

## Another look at the transient behavior of the M/G/1 workload process

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

Fralix, B. H. (2009). *Another look at the transient behavior of the M/G/1 workload process*. (Report Eurandom; Vol. 2009017). Eurandom.

**Document status and date:**

Published: 01/01/2009

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

# Another look at the transient behavior of the $M/G/1$ workload process

Brian H. Fralix  
Department of Mathematical Sciences  
Clemson University  
Clemson, SC USA

June 24, 2009

## Abstract

We use Palm measures, along with a simple approximation technique to derive new explicit expressions for all of the transient moments of the workload process of an  $M/G/1$  queue. These expressions can also be used to derive a closed-form expression for the  $n$ th moment of the stationary workload, which solves the well-known Takács recursion that generates the waiting time moments of an  $M/G/1$  queue that serves customers in a first-come-first-serve manner.

**Keywords:** Palm distribution, transient behavior,  $M/G/1$  queue,  $M/G/1$  workload, preemptive-resume discipline, waiting time recursion

**MSC:** 60K25, 60G55

## 1 Introduction

We consider the workload process of an  $M/G/1$  queue that serves its customers in a work-conserving manner. Examples of such service disciplines include the First-Come-First-Served (FCFS) discipline, the Last-Come-First-Served (LCFS) discipline (both nonpreemptive and preemptive-resume), and the Processor-Sharing discipline. This process has been heavily studied in the literature, and is one of the classical queueing models, so we will not dare attempt to give a complete list of references which feature studies on the workload. Readers interested in a classical study of the workload process can refer to the textbooks of Kleinrock [21], and Prabhu [23].

Our study will focus on the transient, or time-dependent behavior of the workload process. Some examples of papers which discuss the transient behavior of the workload include Takács [25], Ott [22], Cox and Isham [15], and Abate and Whitt [5, 6]. Other recent studies of a more general nature (i.e. for reflected Lévy processes) can be found in the recent works of Es-Saghouani and Mandjes [16] and Andersen and Mandjes [7].

The dynamics of the  $M/G/1$  workload can be described as follows. We assume that customers arrive according to a Poisson process with rate  $\lambda > 0$ , and each customer brings with it to the system a random, arbitrarily distributed amount of work, which is independent of the time at which the customer arrived, along with the arrival times of all other customers visiting the system. This generates a sequence of service times, and we assume that this sequence is independent and identically distributed.

A classical method used to model the  $M/G/1$  workload is to make use of what is known as a one-dimensional reflection map; see, for example, [13]. To be more precise, let  $Y := \{Y(t); t \geq 0\}$  be a stochastic process which is of the form

$$Y(t) = Y(0) + \sum_{n=1}^{N(t)} B_n - t$$

where  $N$  represents our Poisson process with rate  $\lambda$ ,  $\{B_k\}_{k \geq 1}$  our sequence of i.i.d. service times, and  $Y(0)$  is a random variable which we assume is independent of all other random elements in our model. In our paper,  $Y(0)$  will actually be a deterministic value. In particular,  $Y(0)$  can be interpreted as the initial workload found in the queueing system at time zero.

We now define another stochastic process  $W := \{W(t); t \geq 0\}$  which is a functional of  $Y$ . In particular,

$$\begin{aligned} W(t) &= \max(Y(t), Y(t) - \inf_{0 \leq s \leq t} Y(s)) \\ &= Y(t) - \min(0, \inf_{0 \leq s \leq t} Y(s)). \end{aligned}$$

Under this representation,  $W(t)$  can be interpreted as the workload currently found in an  $M/G/1$  queue at time  $t$ : moreover,  $W(t) = Y(t)$  if and only if  $Y(0) > 0$  and  $\inf_{0 \leq s \leq t} Y(s) > 0$ , and  $W(t) = Y(t) - \inf_{0 \leq s \leq t} Y(s)$  if and only if  $Y$  hits zero at some point in  $[0, t]$ , which corresponds to the end of the first busy period of the  $M/G/1$  queue. If the first busy period ends before time  $t$ , then by subtracting the infimum of  $Y$  over  $[0, t]$  we are forcing  $W(t)$  to be zero while the queueing system is empty.

This representation of the workload has proven to be extremely useful in many queueing studies, particularly when attempting to show that the sample paths of the workload can be approximated by a simpler stochastic process, such as a regulated Brownian motion.

It is also well-known that if  $\rho := \lambda E[B_1] < 1$ , the  $M/G/1$  queue is stable, and as  $t \rightarrow \infty$ ,  $W(t)$  converges weakly to a random variable  $W(\infty)$ , where for each  $x \geq 0$

$$P(W(\infty) \leq x) = (1 - \rho) \sum_{n=0}^{\infty} \rho^n P\left(\sum_{k=1}^n B_k^* \leq x\right) \quad (1)$$

where  $\{B_k^*\}_{k \geq 1}$  represents an i.i.d. sequence of random variables, with distribution

$$P(B_1^* \leq x) = \frac{1}{E[B_1]} \int_0^x P(B_1 > t) dt.$$

Identity (1) is known as the Pollaczek-Khintchine formula.

One way to see why the Pollaczek-Khintchine formula holds is to imagine that the workload process corresponds to that of an  $M/G/1$  queue which operates under the preemptive-resume LCFS discipline. Queueing systems which serve customers in this manner operate as follows: when a customer arrives to the system, it immediately begins receiving service, and the customer that was previously receiving service up to that point then waits in the system until the customer that interrupted its moment with the server leaves the system, at which point it rejoins its place with the server. Once the server revisits a customer it previously abandoned, it resumes service at the point where it previously left-off. The best way to visualize such a system is to imagine the customers waiting in a stack, where newly arriving customers to the system are placed on top of the stack, and at each moment in time, the server serves whomever is currently on top of the stack.

Under these assumptions, we see that this is a work-conserving discipline, and it is well-known that the workload process corresponding to this system is the same for all  $M/G/1$  queues which operate under a work-conserving discipline. In particular, in steady-state, the stationary joint distribution of the queue-length (including the customer in service) and the residual service times (which we denote as  $V_k(\infty)$ ) of an  $M/G/1$  queue under preemptive-resume LCFS is

$$P(Q(\infty) = n, V_k(\infty) \leq x_k, 1 \leq k \leq n) = (1 - \rho) \rho^n \prod_{k=1}^n P(R_B \leq x_k)$$

where  $R_B$  denotes the residual, or equilibrium distribution, of  $B$ . Once this is known, a simple conditioning argument can be used to derive the Pollaczek-Khintchine formula. This observation was first discovered by Cooper and Niu [14].

