## CWI

# Tail asymptotics of the M/G/ model 

M.R.H. Mandjes, P. Żuraniewski

PNA-1002

Centrum Wiskunde \& Informatica (CWI) is the national research institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organisation for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.
CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

## Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)
Modelling, Analysis and Simulation (MAS)
Information Systems (INS)

Copyright © 2010, Centrum Wiskunde \& Informatica
P.O. Box 94079, 1090 GB Amsterdam (NL)

Science Park 123, 1098 XG Amsterdam (NL)
Telephone +31 205929333
Telefax +31 205924199

# TAIL ASYMPTOTICS OF THE M/G/ $\infty$ MODEL 

## M. MANDJES AND P. ŻURANIEWSKI


#### Abstract

This paper considers the so-called $M / G / \infty$ 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 $B$. 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., $A(t):=\int_{0}^{t} N(s) \mathrm{d} s$. 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 $\varrho:=\lambda \mathbb{E} B$. A crucial issue is that $A(t)$ cannot be written as the sum of i.i.d. increments, which makes 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).


2000 Mathematics Subject Classification: 60K25, 60K05
Keywords and Phrases: infinite-server queue, tail asymptotics, large deviations

## 1. Introduction

Consider the so-called $\mathrm{M} / \mathrm{G} / \infty$ 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 $\left(B_{i}\right)_{i \in \mathbb{N}}$, 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., $A(t):=\int_{0}^{t} N(s) \mathrm{d} s$. 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

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

[^0]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 $A(t)$ can not be written as the sum of i.i.d. random variables (such that each of them does not depend on $t$ ). If that would have been possible, then the classical Bahadur-Rao asymptotics would have been applicable [3]. The main contribution of our work is to show that it is possible to adapt the proof of the Bahadur-Rao result to allow (some) dependence between the increments, 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 dependent random variables. Here we note that there is already a body of powerful results on rough, logarithmic asymptotics, most notably the celebrated Gärtner-Ellis result [7, 10, 13]; 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 [3], and in particular the proof of this result as presented in [6]; a substantially more general result is due to Höglund [11]. 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 $[5,8]$.

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.

## 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 $\mathrm{M} / \mathrm{G} / \infty$ 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 $\varrho$ [2]. Denote the load imposed on the system in the interval $[0, t]$ by

$$
A(t):=\int_{0}^{t} N(s) \mathrm{d} s
$$

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

Computation of 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 . It is standard that $A_{-}(t)$ 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

$$
\mathbb{P}\left(B^{r}>t\right)=\frac{1}{\mathbb{E} B} \int_{t}^{\infty} \mathbb{P}(B>s) \mathrm{d} s
$$

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

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

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

$$
\begin{equation*}
\log \mathbb{E} e^{\vartheta A_{+}(t)} \rightarrow \varrho\left(\mathbb{E} e^{\vartheta B^{r}}-1\right)=\varrho\left(\frac{M_{B}(\vartheta)-1}{\vartheta \cdot \mathbb{E} B}-1\right), \tag{1}
\end{equation*}
$$

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 $\lambda t$. Then observe that any arrived job enters the system at a time epoch that is uniformly distributed on $(0, t]$ (and the arrival epochs of the individual jobs, conditional on the number of jobs, are independent). Again distinguishing between the jobs that have left at time $t$ and the jobs that are still present, we obtain

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

Routine calculations yield:

$$
\begin{equation*}
\log \mathbb{E} e^{\vartheta A_{-}(t)}=\lambda W_{t}(\vartheta), \tag{2}
\end{equation*}
$$

where

$$
W_{t}(\vartheta):=-t+\int_{0}^{t}(t-s) f_{B}(s) e^{\vartheta s} \mathrm{~d} s+\int_{0}^{t} e^{\vartheta s} \mathbb{P}(B>s) \mathrm{d} s .
$$

Instantaneous input. We now consider a related model, namely the model in which the arriving jobs transmit their traffic 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 $t \geq 0$,

