10^{-3}$
	large $t$	$7.04 \cdot 10^{-1}$	$7.26 \cdot 10^{-2}$	$1.48 \cdot 10^{-2}$	$3.48 \cdot 10^{-3}$
	large $\varrho$	$7.04 \cdot 10^{-1}$	$7.26 \cdot 10^{-2}$	$1.48 \cdot 10^{-2}$	$3.48 \cdot 10^{-3}$
	Norm	$3.08 \cdot 10^{-1}$	$5.68 \cdot 10^{-2}$	$1.26 \cdot 10^{-2}$	$3.07 \cdot 10^{-3}$
$t = 2000$	Sim	$2.27 \cdot 10^{-1}$	$1.30 \cdot 10^{-2}$	$8.69 \cdot 10^{-4}$	$5.75 \cdot 10^{-5}$
	large $t$	$4.40 \cdot 10^{-1}$	$1.48 \cdot 10^{-2}$	$8.69 \cdot 10^{-4}$	$5.89 \cdot 10^{-5}$
	large $\varrho$	$4.40 \cdot 10^{-1}$	$1.48 \cdot 10^{-2}$	$8.69 \cdot 10^{-4}$	$5.89 \cdot 10^{-5}$
	Norm	$2.40 \cdot 10^{-1}$	$1.27 \cdot 10^{-2}$	$7.81 \cdot 10^{-4}$	$5.35 \cdot 10^{-5}$

Based on the standard approximation  $\Phi(x) \approx x^{-1}(2\pi)^{-1/2} \exp(-\frac{1}{2}x^2)$ , we have that the Normal approximation gives

$$\mathbb{P}(A_\varrho(t) \geq \varrho t(1 + \varepsilon)) \approx \frac{1}{\varepsilon \sqrt{\pi \varrho t}} e^{-\frac{1}{4}(\varrho t)\varepsilon^2}. \quad (9)$$

Now consider the long-timescale approximation, with  $a = (\lambda/\mu)(1 + \varepsilon)$ . It is elementary to verify that for small  $\varepsilon$  it holds that  $\sqrt{\lambda(1 + \varepsilon)} - \sqrt{\lambda}$  approximately equals  $\frac{1}{2}\sqrt{\lambda}\varepsilon$ . Inserting this into (8), we obtain expression (9) for large  $\varrho$  and  $t$  and small  $\varepsilon$ . Regarding the large-load approximation, first verify that  $\zeta$  can be approximated by  $1 - (1 + \varepsilon)^{-1/2}$ . With  $1 - (1 + \varepsilon)^{-1/2} \approx \varepsilon/2$  for  $\varepsilon$  small, Eq. (7) yields that in this regime the large-load

**TABLE 2** Numerical results,  $\varepsilon = 0.2$ 

		$\varrho = 1$	$\varrho = 5$	$\varrho = 8$
$t = 50$	Sim	$1.52 \cdot 10^{-1}$	$1.45 \cdot 10^{-2}$	$2.82 \cdot 10^{-3}$
	large $t$	$2.51 \cdot 10^{-1}$	$1.75 \cdot 10^{-2}$	$3.43 \cdot 10^{-3}$
	large $\varrho$	$2.48 \cdot 10^{-1}$	$1.73 \cdot 10^{-2}$	$3.39 \cdot 10^{-3}$
	Norm	$1.56 \cdot 10^{-1}$	$1.19 \cdot 10^{-2}$	$2.14 \cdot 10^{-3}$
$t = 100$	Sim	$8.85 \cdot 10^{-2}$	$1.14 \cdot 10^{-3}$	$6.23 \cdot 10^{-5}$
	large $t$	$1.13 \cdot 10^{-1}$	$1.27 \cdot 10^{-3}$	$6.35 \cdot 10^{-5}$
	large $\varrho$	$1.12 \cdot 10^{-1}$	$1.26 \cdot 10^{-3}$	$6.31 \cdot 10^{-5}$
	Norm	$7.76 \cdot 10^{-2}$	$7.41 \cdot 10^{-4}$	$2.91 \cdot 10^{-5}$
$t = 150$	Sim	$4.49 \cdot 10^{-2}$	$1.02 \cdot 10^{-4}$	$1.30 \cdot 10^{-6}$
	large $t$	$5.83 \cdot 10^{-2}$	$1.06 \cdot 10^{-4}$	$1.36 \cdot 10^{-6}$
	large $\varrho$	$5.81 \cdot 10^{-2}$	$1.06 \cdot 10^{-4}$	$1.35 \cdot 10^{-6}$
	Norm	$4.11 \cdot 10^{-2}$	$5.10 \cdot 10^{-5}$	$4.43 \cdot 10^{-7}$