Moreover, the moments of  $W(\infty)$  satisfy the following recursion:

$$E[W(\infty)^k] = \frac{\lambda}{1-\rho} \sum_{i=1}^k \binom{k}{i} \frac{E[B_1^{i+1}]}{(i+1)} E[W(\infty)^{k-i}]. \quad (2)$$

These results were first discovered by Takács [25], and are also discussed in the textbook of Kleinrock [21].

Our study of the transient behavior of the  $M/G/1$  workload is closely aligned with what was carried out in [6], but it has also been influenced by [14]. In [6], the authors showed that the transient moments of the workload can be related to one another through a system of differential equations. To be precise, if we (using their notation) define  $m_k(t) = m_k(t, x) = E[W(t)^k | W(0) = x]$ , then the relationship can be summarized in the following theorem.

**Theorem 1.1** (*Theorem 1 of [6]*).

(a)  $m_1'(t) = \rho - 1 + P(W(t) = 0 | W(0) = x)$ .

(b) If  $m_k(t) < \infty$  for some  $k$ ,  $k \geq 2$ , then the derivative  $m_k'(t)$  exists and

$$m_k'(t) = \rho E[B^k] - (1-\rho)k m_{k-1}(t) + \rho \sum_{j=2}^{k-1} \binom{k}{j} E[B^j] m_{k-j}(t). \quad (3)$$

(c) If  $m_{k+1}(0) < \infty$  and  $E[B^{k+1}] < \infty$  for some  $k$ ,  $k \geq 1$ , then  $m_{k+1}'(t) \rightarrow 0$  as  $t \rightarrow \infty$  and

$$m_k(\infty) = \frac{\lambda}{1-\rho} \sum_{j=1}^k \binom{k}{j} \frac{E[B^{j+1}]}{j+1} m_{k-j}(\infty). \quad (4)$$

Furthermore, the authors show in [6] that (4) is a consequence of (3), and this recursion is just (2). They also show that, given  $W(0) = 0$ ,  $W(t)$  is stochastically increasing in  $t$ : hence, the  $k$ th moment of  $W(t)$  can be expressed as the  $k$ th moment of  $W(\infty)$ , times a cumulative distribution function. Once this is known, the study of the transient behavior reduces to the study of the cumulative distribution function, and how it can be approximated. A crucial point in the analysis is that all of these distribution functions depend on the idle probabilities  $P(W(t) = 0 | W(0) = 0)$ ,  $t \geq 0$ , which at the time could only be expressed in terms of transforms. Not only that, only the first four moments are derived explicitly in [6]: while the method could be used in principle to derive any moment of interest, it is not clear how to use this method to derive an explicit formula which describes all of the moments of  $W(t)$ .

It is also possible to derive (3) by making use of the “BAR” (Basic Adjoint Relation) approach; see Chen [12]. This was originally used in queueing to derive the stationary distributions of multi-dimensional regulated Brownian motion, and it is shown in [12] that the method is also useful in a transient setting as well.

Our contribution to the transient literature on the  $M/G/1$  workload is as follows: we show that all of the moments of  $W(t)$  can be expressed in terms of cumulative distribution functions, which correspond to different types of independent sums of residual busy periods, which are generated by various residuals of a typical service time. For instance, we will show in the next section that

$$E[W(t) | W(0) = x] = \int_0^x P(\tau_u > t) du + \lambda E[B] \int_0^t P(\tau_{R_B} > s) ds \quad (5)$$

where  $\tau_x$  represents the amount of time it takes the workload to go from level  $x$  to level zero, and

$$\tau_X = X + \sum_{k=1}^{N(X)} \tau_{B_k}$$

where  $X$  is independent of all other random elements present in the random summation.

From this expression we observe the well-known fact that if  $E[B^2] < \infty$ , the first moment of  $W(t)$  converges to the first moment of  $W(\infty)$ , as  $t \rightarrow \infty$ . Moreover, under this condition we find that

$$E[W(t)|W(0) = x] = \int_0^x P(\tau_u > t) du + \frac{\lambda E[B^2]}{2(1-\rho)} P(R_{\tau_{RB}} \leq t). \quad (6)$$

The form of this moment is very similar to the transient moments of the  $M/G/1$  queue that operates under the preemptive Last-Come-First-Served discipline, which we provide for the reader's convenience (details can be found in [19]):

$$E[Q(t)|Q(0) = n_0] = \sum_{k=1}^{n_0} P(\tau_1 + \dots + \tau_k > t) + \frac{\rho}{1-\rho} P(R_{\tau_B} \leq t). \quad (7)$$

Similar expressions also hold for higher moments of  $W(t)$ , as we will show below. Indeed, these expressions are very similar to the type of expressions derived in [1, 2, 3, 4] for regulated Brownian motion, and the  $M/M/1$  queue length process.

What is also interesting about our argument is that it makes heavy use of the idea of relating the workload process to a queueing system which operates under the Last-Come-First-Served preemptive-resume discipline. In [14], the authors used such an idea to give a probabilistic description of the Pollaczek-Khintchine formula. Similarly, in our work we will also show that the sample paths of the  $M/G/1$  workload can be approximated with (properly scaled) sample paths of an  $M/G/1$  system, with batch arrivals, which operates under the preemptive-resume LCFS discipline. Queueing systems which behave in a preemptive-resume manner are generally much easier to analyze than those which behave otherwise, and so this allows us to see that the workload process is quite tractable as well.

We begin our discussion in Section 2, where we briefly define the notion of a Palm measure, which will be the main technical tool used throughout this study. Section 3 will focus on the derivation of the first moment, and this expression will be used to derive the covariance function of the stationary version of the workload process. In Section 4, we will show how our technique can also be used to derive all of the moments of the transient workload. The reader will see that our expressions are explicit, and so they can be used to derive closed-form expressions for the steady-state workload as well, which solves the well-known recursion (2) discussed above. Our transient moments can be expressed in terms of cumulative distribution functions (cdfs) which represent various types of independent sums of busy periods generated by residuals of a typical service time, and we will conclude in Section 5 by showing how our Palm approach can be used to derive all of the moments of these cdfs as well. Being able to derive these moments is important, as they allow us to create simple approximations of our cdfs, which leads to simple approximations of the moments of the transient workload.