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

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 $t \rightarrow \infty$,

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

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

$$
\frac{\bar{A}(t)-\varrho t}{\sigma \sqrt{t}} \rightarrow{ }_{\mathrm{d}} N(0,1)
$$

with $N(0,1)$ being a standard-Normal random variable, and ' $\rightarrow \mathrm{d}$ ' denoting convergence in distribution.
The next question is: does $A(t)$ obey the same law of large numbers and centrallimit 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 (1) that

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

Bearing in mind (2), and using

$$
\lim _{t \rightarrow \infty} \lambda \int_{0}^{t} e^{\vartheta s / \sqrt{t}} \mathbb{P}(B>s) \mathrm{d} s=\varrho
$$

and

$$
\begin{aligned}
& \lim _{t \rightarrow \infty}\left(-\lambda t+\lambda \int_{0}^{t}(t-s) f_{B}(s) e^{\vartheta s / \sqrt{t}} \mathrm{~d} s-\vartheta \varrho \sqrt{t}\right) \\
& \quad=\lim _{t \rightarrow \infty}\left(-\lambda t+\lambda \int_{0}^{t}(t-s) f_{B}(s)\left(1+\frac{\vartheta s}{\sqrt{t}}+\frac{\vartheta^{2} s^{2}}{2 t}\right) \mathrm{d} s-\vartheta \varrho \sqrt{t}\right) \\
& \quad=-\varrho+\frac{\sigma^{2}}{2} \vartheta^{2}
\end{aligned}
$$

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

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 Eqn. (2) implies that

$$
\lim _{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E} e^{\vartheta A_{-}(t)}=\lim _{t \rightarrow \infty}\left(-\lambda+\lambda \int_{0}^{\infty}\left(1-\frac{r}{t}\right) 1_{[0, t)}(r) f_{B}(r) e^{\vartheta r} \mathrm{~d} r\right)
$$

which equals, by virtue of 'monotone convergence',

$$
-\lambda+\lambda \int_{0}^{\infty} f_{B}(r) e^{\vartheta r} \mathrm{~d} r=\Lambda(\vartheta)
$$

From (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 [6, Thm. 2.3.6] 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, then we can write

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

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 Poisson $(\lambda)$ number of independent jobs (each of them distributed as the random variable $B$ ), it follows that $\log \mathbb{E} \exp \left(\vartheta X_{i}\right)=\Lambda(\vartheta)$. Now the so-called Bahadur-Rao result [3] can be applied to find the exact asymptotics of $\bar{A}(t) / t$ :

$$
\lim _{t \rightarrow \infty} \sqrt{t} e^{t I(a)} \cdot \mathbb{P}\left(\frac{\bar{A}(t)}{t}>a\right)=C_{a}
$$

for some constant $C_{a}$, given by

$$
C_{a}:=\left(\eta \sqrt{2 \pi \Lambda^{\prime \prime}(\eta)}\right)^{-1}=\left(\eta \sqrt{2 \pi \lambda M_{B}^{\prime \prime}(\eta)}\right)^{-1}
$$

where $\eta \equiv \eta_{a}$ solves $\Lambda^{\prime}(\eta)=a$. The goal of this note is to find the exact asymptotics of $A(t) / t$, with as main question: does $A(t) / t$ have the same exact asymptotics $\bar{A}(t) / t$ ? 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 consist of $t$ i.i.d. increments, being the reason why we cannot use Bahadur-Rao to compute the exact asymptotics of $A(t) / t$. This is, however, not entirely true. It is true that the $A(i-1, i)$ (with $A(\tau, \sigma)$ the amount of traffic entering between $\tau$ and $\sigma$ in our gradual-input system) are not independent (they are identically distributed), but, interestingly, we can still write $A_{-}(t)$ as the sum of $t$ i.i.d. increments (assuming for ease $t$ to be integer). 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 $T_{i j}$. Then $X_{i}$ is the amount of traffic generated by these $N_{i}$ jobs in $(0, t]$. This can be alternatively written as