**TABLE 3** Numerical results,  $\varepsilon = 0.5$ 

		$\varrho = 1$	$\varrho = 2$	$\varrho = 3$	$\varrho = 4$	$\varrho = 5$
$t = 10$	Sim	$1.20 \cdot 10^{-1}$	$6.07 \cdot 10^{-2}$	$3.15 \cdot 10^{-2}$	$1.53 \cdot 10^{-2}$	$7.42 \cdot 10^{-3}$
	large $t$	$2.06 \cdot 10^{-1}$	$8.35 \cdot 10^{-2}$	$3.91 \cdot 10^{-2}$	$1.94 \cdot 10^{-2}$	$9.97 \cdot 10^{-3}$
	large $\varrho$	$1.94 \cdot 10^{-1}$	$7.81 \cdot 10^{-2}$	$3.63 \cdot 10^{-2}$	$1.79 \cdot 10^{-2}$	$9.13 \cdot 10^{-3}$
	Norm	$1.19 \cdot 10^{-1}$	$4.78 \cdot 10^{-2}$	$2.06 \cdot 10^{-2}$	$9.21 \cdot 10^{-3}$	$4.20 \cdot 10^{-3}$
$t = 25$	Sim	$4.46 \cdot 10^{-2}$	$9.53 \cdot 10^{-3}$	$2.11 \cdot 10^{-3}$	$5.58 \cdot 10^{-4}$	$1.37 \cdot 10^{-4}$
	large $t$	$6.10 \cdot 10^{-2}$	$1.16 \cdot 10^{-2}$	$2.55 \cdot 10^{-3}$	$5.93 \cdot 10^{-4}$	$1.43 \cdot 10^{-4}$
	large $\varrho$	$5.96 \cdot 10^{-2}$	$1.13 \cdot 10^{-2}$	$2.48 \cdot 10^{-3}$	$5.75 \cdot 10^{-4}$	$1.38 \cdot 10^{-4}$
	Norm	$3.56 \cdot 10^{-2}$	$5.36 \cdot 10^{-3}$	$8.89 \cdot 10^{-4}$	$1.54 \cdot 10^{-4}$	$2.74 \cdot 10^{-5}$
$t = 50$	Sim	$1.06 \cdot 10^{-2}$	$5.82 \cdot 10^{-4}$	$3.89 \cdot 10^{-5}$	$2.81 \cdot 10^{-6}$	$1.68 \cdot 10^{-7}$
	large $t$	$1.22 \cdot 10^{-2}$	$6.56 \cdot 10^{-4}$	$4.08 \cdot 10^{-5}$	$2.69 \cdot 10^{-6}$	$1.83 \cdot 10^{-7}$
	large $\varrho$	$1.21 \cdot 10^{-2}$	$6.48 \cdot 10^{-4}$	$4.02 \cdot 10^{-5}$	$2.65 \cdot 10^{-6}$	$1.80 \cdot 10^{-7}$
	Norm	$5.78 \cdot 10^{-3}$	$1.78 \cdot 10^{-4}$	$6.10 \cdot 10^{-6}$	$2.20 \cdot 10^{-7}$	$8.17 \cdot 10^{-9}$

approximation equals (9) as well; here it is also used that  $U_i''(\zeta) \approx 2t(1 + \varepsilon)^{3/2} \approx 2t$ .

In Table 2 we present results for  $\varepsilon = 0.2$ . We observe that in this parameter setting the large-deviations-based approximations outperform, for small probabilities, the Normal approximation. Finally in Table 3 the case  $\varepsilon = 0.5$  is considered. It shows that the long-timescale and large-load approximations differ more than in the previous tables, particularly for ‘moderate’ values of  $t$ . The Normal approximation is in some cases very inaccurate: it is in specific cases more than one order of magnitude off. In both tables we again used 10% relative efficiency and 95% confidence.