## 2 Palm Measures

We begin by assuming that all of the random elements discussed throughout this study are defined on a probability space  $(\Omega, \mathcal{F}, P)$ , where  $\Omega$  represents our sample space which is assumed to be a complete, separable metric space,  $\mathcal{F}$  represents the  $\sigma$ -field generated by the open sets with respect to the metric on  $\Omega$ , and  $P$  is a probability measure on  $(\Omega, \mathcal{F})$ . These topological assumptions on  $(\Omega, \mathcal{F})$

are needed in order for the Palm measures to exist; see Chapter 10 of [20] for further discussion. Readers should not be alarmed by the fact that we are placing such assumptions on our space, as the vast majority of queueing models found in the literature can exist in such a setting.

Each point process  $\{N(t); t \geq 0\}$  induces a family of probability measures  $\{P_t\}_{t \in [0, \infty)}$ , which are known as Palm measures.

**Definition 2.1** *The collection of Palm measures  $\{P_t\}_{t \in [0, \infty)}$  induced by a point process  $N$  are probability measures with the property that (i) for each fixed Borel set  $A \in \mathcal{F}$ ,  $P_t(A)$  is a measurable function in  $t$ , and (ii) for each fixed  $t$ ,  $P_t$  is a probability measure on  $(\Omega, \mathcal{F})$ . Moreover, for each Borel set  $B \subset \mathbb{R}$  and each  $A \in \mathcal{F}$ , we find that*

$$E[N(B)\mathbf{1}(A)] = \int_B P_t(A)\mu(dt)$$

where  $\mu$  represents the mean measure of  $N$ .

A very important consequence of this definition is what is known as the Campbell-Mecke formula, which we now present.

**Theorem 2.1** *Suppose  $X(t, \omega)$  is a measurable function on  $[0, \infty) \times \Omega$ . Then*

$$E \left[ \int_0^\infty X(t, \omega) N(dt) \right] = \int_0^\infty E_t[X(t, \omega)] \mu(dt).$$

This is, quite possibly, the most important tool used when applying Palm measures to a particular problem, and we will find that it is an invaluable tool in our setting as well. Readers interested in Palm measures should refer to the research monograph of Kallenberg [20]: examples of its use in queueing applications include Blaszczyszyn [10], Blaszczyszyn et al. [11], Rolski [24], Fralix and Riaño [19], and Fralix [18].

We will also make use of higher-order Palm measures in our paper, which are of the form  $P_{s_1, \dots, s_n}$ , where  $s_1, \dots, s_n$  are points in  $[0, \infty)$ . We interpret, for a given  $A \in \mathcal{F}$ ,  $P_{s_1, \dots, s_n}(A)$  as the probability of the event  $A$ , given that  $N$  has points at  $s_1, s_2, \dots, s_n$ . Moreover, in our case  $N$  will be a Poisson process (under  $P$ ), and under  $P_{s_1, \dots, s_n}$ ,  $N$  is basically Poisson, with extra points at the locations  $s_1, \dots, s_n$ . We refer readers that are interested in properties of these higher-moment Palm measures to Chapter 12 of [20].

### 3 The First Moment, and the Covariance Formula

We begin by computing the first moment of  $W(t)$ , given  $W(0) = x$ , for some  $x \geq 0$ . This will illustrate the main idea used in all of the computations performed throughout this paper, without unnecessarily complicating the matter with combinatorial arguments, iterated integrals, and residual calculations, which will be required for higher moments.

To compute our expressions for the moments of  $W(t)$ , we will need to make use of a few properties of residuals of random variables.

**Definition 3.1** *Suppose  $X$  is a nonnegative random variable. The residual of  $X$  is a nonnegative random variable  $R_X$  such that, for each  $t \geq 0$ ,*

$$P(R_X \leq t) = \frac{1}{E[X]} \int_0^t P(X > y) dy$$

We will also need to calculate “residuals of residuals”, and so we define  $R_{n,X}$  as the  $n$ th residual of  $X$ , i.e.  $R_{n,X} = R_{1, R_{n-1,X}} = R_{R_{n-1,X}}$ . The use of the  $n$ th residual of a random variable will prove to be useful later, when we derive the higher moments of the workload.

Consider the following collection of queueing processes  $\{Q_M(t); t \geq 0\}$ , for  $M \geq 1$ . For each  $Q_M$ , we assume that groups of customers arrive according to a Poisson process  $N$ , with rate  $\lambda$ . We signify the points of  $N$  with the sequence  $\{T_k\}_{k \geq 1}$ , where  $0 < T_1 < T_2 < \dots$ . To be more precise, at each one of these Poisson arrival times  $T_n$  we assume that  $C_{n,M}$  customers arrive, where each customer brings with it a deterministic amount of work which is of size  $1/2^M$ . In particular, we define  $C_{n,M}$  as follows:

$$C_{n,M} = \sum_{k=1}^{\infty} k \mathbf{1} \left( \frac{k-1}{2^M} < B_n \leq \frac{k}{2^M} \right).$$

The queue then processes all work in a preemptive-resume LCFS manner; hence, we assume that associated with each customer in the batch is a number  $k$ ,  $1 \leq k \leq C_{n,M}$ , where customer 1 is served first, customer 2 second, and so on. Moreover, for each time  $s$  let  $C_M(s)$  denote the size of the last batch to arrive to the system at or before time  $s$ , and let  $W_{k,M}(s)$  denote the sojourn time of the  $k$ th customer within the last batch that arrived at or before time  $s$ . If there is no  $k$ th customer in the batch, we set  $W_{k,M}(s) = 0$ .

We will also assume that at time zero, there are  $m_{0,M} = \sup\{n \geq 1; n/2^M \leq x\}$  customers present in the system. At time zero, the amount of work being held by each of these customers is also equal to  $1/2^M$ , and they are ordered in the same manner as a typical batch, with  $W_{k,M}$  denoting the sojourn time of the  $k$ th customer in the group.

Notice that now  $Q_M$  is an integer-valued process, and due to our choice of service discipline we can express it in the following way: for each time  $t \geq 0$ ,

$$Q_M(t) = \sum_{k=1}^{m_{0,M}} \mathbf{1}(W_{k,M} > t) + \int_0^t \sum_{k=1}^{C_M(s)} \mathbf{1}(W_{k,M}(s) > t-s) N(ds)$$

Our first result shows that, when properly scaled, the sample paths of the  $Q_M$  processes approach, as  $M \rightarrow \infty$ , the workload process. This is a simple and very intuitive (but not very deep) proposition, and it will be of great use when we derive the transient moments of the workload.