$$
X_{i}:=\sum_{j=1}^{N_{i}} \min \left\{t-T_{i j}, B_{i j}\right\}
$$

with the $B_{i j}$ 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 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. 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 (2), we get that

$$
\Lambda_{t}(\vartheta):=\log \mathbb{E} e^{\vartheta X(t)}=-\lambda+\lambda \int_{0}^{t}\left(1-\frac{s}{t}\right) f_{B}(s) e^{\vartheta s} \mathrm{~d} s+\lambda \cdot \frac{1}{t} \int_{0}^{t} e^{\vartheta s} \mathbb{P}(B>s) \mathrm{d} s
$$

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

$$
\begin{equation*}
\Lambda(\vartheta)-\Lambda_{t}(\vartheta) \sim \frac{\lambda}{t}\left(\int_{0}^{\infty} s f_{B}(s) e^{\vartheta s} \mathrm{~d} s-\int_{0}^{\infty} e^{\vartheta s} \mathbb{P}(B>s) \mathrm{d} s\right)=\frac{\lambda}{t} \cdot \varphi(\vartheta) ; \tag{3}
\end{equation*}
$$

here

$$
\varphi(\vartheta):=M_{B}^{\prime}(\vartheta)+\frac{1}{\vartheta}-\frac{1}{\vartheta} \cdot M_{B}(\vartheta) \geq 0
$$

where the last inequality (being valid for $\vartheta>0$ ) follows from $M_{B}(\vartheta)-\vartheta M_{B}^{\prime}(\vartheta) \leq$ $M_{B}(0)=1$, due to the convexity of $M_{B}(\vartheta)$. Likewise,
$\Lambda^{\prime}(\vartheta)-\Lambda_{t}^{\prime}(\vartheta) \sim \frac{\lambda}{t} \cdot \psi(\vartheta)$, with $\psi(\vartheta):=M_{B}^{\prime \prime}(\vartheta)-\frac{1}{\vartheta^{2}}+\frac{1}{\vartheta^{2}} \cdot M_{B}(\vartheta)-\frac{1}{\vartheta} \cdot M_{B}^{\prime}(\vartheta)$,
and

$$
\begin{equation*}
\frac{\Lambda^{\prime}(\vartheta)}{\Lambda(\vartheta)}-\frac{\Lambda_{t}^{\prime}(\vartheta)}{\Lambda_{t}(\vartheta)} \sim \frac{\lambda}{t} \cdot \zeta(\vartheta), \text { with } \zeta(\vartheta):=\frac{\psi(\vartheta)}{\Lambda(\vartheta)}-\frac{\varphi(\vartheta)}{\Lambda(\vartheta)} \cdot \frac{\Lambda^{\prime}(\vartheta)}{\Lambda(\vartheta)} \tag{4}
\end{equation*}
$$

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 $g(x)=f(x) \cdot \exp \left(\eta x-\Lambda_{t}(\eta)\right)$. The standard change-of-measure identity yields, in self-evident notation:

$$
\begin{aligned}
& \mathbb{P}\left(\frac{A_{-}(t)}{t} \geq a\right)=\mathbb{E}_{g}\left(e^{-\eta A_{-}(t)+t \Lambda_{t}(\eta)} \cdot 1_{\left\{A_{-}(t) \geq a t\right\}}\right) \\
& \quad=e^{t\left(\Lambda_{t}(\eta)-\eta a_{t}\right)} \cdot \mathbb{E}_{g}\left(e^{-\eta\left(A_{-}(t)-t a_{t}\right)} \cdot 1_{\left\{A_{-}(t) \geq a t\right\}}\right) ;
\end{aligned}
$$

here we use the notation $a_{t}:=\Lambda_{t}^{\prime}(\eta) / \Lambda_{t}(\eta)$. Now note that (3) and (4) imply that

$$
\lim _{t \rightarrow \infty} \frac{\exp \left(t\left(\Lambda_{t}(\eta)-\eta a_{t}\right)\right)}{\exp (t(\Lambda(\eta)-\eta a))}=e^{-\lambda \varphi(\eta)} \cdot e^{\eta \lambda \zeta(\eta)}
$$