## Heavy Tails

The analysis presented in this paper relates to the case of light-tailed jobs, as we assume the moment generating function to be finite in an open neighborhood of 0. In the heavy-tailed case the event under consideration will be essentially caused by a number of jobs present during a substantial part of the interval  $[0, t]$ . We now sketch the heuristics of this scenario; a rigorous treatment is beyond the scope of this paper. We focus on the case that  $\mathbb{P}(B > x) \sim L(x)x^{-\alpha}$  for some  $\alpha > 1$  and a slowly-varying  $L(\cdot)$ , i.e.,  $L(tx)/L(x) \rightarrow 1$  as  $x \rightarrow \infty$ , for any  $t > 0$ . In this situation we say that the tail distribution of  $B$  is regularly varying with index  $\alpha$ ; it is known that  $B^r$  is regularly varying with index  $\alpha - 1$ .

Due to the law of large numbers, with overwhelming probability an amount of traffic in the order of  $\varrho t$  will be generated in the interval  $[0, t]$ . In order to make sure that  $\varrho t(1 + \varepsilon)$  is generated, the number of ‘long

jobs', on top of the volume  $\varrho t$ , should be  $\lceil \varrho \varepsilon \rceil$ . The probability of this scenario roughly equals

$$\left( \mathbb{P} \left( B^r > \frac{\varrho \varepsilon t}{\lceil \varrho \varepsilon \rceil} \right) \right)^{\lceil \varrho \varepsilon \rceil},$$

which is regularly varying with index  $\lceil \varrho \varepsilon \rceil \cdot (\alpha - 1)$ . To make this statement rigorous, one has to show that all other scenarios that lead to the event of interest are asymptotically negligible.

## ACKNOWLEDGMENTS

Part of this work was done while MM was at Stanford University, US, and PZ was at Korteweg-de-Vries Institute, University of Amsterdam, the Netherlands. PZ acknowledges support of the academic computer centre CYFRONET AGH (MNiSW/IBM-BC-HS21/AGH/052/2009) and the Polish Ministry of Science and Higher Education (10.420.03; 28.28.120.7026).

## REFERENCES

1. Abramowitz, M.; Stegun, I. *Handbook of Mathematical Functions*; Dover Publications: New York, 1970.
2. Bahadur, R.; Rao, R.R. On deviations of the sample mean. *Ann. Math. Stat.* **1960**, *31*, 1015–1027.
3. Beneš, V. *Mathematical Theory of Connecting Networks and Telephone Traffic*; Academic Press: New York, 1965.
4. Bonin, O. Large deviations theorems for weighted compound Poisson sums. *Probab. Math. Stat.* **2003**, *23*, 357–368.
5. Dembo, A.; Zeitouni, O. *Large Deviations Techniques and Applications*, 2nd Ed.; Springer Verlag: New York, 1998.
6. Ellis, R. Large deviations for a general class of random vectors. *Ann. Probab.* **1984**, *12*, 1–12.
7. Embrechts, P.; Jensen, J.; Maejima, M.; Teugels, J. Approximations for Compound Poisson and Pólya processes. *Adv. Appl. Probab.* **1985**, *17*, 623–637.
8. Feller, W. *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd Ed.; Wiley: New York, 1970.
9. Gärtner, J. On large deviations from the invariant measure. *Th. Probab. Appl.* **1977**, *22*, 24–39.
10. Höglund, T. A unified formulation of the central limit theorem for small and large deviations from the mean. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **1979**, *49*, 105–117.
11. Kelly, F. *Reversibility and Stochastic Networks*; Wiley: Chichester, 1979.
12. Mandjes, M.; van Uitert, M. Sample-path large deviations for tandem and priority queues with Gaussian inputs. *Ann. Appl. Probab.* **2005**, *15*, 1193–1226.
13. Plachky, D.; Steinebach, J. A theorem about probabilities of large deviations with an application to queueing theory. *Per. Math. Hungarica* **1975**, *6*, 343–345.
14. Robert, P. *Stochastic Networks and Queues*; Springer: New York, 2003.
15. Shiryayev, A. *Probability*, 2nd Ed.; Springer: New York, 1996.