**Proposition 3.1** *As  $M \rightarrow \infty$ , for each  $t \geq 0$ ,*

$$\sup_{0 \leq s \leq t} \left| \frac{Q_M(s)}{2^M} - W(s) \right| \rightarrow 0.$$

*In other words,  $Q_M/2^M$  converges to  $W$  uniformly on compact sets.*

**Proof** Clearly, after some thought we see that, since the arrival times are exactly the same for each  $M \geq 1$ ,

$$\sup_{0 \leq s \leq t} \left| \frac{Q_M(s)}{2^M} - W(s) \right| \leq \frac{2(N(t) + 1)}{2^M}$$

which approaches 0 as  $M \rightarrow \infty$ . If the reader does not see the upper bound on the supremum, consider the case where  $N(t) = 0$ . In this case, it is clear that, for each  $M$ , the maximum difference between the scaled queue-length and the workload is just

$$\max \left( \frac{Q_M(0)}{2^M} - W(0), Q_M(1/2^M -) - W(1/2^M -) \right) \leq \frac{2}{2^M}$$

since it is clear that  $\frac{Q_M(0)}{2^M} - W(0) = \frac{Q_M(1/2^M)}{2^M} - W(1/2^M)$ . The same type of logic can be used for any realization of  $N(t)$ , which is always finite.  $\diamond$

One consequence of this proposition is that the hitting times of the scaled queue-length processes converge to the corresponding hitting-times of the workload. To be more precise, define, for each  $M \geq 1$ ,

$$\tau_{k,M} = \inf\{t > 0 : Q_M(0) = k, Q_M(t) = 0\}$$

and

$$\tau_x = \inf\{t > 0 : W(0) = x, W(t) = 0\}.$$

Note that, in the definition of  $\tau_x$ , we are placing the condition that  $W(0) = x$  in the statement; this is merely to remind the reader that this random variable is the amount of time it takes the workload to reach state zero, starting from state  $x$  (similarly for the  $\tau_{k,M}$  definition). With this proposition in mind, an application of the continuous mapping theorem (see [26]) shows that if we let  $k(M, y) = \sup\{k \geq 1 : k/2^M \leq y\}$ , then  $\tau_{k(M,y),M}$  converges weakly, as  $M \rightarrow \infty$ , to  $\tau_y$ .

We are now ready to compute the first moment of the workload. For convenience, we will assume that  $E[B^2] < \infty$ . This assumption is not really needed (as can be seen from the proof), but it does allow us to rewrite an integral as a cumulative distribution function.

**Theorem 3.1** *For each  $t \geq 0$ ,  $x \geq 0$ ,*

$$E[W(t)|W(0) = x] = \int_0^x P(\tau_y > t)dy + \frac{\lambda E[B^2]}{2(1-\rho)} P(R_{\tau_{RB}} \leq t).$$

**Proof** We will first assume that the service time distribution of each  $B_k$  is absolutely continuous. Notice that the first moment of each scaled queue-length process  $Q_M$  is as follows:

$$\begin{aligned} E\left[\frac{Q_M(t)}{2^M}\right] &= \sum_{k=1}^{m_{0,M}} P(\tau_{k,M} > t) \frac{1}{2^M} + \frac{1}{2^M} E\left[\int_0^t \sum_{k=1}^{C_M(s)} \mathbf{1}(W_{k,M}(s) > t-s) N(ds)\right] \\ &= \sum_{k=1}^{m_{0,M}} P(\tau_{k,M} > t) \frac{1}{2^M} + \lambda \int_0^t E_s \left[\frac{1}{2^M} \sum_{k=1}^{C_M(s)} \mathbf{1}(W_{k,M}(s) > t-s)\right] ds. \end{aligned} \quad (8)$$

However, after some thought we see that under  $P_s$ , the joint distribution of  $C_M(s)$  and  $\tau_{k,M}$  is the same as the joint distribution of a typical batch size, and the busy period generated by a deterministic amount of work of size  $k/2^M$ . Thus, under  $E_s$  the events  $\{C_M(s) \geq k\}$  and  $\{\tau_{k,M} > t-s\}$  are independent, and so

$$E_s \left[\frac{1}{2^M} \sum_{k=1}^{C_M(s)} \mathbf{1}(\tau_{k,M} > t-s)\right] = E_s \left[\frac{1}{2^M} \sum_{k=1}^{C_M(s)} P(\tau_{k,M} > t-s)\right].$$

Recall that as  $M \rightarrow \infty$ ,  $\tau_{k(M,y),M}$  converges weakly to  $\tau_y$ , which implies that  $P(\tau_{k(M,y),M} > t-s)$  converges to  $P(\tau_y > t-s)$ , for all  $y \neq t-s$ , due to the fact that each random variable  $\tau_x$  has an atom at  $x$ . Thus, for each realization of  $C_M(s)$  under  $E_s$ , we can apply the dominated convergence theorem to conclude that

$$\frac{1}{2^M} \sum_{k=1}^{C_M(s)} P(\tau_{k,M} > t-s) \rightarrow \int_0^B P(\tau_y > t-s) dy \quad (9)$$

Combining (8) with (9) yields, as  $M \rightarrow \infty$ ,



$$\begin{aligned}
E \left[ \frac{Q_M(t)}{2^M} \right] &\rightarrow \int_0^x P(\tau_y > t) dy + \lambda \int_0^t E \left[ \int_0^B P(\tau_y > t-s) dy \right] ds \\
&= \int_0^x P(\tau_y > t) dy + \lambda \int_0^t E \left[ \int_0^\infty P(\tau_y > t-s) P(B > y) dy \right] ds \\
&= \int_0^x P(\tau_y > t) dy + \lambda E[B] \int_0^t P(\tau_{R_B} > t-s) ds \\
&= \int_0^x P(\tau_y > t) dy + \lambda E[B] E[\tau_{R_B}] P(R_{\tau_{R_B}} \leq t). \tag{10}
\end{aligned}$$

However, clearly

$$E[\tau_{R_B}] = \frac{E[R_B]}{1-\rho} = \frac{E[B^2]}{2E[B](1-\rho)}$$

and plugging this into (10) completes the proof. A standard phase-type approximation argument (see Theorem 4.2 of Asmussen [8]) can be used to derive the first moment for general service times.  $\diamond$

**Remark** By letting  $t \rightarrow \infty$ , we arrive at the first moment of the stationary workload:

$$\lim_{t \rightarrow \infty} E[W(t)|W(0) = x] = \frac{\lambda E[B^2]}{2(1-\rho)}.$$