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

$$
Y_{i}(t):=\frac{X_{i}(t)-a_{t}}{\sqrt{\Lambda_{t}^{\prime \prime}(\eta)}}, \quad W(s, t):=\frac{1}{\sqrt{t}} \cdot \sum_{i=1}^{t} Y_{i}(s)
$$

Hence, with $F_{s, t}(\cdot)$ the distribution function of $W(s, t)$ under $g$,

$$
\begin{aligned}
& \mathbb{E}_{g}\left(e^{-\eta\left(A_{-}(t)-t a_{t}\right)} \cdot 1_{\left\{A_{-}(t) \geq a t\right\}}\right) \\
& \quad=\mathbb{E}_{g}\left(e^{-\eta \sqrt{\Lambda_{t}^{\prime \prime}(\eta)} \sqrt{t} W(t, t)} \cdot 1_{\left\{W(t, t) \geq\left(a-a_{t}\right) \sqrt{t} / \sqrt{\Lambda_{t}^{\prime \prime}(\eta)}\right\}}\right) \\
& \quad=\int_{\left(a-a_{t}\right) \sqrt{t} / \Lambda_{t}^{\prime \prime}(\eta)}^{\infty} e^{-\eta \sqrt{\Lambda_{t}^{\prime \prime}(\eta)} \sqrt{t} x} \mathrm{~d} F_{t, t}(x) .
\end{aligned}
$$

Define by $\Phi(\cdot)$ the distribution function of a standard normal random variable, and by $\phi(\cdot)$ the corresponding density. Now define and $\xi(x):=\left(1-x^{2}\right) \phi(x)$ and $\gamma:=$ $\sup _{x} \xi(x) \in(0, \infty)$. Observe that for any $s$, due to the fact that $F_{s, t}(x)$ decreases in $s$, it holds that

$$
\begin{align*}
& \lim _{t \rightarrow \infty} \sqrt{t}\left(\sup _{x}\left(F_{t, t}(x)-\Phi(x)-\frac{m_{3}}{6 \sqrt{t}} \xi(x)\right)\right)  \tag{5}\\
& \leq \lim _{t \rightarrow \infty} \sqrt{t}\left(\sup _{x}\left(F_{s, t}(x)-\Phi(x)-\frac{m_{3, s}}{6 \sqrt{t}} \xi(x)\right)\right)+ \\
& \quad \sup _{x}\left(\frac{1}{6} \cdot\left(m_{3, s}-m_{3}\right) \xi(x)\right) \\
& =\frac{\gamma}{6}\left(m_{3, s}-m_{3}\right) ;
\end{align*}
$$

here $m_{3}$ is the third moment of $Y_{i}(\infty)$, and $m_{3, s}$ the third moment of $Y_{i}(s)$, both under $g$. As we can pick $s$ arbitrarily, we can let $s$ grow to $\infty$, and we obtain that the limit (5) is bounded above by 0 . In addition, $F_{t, t}(x) \geq F_{\infty, t}(x)$, and we know from the Berry-Esseen-type estimates in [9, Eqn. XVI.4.1] that

$$
\sqrt{t}\left(\sup _{x}\left(F_{\infty, t}(x)-\Phi(x)-\frac{m_{3}}{6 \sqrt{t}} \xi(x)\right)\right) \rightarrow 0
$$