♣

**Remark** It is also easy to see that if  $x = 0$ ,  $E[W(t)]$  is not only increasing, but also concave in  $t$ . This follows immediately from the expression, since the derivative is decreasing in  $t$ . These type of properties were proved in a much more general setting in [7]; however, for our specific Lévy process we are actually able to see the result immediately from our derived expression for the first moment. ♣

**Remark** We should also mention that an alternative expression for the first moment can also be found in Theorem 7.3 of [2], which holds for an arbitrary reflected Lévy process. It would be interesting to see a direct argument which shows how the two expressions are related to one another. ♣

We will now use this result to derive an explicit expression for the covariance function of the stationary version of the workload process  $\{W_s(t); t \in \mathbb{R}\}$ . Clearly, since the workload process is Markovian, we see that

$$\begin{aligned}
E[W_s(0)W_s(t)] &= \int_0^\infty x E[W(t)|W(0) = x] dP(W_s(0) \leq x) \\
&= \int_0^\infty x \int_0^x P(\tau_y > t) dy dP(W_s(0) \leq x) + \int_0^\infty x \frac{\lambda E[B^2]}{2(1-\rho)} P(R_{\tau_{R_B}} \leq t) dP(W_s(0) \leq x) \\
&= \int_0^\infty \int_y^\infty x P(\tau_y > t) dP(W_s(0) \leq x) dy + \left[ \frac{\lambda E[B^2]}{2(1-\rho)} \right]^2 P(R_{\tau_{R_B}} \leq t). \tag{11}
\end{aligned}$$

Hence, the covariance formula  $r(t) = E[W_s(0)W_s(t)] - E[W_s(0)]^2$  is just

$$r(t) = \int_0^\infty \int_y^\infty x P(\tau_y > t) dP(W_s(0) \leq x) dy + \left[ \frac{\lambda E[B^2]}{2(1-\rho)} \right]^2 P(R_{\tau_{R_B}} \leq t) - \left[ \frac{\lambda E[B^2]}{2(1-\rho)} \right]^2$$

However, we recall that, for each  $y$ ,

$$\int_y^\infty x dP(W_s(0) \leq x) = P(A_{W_s(0)} > y) E[W_s(0)]$$

where, for a given nonnegative random variable  $X$ ,  $A_X$  represents the total life distribution of  $X$ , i.e. for each  $y \geq 0$ ,

$$P(A_X \leq y) = \frac{\int_0^y x dP(X \leq x)}{E[X]}.$$

Therefore, we see from (11) that

$$\begin{aligned} r(t) &= E[W_s(0)] \int_0^\infty P(\tau_y > t) P(C_{W_s(0)} > y) dy - \left( \frac{\lambda E[B^2]}{2(1-\rho)} \right)^2 P(R_{\tau_{R_B}} > t) \\ &= E[W_s(0)^2] P(\tau_{C_{W_s(0)}} > t) - \left( \frac{\lambda E[B^2]}{2(1-\rho)} \right)^2 P(R_{\tau_{R_B}} > t). \end{aligned}$$

It is stated in Es-Saghouani and Mandjes [16] that the asymptotics (as  $t \rightarrow \infty$ ) of the covariance function of the  $M/G/1$  workload resemble those of the busy period, and our results show why this is the case. It is not, however, immediately obvious from the expression to draw the same conclusions regarding the behavior of  $r(t)$  (i.e. whether or not  $r$  is positive, decreasing or convex) from this particular form of  $r$ .

## 4 Higher Moments

Throughout this section, we will assume that  $W(0) = 0$ . The results can be generalized to any initial condition, by following a similar proof technique, and we leave the details to the interested reader.

**Theorem 4.1** *Suppose that  $W(0) = 0$ . Then, the  $n$ th moment of  $W(t)$  can be expressed in the following way:*

$$E[W(t)^n] = \sum_{m=1}^n \sum_{r_k \geq 1, r_1 + \dots + r_m = n} \left( \prod_{l=1}^m \binom{n-1 - \sum_{j=1}^{l-1} r_j}{r_l - 1} \right) \left( \prod_{l=1}^m \frac{\lambda E[B^{r_l+1}] l}{(r_l + 1)(1-\rho)} \right) P\left( \sum_{l=1}^m R_{\tau_{R_{r_l, B}}} \leq t \right)$$

Moreover, letting  $t \rightarrow \infty$  in  $E[W(t)^n]$  gives us the stationary moments, which we now state as a corollary.

**Corollary 4.1** *The moments of the stationary distribution of the workload process are as follows: for  $n \geq 1$ ,*

$$E[W(\infty)^n] = \sum_{m=1}^n \sum_{r_k \geq 1, r_1 + \dots + r_m = n} \left( \prod_{l=1}^m \binom{n-1 - \sum_{j=1}^{l-1} r_j}{r_l - 1} \right) \left( \prod_{l=1}^m \frac{\lambda E[B^{r_l+1}] l}{(r_l + 1)(1-\rho)} \right).$$

**Remark** We can also use these moments to derive the moments of the steady-state distribution of the number of customers waiting for service in an  $M/G/1$  queue. From the distributional Little's law (see e.g. Bertsimas and Nakazato [9]) we find that, if we assume the queue operates under the FCFS discipline, the stationary workload is equal in distribution the steady-state waiting time, and the steady-state queue-length and steady-state waiting times are related to one another in the following way:

$$E[z^{Q(\infty)}] = E[e^{-\lambda(1-z)W(\infty)}].$$

After differentiating both sides of the equality  $n$  times with respect to  $z$ , and letting  $z \rightarrow 1$ , we find that

$$E[Q(\infty)(Q(\infty) - 1) \cdots (Q(\infty) - n + 1)] = \lambda^n E[W(\infty)^n]$$

which proves the claim. ♣

**Proof** Our goal is to derive the  $n$ th moment of  $W(t)$ , for all  $t$ . We will again assume that the services are absolutely continuous, but as we saw for the first moment, the general case will also follow from a standard phase-type approximation argument. Clearly,

$$\begin{aligned} E \left[ \left( \frac{Q_M(t)}{2^M} \right)^n \right] &= \frac{1}{2^{Mn}} E \left[ \int_0^t \cdots \int_0^t \prod_{k=1}^n \sum_{l=1}^{C_M(s_k)} \mathbf{1}(W_l(s_k) > t - s_k) N(ds_n) N(ds_{n-1}) \cdots N(ds_1) \right] \\ &= \frac{1}{2^{Mn}} \int_0^t \cdots \int_0^t E_{s_1, s_2, \dots, s_n} \left[ \prod_{k=1}^n \sum_{l=1}^{C_M(s_k)} \mathbf{1}(W_l(s_k) > t - s_k) \right] \mu_{s_1, \dots, s_{n-1}}(ds_n) \cdots \mu_{s_1}(ds_2) \mu(ds_1) \end{aligned}$$

where for each Borel set  $A$ ,  $\mu(A) = \lambda|A|$  ( $|\cdot|$  representing Lebesgue measure), and  $\mu_{s_1, \dots, s_n}(A) = \lambda|A| + \sum_{k=1}^n \mathbf{1}(s_k \in A)$ . In other words,  $\mu_{s_1, \dots, s_n}$  represents a measure that is basically proportional to Lebesgue measure, with deterministic unit atoms at the points  $s_1, \dots, s_n$ .

To compute this integral, we will need to rewrite it in a way such that we are only integrating with respect to  $\mu$ , i.e. a constant multiple of Lebesgue measure. We will first give the following identity:

$$\begin{aligned} &\int_0^t \cdots \int_0^t E_{s_1, s_2, \dots, s_n} \left[ \prod_{k=1}^n \sum_{l=1}^{C_M(s_k)} \mathbf{1}(W_l(s_k) > t - s_k) \right] \mu_{s_1, \dots, s_{n-1}}(ds_n) \cdots \mu_{s_1}(ds_2) \mu(ds_1) \\ &= \sum_{m=1}^n \sum_{r_k \geq 1, r_1 + \dots + r_m = n} \lambda^m \left[ \prod_{l=1}^m \binom{n-1 - \sum_{j=1}^{l-1} r_j}{r_l - 1} \right] \\ &\times \int_{(0,t]} \cdots \int_{(0,t] - \{s_1, \dots, s_{m-1}\}} E_{s_1, \dots, s_m} \left[ \prod_{k=1}^m \left( \sum_{l=1}^{C_M(s_k)} \mathbf{1}(W_l(s_k) > t - s_k) \right)^{r_k} \right] ds_m \cdots ds_1. \quad (12) \end{aligned}$$

This can be seen by applying a counting argument on the indices.

To finish the calculation, we will need to compute each of the integrals that are of the form

$$\int_{(0,t]} \cdots \int_{(0,t] - \{s_1, \dots, s_{m-1}\}} E_{s_1, \dots, s_m} \left[ \prod_{k=1}^m \left( \sum_{l=1}^{C_M(s_k)} \mathbf{1}(W_l(s_k) > t - s_k) \right)^{r_k} \right] ds_m \cdots ds_1. \quad (13)$$

An important thing to notice is that this integral can be expressed as the sum of  $m!$  integrals which are of the form

$$\int_0^t \int_0^{s_1} \cdots \int_0^{s_{m-1}} E_{s_1, \dots, s_m} \left[ \prod_{k=1}^m \left( \sum_{l=1}^{C_M(s_k)} \mathbf{1}(W_l(s_k) > t - s_k) \right)^{r_k} \right] ds_m \cdots ds_1 \quad (14)$$

where  $m!$  refers to the number of permutations of  $s_1, \dots, s_m$ . The reader may think that this integral will differ between permutations, due to the presence of the  $r_i$  terms in the exponent, but we will see very soon that this is not the case. Based on how (14) is written, we see that  $s_m < s_{m-1} < \dots < s_1$ , and so from the independent increments property of a marked Poisson process, along with the fact that the customers are served in a preemptive-resume LCFS manner, we find that

$$\begin{aligned} & E_{s_1, \dots, s_m} \left[ \prod_{k=1}^m \left( \sum_{l=1}^{C_M(s_k)} \mathbf{1}(W_l(s_k) > t - s_k) \right)^{r_k} \right] \\ &= E_{s_1, \dots, s_m} \left[ \prod_{k=1}^m \left( \sum_{l=1}^{C_M(s_k)} \mathbf{1}(W_l(s_k) > s_{k-1} - s_k) \right)^{r_k} \right] \\ &= \prod_{k=1}^m E_{s_k} \left[ \left( \sum_{l=1}^{C_M(s_k)} \mathbf{1}(W_l(s_k) > s_{k-1} - s_k) \right)^{r_k} \right] \end{aligned}$$

where we are following the convention that  $s_0 = t$ . Thus, to compute the integral it suffices to compute each of the expectations

$$E_{s_k} \left[ \left( \sum_{l=1}^{C_M(s_k)} \mathbf{1}(W_l(s_k) > s_{k-1} - s_k) \right)^{r_k} \right].$$

However, after rearranging terms and performing some algebra we see that

$$\begin{aligned} & \left( \sum_{l=1}^{C_M(s_k)} \mathbf{1}(W_l(s_k) > s_{k-1} - s_k) \right)^{r_k} \\ &= \sum_{l=1}^{C_M(s_k)} ((C_M(s_k) - l + 1)^{r_k} - (C_M(s_k) - l)^{r_k}) \mathbf{1}(W_l(s_k) > s_{k-1} - s_k). \end{aligned} \quad (15)$$

Moreover, as was true in the previous section, on the set where  $C_M(s_k) \geq l$ ,  $W_l(s_k) \stackrel{d}{=} \tau_{l,M}$ , and so by independence under  $P_{s_k}$

$$\begin{aligned} & E_{s_k} \left[ \left( \sum_{l=1}^{C_M(s_k)} \mathbf{1}(W_l(s_k) > s_{k-1} - s_k) \right)^{r_k} \right] \\ &= E_{s_k} \left[ \sum_{l=1}^{C_M(s_k)} ((C_M(s_k) - l + 1)^{r_k} - (C_M(s_k) - l)^{r_k}) P(\tau_{l,M} > s_{k-1} - s_k) \right] \end{aligned}$$

and as  $M \rightarrow \infty$ , we see that after including our space-scaling factor, two applications of the dominated convergence theorem give

$$\begin{aligned} & \frac{1}{2^{Mr_k}} E \left[ \sum_{l=1}^{C_M(s_k)} ((C_M(s_k) - l + 1)^{r_k} - (C_M(s_k) - l)^{r_k}) P(\tau_{l,M} > s_{k-1} - s_k) \right] \\ & \rightarrow E \left[ \int_0^B P(\tau_y > s_{k-1} - s_k) r_k (B - y)^{r_k - 1} dy \right] \end{aligned}$$

$$\begin{aligned}
&= E \left[ \int_0^\infty P(\tau_y > s_{k-1} - s_k) r_k (B - y)^{r_k - 1} \mathbf{1}(B > y) dy \right] \\
&= \int_0^\infty P(\tau_y > s_{k-1} - s_k) r_k E[(B - y)^{r_k - 1} \mathbf{1}(B > y)] dy.
\end{aligned} \tag{16}$$