We conclude that (5) equals 0 . Now we can finish the proof as in [6, Thm. 3.7.4], as we have the counterpart of their Eqn. (3.7.8), but now for i.i.d. random variables that (mildly) depend on $t$. As a result, we can essentially replace the distribution function $F_{t, t}(\cdot)$ by that of a standard Normal random variable. It follows that, for $t \rightarrow \infty$, with $N\left(\mu, \sigma^{2}\right)$ denoting, as usual a Normally distributed random variable with mean $\mu$ and variance $\sigma^{2}$,

$$
\begin{aligned}
& \int_{\left(a-a_{t}\right) \sqrt{t} / \sqrt{\Lambda_{t}^{\prime \prime}(\eta)}}^{\infty} e^{-\eta \sqrt{\Lambda_{t}^{\prime \prime}(\eta)} \sqrt{t} x} \mathrm{~d} F_{t, t}(x) \sim \int_{\left(a-a_{t}\right) \sqrt{t} / \sqrt{\Lambda_{t}^{\prime \prime}(\eta)}}^{\infty} e^{\eta \sqrt{\Lambda_{t}^{\prime \prime}(\eta)} \sqrt{t} x} \mathrm{~d} \Phi(x) \\
& \quad=\frac{1}{\sqrt{2 \pi}} e^{\eta^{2} \Lambda_{t}^{\prime \prime}(\eta) t} \int_{\left(a-a_{t}\right) \sqrt{t} / \sqrt{\Lambda_{t}^{\prime \prime}(\eta)}}^{\infty} e^{-\frac{1}{2}\left(x+\eta \sqrt{\Lambda_{t}^{\prime \prime}(\eta)} \sqrt{t}\right)^{2}} \mathrm{~d} x \\
& \quad=e^{\eta^{2} \Lambda_{t}^{\prime \prime}(\eta) t} \mathbb{P}\left(N\left(-\eta \sqrt{\Lambda_{t}^{\prime \prime}(\eta)} \sqrt{t}, 1\right)>\frac{\left(a-a_{t}\right) \sqrt{t}}{\Lambda_{t}^{\prime \prime}(\eta)}\right) \\
& \quad \sim\left(1-\Phi\left(\eta \sqrt{\Lambda_{t}^{\prime \prime}(\eta)} \sqrt{t}+\frac{\lambda \zeta(\eta)}{\sqrt{\Lambda_{t}^{\prime \prime}(\eta)} \sqrt{t}}\right)\right) e^{\frac{1}{2} \eta^{2} \Lambda_{t}^{\prime \prime}(\eta) t} \sim \frac{e^{-\eta \lambda \zeta(\eta)}}{\eta \sqrt{2 \pi \Lambda^{\prime \prime}(\eta)}} \cdot \frac{1}{\sqrt{t}} ;
\end{aligned}
$$

in the last step the standard estimate $1-\Phi(x) \sim(2 \pi x)^{-1 / 2} e^{-x^{2} / 2}$ (as $x \rightarrow \infty$ ) was applied [1]. 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

$$
\lim _{t \rightarrow \infty} \sqrt{t} e^{t I(a)} \cdot \mathbb{P}\left(\frac{A_{-}(t)}{t}>a\right)=C_{a} \cdot e^{-\lambda \varphi(\eta)}
$$

Remark 1. It is an easy exercise in calculus to show that we can rewrite

$$
\lambda \varphi(\eta)=a-\frac{\lambda}{\eta}\left(M_{B}(\eta)-1\right)
$$

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 $\check{X}_{1}$ up to $\check{X}_{t}$, where

$$
\begin{equation*}
\check{X}_{i}:=\sum_{j=1}^{N_{i}} \min \left\{t-T_{i j}, B_{i j}\right\}+\sum_{j=1}^{M_{i}} \min \left\{t, B_{i j}^{r}\right\}, \tag{6}
\end{equation*}
$$

with $N_{i}$ and the $B_{i j}$ as before, and $M_{i}$ a Poisson random variable with mean $\varrho / t$ and the $B_{i j}^{r}$ i.i.d., distributed as the random variable $B^{r}$; in addition, both sums in (6) 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

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

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

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

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

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

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 $e^{-\lambda \varphi(\eta)}$. $\mathbb{E} e^{\eta A_{+}(\infty)}$ is always smaller or larger than 1 . This makes sense as in the gradualinput 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 that the criterion $e^{-\lambda \varphi(\eta)} \cdot \mathbb{E} e^{\eta A_{+}(\infty)}<1$ can be rewritten as

$$
\frac{M_{B}(\eta)-1}{\eta}>\frac{M_{B}^{\prime}(\eta)+M_{B}^{\prime}(0)}{2}=\frac{M_{B}^{\prime}(\eta)+\mathbb{E} B}{2}
$$

which can in turn be written as

$$
\begin{equation*}
H_{1}(\eta):=2\left(M_{B}(\eta)-1\right)<\eta\left(M_{B}^{\prime}(\eta)+\mathbb{E} B\right)=: H_{2}(\eta) . \tag{7}
\end{equation*}
$$

The rest of the proof is devoted to establishing this inequality for $\eta>0$.
To this end, first observe that $M_{B}^{\prime}(\cdot)$ is convex. Hence $M_{B}^{\prime \prime}(\eta)>\left(M_{B}^{\prime}(\eta)-M_{B}^{\prime}(0)\right) / \eta$, which is equivalent to

$$
M_{B}^{\prime}(\eta)-\eta M_{B}^{\prime \prime}(\eta)<\mathbb{E} B
$$

This implies (7), as $H_{1}(0)=H_{2}(0)=2$ and

$$
H_{1}^{\prime}(\eta)=2 M_{B}^{\prime}(\eta)<M_{B}^{\prime}(\eta)+\eta M_{B}^{\prime \prime}(\eta)+\mathbb{E} B=H_{2}^{\prime}(\eta) .
$$

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

$$
U_{t}(\vartheta):=V_{t}(\vartheta)+W_{t}(\vartheta) ;
$$

recall that $V_{t}(\vartheta)$ and $W_{t}(\vartheta)$ were defined in Section 2. The Bahadur-Rao result then entails that, for $s>t$,

$$
\begin{equation*}
\lim _{\varrho \rightarrow \infty} \sqrt{\varrho} e^{\varrho J(s)} \cdot \mathbb{P}\left(\frac{A_{\varrho}(t)}{\varrho} \geq s\right)=K_{s} \tag{8}
\end{equation*}
$$

where $J(s):=\sup _{\vartheta}\left(\vartheta s-U_{t}(\vartheta)\right)$, and

$$
K_{s}:=\left(\zeta \sqrt{2 \pi U_{t}^{\prime \prime}(\zeta)}\right)^{-1}
$$

where $\zeta \equiv \zeta_{s}$ solves $U_{t}^{\prime}(\vartheta)=s$.
Numerical example: $M / M / \infty$. In this example we consider the case of exponentially distributed jobs. The goal is to compare the long-timescale asymptotics of Thm. 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

$$
\Lambda^{\prime \prime}(\eta)=\lambda M_{B}^{\prime \prime}(\eta)=2 \frac{a^{3 / 2}}{\sqrt{\lambda \mu}}
$$

and

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

Noting that

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

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

$$
\begin{align*}
\mathbb{P}\left(\frac{A_{\varrho}(t)}{t} \geq a\right) & \sim \frac{C_{a}}{\sqrt{t}} e^{-(t+1 / \mu)(\sqrt{\mu a}-\sqrt{\lambda})^{2}}  \tag{9}\\
& \sim \mathbb{P}\left(\frac{\overline{A_{\varrho}}(t)}{t} \geq a\right) \cdot e^{-(1 / \mu)(\sqrt{\mu a}-\sqrt{\lambda})^{2}}
\end{align*}
$$

|  |  | $\varrho=5$ | $\varrho=50$ | $\varrho=100$ | $\varrho=150$ |
| :--- | :--- | :---: | :---: | :---: | :---: |
| $t=100$ | 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 | $2.40 \cdot 10^{-1}$ | $1.27 \cdot 10^{-2}$ | $7.81 \cdot 10^{-4}$ | $5.35 \cdot 10^{-5}$ |

Table 1. Numerical results, $\varepsilon=0.01$

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

$$
V_{t}(\vartheta)=\frac{\vartheta}{\mu-\vartheta}\left(1-e^{-(\mu-\vartheta) t}\right), \quad W_{t}(\vartheta)=\frac{\vartheta}{(\mu-\vartheta)^{2}}\left(1-e^{-(\mu-\vartheta) t}\right)+t \frac{\vartheta}{\mu-\vartheta}
$$

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

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

Straightforward algebra leads to expressions for $U_{t}^{\prime}(\vartheta)$ and $U_{t}^{\prime \prime}(\vartheta)$. Now Eqn. (8) enables us to compute the large-load asymptotics.