If we now apply Lemma 6.2 from the appendix to (16), we find that

$$\begin{aligned}
\int_0^\infty P(\tau_y > s_{k-1} - s_k) r_k E[(B - y)^{r_k - 1} \mathbf{1}(B > y)] dy &= \int_0^\infty P(\tau_y > s_{k-1} - s_k) r_k E[B^{r_k - 1}] P(R_{r_k - 1, B} > y) dy \\
&= E[B^{r_k - 1}] E[R_{r_k - 1, B}] r_k P(\tau_{R_{r_k, B}} > s_{k-1} - s_k) \\
&= E[B^{r_k}] P(\tau_{R_{r_k, B}} > s_{k-1} - s_k).
\end{aligned}$$

Thus, after plugging this into (14) and integrating,

$$\begin{aligned}
&\int_0^t \int_0^{s_1} \cdots \int_0^{s_{m-1}} E_{s_1, \dots, s_m} \left[ \prod_{k=1}^m \left( \sum_{l=1}^{C_M(s_k)} \mathbf{1}(W_l(s_k) > t - s_k) \right)^{r_k} \right] ds_m \cdots ds_1 \\
&= \int_0^t \int_0^{s_1} \cdots \int_0^{s_{m-1}} \prod_{k=1}^m E[B^{r_k}] P(\tau_{R_{r_k, B}} > s_{k-1} - s_k) ds_m \cdots ds_1 \\
&= \left( \prod_{k=1}^m E[B^{r_k}] \right) \left( \prod_{k=1}^m E[\tau_{R_{r_k, B}}] \right) P \left( \sum_{k=1}^m R_{\tau_{R_{r_k, B}}} \leq t \right) \\
&= \left( \prod_{k=1}^m E[B^{r_k}] \right) \left( \prod_{k=1}^m \frac{E[R_{r_k, B}]}{1 - \rho} \right) P \left( \sum_{k=1}^m R_{\tau_{R_{r_k, B}}} \leq t \right) \\
&= \left( \prod_{k=1}^m \frac{E[B^{r_k + 1}]}{(r_k + 1)(1 - \rho)} \right) P \left( \sum_{k=1}^m R_{\tau_{R_{r_k, B}}} \leq t \right).
\end{aligned}$$

This calculation proves that each of the  $m!$  integrals we alluded to above whose sum gives (13) are equal to one another. Finally, plugging this expression into (12) completes the proof.  $\diamond$

## 5 The Moments of the Busy Period

It is also worth noting at this point that all of the moments of the cdfs found in the moment expressions of  $W(t)$  can be computed as well, once we know the moments of the  $M/G/1$  busy period, and these can be computed in a recursive manner; see [4]. To do this, it suffices to derive a formula for the moments of  $\tau_x$ .

Clearly  $\tau_x$  can be represented in the following way:

$$\tau_x = x + \int_0^x \tau(s) N(ds)$$

where  $\tau(s)$  is the busy period induced by the last arrival at or before time  $s$ . Notice that “time” (on  $[0, x]$ ) has a different interpretation in this context; however, the idea behind this equation is exactly the same as the classical branching-process interpretation used when computing the busy period of the  $M/G/1$  queue (see Feller [17] for details). In particular, under the measure  $P_{s_1, \dots, s_n}$ ,  $\tau(s_1), \dots, \tau(s_n)$  are independent, and are equal in distribution to a typical busy period.

Also, for each integer  $n \geq 1$ ,

$$E[\tau_x^n] = \sum_{k=0}^n \binom{n}{k} x^{n-k} E \left[ \left( \int_0^x \tau(s) N(ds) \right)^k \right]. \quad (17)$$

By applying the Campbell-Mecke formula  $k$  times, we can mimic the argument used in the calculation of the moments of the workload to find that

$$\begin{aligned} & E \left[ \left( \int_0^x \tau(s) N(ds) \right)^k \right] \\ &= \int_0^x \cdots \int_0^x E_{s_1, s_2, \dots, s_k} \left[ \prod_{l=1}^k B(s_l) \right] \mu_{s_1, \dots, s_{k-1}}(ds_k) \cdots \mu(ds_1) \\ &= \sum_{j=1}^k \sum_{r_i \geq 1, r_1 + \dots + r_j = k} \lambda^j \left( \prod_{l=1}^j \binom{k-1 - \sum_{m=1}^{l-1} r_m}{r_l - 1} \right) x^j \prod_{l=1}^j E[\tau^{r_l}]. \end{aligned} \quad (18)$$

By plugging (18) into (17), we see that we now have an expression for all of the moments of  $\tau_x$ , purely in terms of the moments of the busy period. Expressions for the first four moments of  $\tau_x$  can also be found in Theorem 7 of [6].

## 6 Appendix

We will now state an expression for the moments of the residuals, which is a well-known result, but for the reader's convenience we will also provide a proof.

**Lemma 6.1** *For integers  $n \geq 0$ ,  $k \geq 0$ , we have that*

$$E[R_{k,B}^n] = \frac{E[B^{n+k}]}{\binom{n+k}{n} E[B^k]}.$$

**Proof** We will prove this statement by induction. Suppose that  $k = 0$ . Then

$$E[B^n] = E[R_{0,B}^n] = \frac{E[B^{n+0}]}{\binom{n+0}{n} E[B^0]} = E[B^n].$$

Now assume that it holds for all  $j \leq k$ . Then

$$\begin{aligned} E[R_{k+1,B}^n] &= \frac{E[R_{k,B}^{n+1}]}{(n+1)E[R_{k,B}]} = \frac{E[B^{n+k+1}]}{\binom{n+k+1}{n+1} E[B^k] (n+1) \frac{E[B^{k+1}]}{\binom{k+1}{k} E[B^k]}} \\ &= \frac{E[B^{n+k+1}]}{\frac{(n+k+1)!}{(n+1)! k!} \frac{n+1}{k+1} E[B^{k+1}]} \\ &= \frac{E[B^{n+k+1}]}{\frac{(n+k+1)!}{n!(k+1)!} E[B^{k+1}]} \end{aligned}$$

which completes the proof.  $\diamond$

The following result gives an alternative representation for the tail of the  $n$ th residual of a random variable.