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

$$
\mathbb{P}\left(A_{\varrho}(t) \geq \varrho t(1+\varepsilon)\right) \approx \Phi\left(-\frac{\varrho t \varepsilon}{\sqrt{\mathbb{V a r} A_{\varrho}(t)}}\right)
$$

with $\Phi(\cdot)$ being the standard Normal distribution function. From the above formulae we also obtain that $\mathbb{V a r} A_{\varrho}(t)=\Lambda_{t}^{\prime \prime}(0)=2 \varrho\left(t-1+e^{-t}\right)$, cf. [4, 12].

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

$$
\begin{equation*}
\mathbb{P}\left(A_{\varrho}(t) \geq \varrho t(1+\varepsilon)\right) \approx \frac{1}{\varepsilon \sqrt{\pi \varrho t}} e^{-\frac{1}{4}(\varrho t) \varepsilon^{2}} \tag{10}
\end{equation*}
$$

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 (9), we obtain expression (10) 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, Eqn. (8)

|  |  | $\varrho=1$ | $\varrho=5$ | $\varrho=8$ |
| :--- | :--- | :---: | :---: | :---: |
| $t=50$ | 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 2. Numerical Results, $\varepsilon=0.2$
yields that in this regime the large-load approximation equals (10) as well; here it is also used that $U_{t}^{\prime \prime}(\zeta) \approx 2 t(1+\varepsilon)^{3 / 2} \approx 2 t$.
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.

Heary 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., $L(t x) / L(x) \rightarrow 1$ 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.

|  |  | $\varrho=1$ | $\varrho=2$ | $\varrho=3$ | $\varrho=4$ | $\varrho=5$ |
| :--- | :--- | :---: | :---: | :---: | :---: | :---: |
| $t=10$ | 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 | $5.78 \cdot 10^{-3}$ | $1.78 \cdot 10^{-4}$ | $6.10 \cdot 10^{-6}$ | $2.20 \cdot 10^{-7}$ | $8.17 \cdot 10^{-9}$ |

Table 3. Numerical Results, $\varepsilon=0.5$

## REFERENCES

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

Korteweg-de Vries Institute for Mathematics, University of Amsterdam, the Netherlands; Eurandom, Eindhoven University of Technology, the Netherlands; CWI, Amsterdam, the Netherlands
E-mail address: m.r.h.mandjes@uva.nl
Dept. of Applied Mathematics, AGH University of Science and Technology, Kraków, Poland.
E-mail address: piotr.zuraniewski@agh.edu.pl

Centrum Wiskunde \& Informatica (CWI) is the national research institute for mathematics and computer science in the Netherlands. The institute's strategy is to concentrate research on four broad, societally relevant themes: earth and life sciences, the data explosion, societal logistics and software as service.

Centrum Wiskunde \& Informatica (CWI) is het nationale onderzoeksinstituut op het gebied van wiskunde en informatica. De strategie van het instituut concentreert zich
op vier maatschappelijk relevante
onderzoeksthema's: aard- en
levenswetenschappen, de data-explosie,
maatschappelijke logistiek en software als service.

Bezoekadres:
Science Park 123
Amsterdam
Postadres:
Postbus 94079, 1090 GB Amsterdam
Telefoon 0205929333
Fax 0205924199
info@cwi.nl
www.cwi.nl

Centrum Wiskunde \& Informatica


[^0]:    Date: January 25, 2010.
    Part of this work was done while MM was at Stanford University, Stanford, CA 94305, US, and PZ was at Korteweg-de Vries Institute, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Amsterdam, the Netherlands.