**Lemma 6.2** For each integer  $n \geq 0$ , and each nonnegative real number  $x$ ,

$$E[(B - x)^n \mathbf{1}(B > x)] = E[B^n]P(R_{n,B} > x).$$

**Proof** The statement is trivial if  $n = 0$ . Suppose now that the statement holds for all  $k \leq n$ . Then

$$\begin{aligned} E[(B - x)^{n+1} \mathbf{1}(B > x)] &= \int_x^\infty (y - x)^{n+1} dP(B \leq y) \\ &= \int_x^\infty \int_x^y (y - x)^n du dP(B \leq y) \\ &= \int_x^\infty \int_u^\infty (y - x)^n dP(B \leq y) du \\ &= \int_x^\infty \int_u^\infty \sum_{k=0}^n \binom{n}{k} (y - u)^k (u - x)^{n-k} dP(B \leq y) du \\ &= \int_x^\infty \sum_{k=0}^n \binom{n}{k} E[B^k] P(R_{k,B} > u) (u - x)^{n-k} du. \end{aligned} \quad (19)$$

However, if we apply Lemma 6.1, we find that

$$\begin{aligned} \int_x^\infty (u - x)^{n-k} P(R_{k,B} > u) du &= \int_x^\infty \int_x^y (u - x)^{n-k} du dP(R_{k,B} \leq y) \\ &= \int_x^\infty \frac{(y - x)^{n-k+1}}{n - k + 1} dP(R_{k,B} \leq y) \\ &= \frac{E[R_{k,B}^{n-k+1}] P(R_{n+1,B} > x)}{n - k + 1} \\ &= \frac{E[B^{n+1}] P(R_{n+1,B} > x)}{\binom{n+1}{n-k+1} (n - k + 1) E[B^k]}. \end{aligned} \quad (20)$$

Plugging (20) into (19) and simplifying gives

$$\begin{aligned} E[(B - x)^{n+1} \mathbf{1}(B > x)] &= E[B^{n+1}] P(R_{n+1,B} > x) \sum_{k=0}^n \frac{1}{n+1} \\ &= E[B^{n+1}] P(R_{n+1,B} > x) \end{aligned}$$

which completes the proof.  $\diamond$

**Acknowledgments** This research began while the author was a Postdoctoral Fellow at EURANDOM, Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands. He would like to thank Onno Boxma for providing comments on a previous draft, and for showing him [14]. He would also like to thank Ward Whitt for his comments on a previous work [19], which initiated this study.

## References

- [1] J. Abate and W. Whitt (1987). Transient behavior of regulated Brownian motion, I: starting at the origin. *Advances in Applied Probability* **19**, 560-598.
- [2] J. Abate and W. Whitt (1987). Transient behavior of regulated Brownian motion, II: non-zero initial conditions. *Advances in Applied Probability* **19**, 599-631.

- [3] J. Abate and W. Whitt (1987). Transient behavior of the  $M/M/1$  queue: starting at the origin. *Queueing Systems* **2**, 41-65.
- [4] J. Abate and W. Whitt (1988). Transient behavior of the  $M/M/1$  queue via Laplace transforms. *Advances in Applied Probability* **20**, 145-178.
- [5] J. Abate and W. Whitt (1988). The correlation functions of  $M/M/1$  and RBM. *Stochastic Models* **4**, 315-359.
- [6] J. Abate and W. Whitt (1994). Transient behavior of the  $M/G/1$  workload process. *Operations Research* **42**, 750-764.
- [7] L. N. Andersen and M. Mandjes (2009). Structural properties of reflected Lévy processes. *Queueing Systems*, to appear.
- [8] S. Asmussen (2003). *Applied Probability and Queues*. Springer-Verlag, New York.
- [9] D. Bertsimas and D. Nakazato (1995). The distributional Little's law and its applications. *Operations Research* **43**, 298-310.
- [10] B. Blaszczyszyn (1995). Factorial moment expansion for stochastic systems. *Stochastic Processes and their Applications* **56**, 321-335.
- [11] B. Blaszczyszyn, T. Rolski and V. Schmidt (1995). Light-traffic approximation in queues and related stochastic models. *Advances in Queueing: Theory, Methods, and Open Problems*, CRC Press, 379-406
- [12] H. Chen (1999). Basic adjoint relation for transient and stationary analysis of some Markov processes. *Annals of Operations Research* **87**, 273-303.
- [13] H. Chen and D.D. Yao (2001). *Fundamentals of Queueing Networks*. Springer, New York.
- [14] R. B. Cooper and S-C Niu. (1986). Beneš's formula for  $M/G/1$ -FIFO 'explained' by preemptive-resume LIFO. *Journal of Applied Probability* **23**, 550-554.
- [15] D. R. Cox and V. Isham. (1986). The virtual waiting-time and related processes. *Advances in Applied Probability* **18**, 558-573.
- [16] A. Es-Saghouani and M. Mandjes (2008). On the correlation structure of Levy-driven queues. *Journal of Applied Probability* **45**, 940-952.
- [17] W. Feller (1971). *An Introduction to Probability Theory and its Applications, Volume 2*. John Wiley and Sons, New York.
- [18] B. H. Fralix (2009). A time-dependent view of the ASTA problem, with applications to preemptive queues and birth-death processes. Submitted.
- [19] B. H. Fralix and G. Riaño (2008). Another look at transient versions of Little's law, and  $M/G/1$ -preemptive LCFS queues. Submitted: a version of this paper also appears as EURANDOM report 2008-42.
- [20] O. Kallenberg (1983). *Random Measures*. Akademie-Verlag, Berlin.
- [21] L. Kleinrock (1975). *Queueing Systems, Volume 1: Theory*. John Wiley and Sons, New York.
- [22] T. J. Ott (1977). The covariance function of the virtual waiting-time process in an  $M/G/1$  queue. *Advances in Applied Probability* **9**, 158-168.
- [23] N. U. Prabhu (1999). *Stochastic Storage Processes*. Springer, New York.
- [24] T. Rolski (1989). Relationships between characteristics in periodic Poisson queues. *Queueing Systems* **4**, 17-26.



- [25] L. Takács (1962). A single-server queue with Poisson input. *Operations Research* **10**, 388-394.
- [26] W. Whitt (2002). *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer, New York.