An M/M/1 queue in a semi-Markovian environment

Philippe Nain
INRIA
B.P. 93
06902 Sophia Antipolis
France
nain@sophia.inria.fr

Rudesindo Núñez-Queija*
CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands
sindo@cwi.nl

ABSTRACT

We consider an M/M/1 queue in a semi-Markovian environment. The environment is modeled by a two-state semi-Markov process with arbitrary sojourn time distributions $F_0(x)$ and $F_1(x)$. When in state $i = 0, 1$, customers are generated according to a Poisson process with intensity λ_i and customers are served according to an exponential distribution with rate μ_i. Using the theory of Riemann-Hilbert boundary value problems we compute the z-transform of the queue-length distribution when either $F_0(x)$ or $F_1(x)$ has a rational Laplace-Stieltjes transform and the other may be a general — possibly heavy-tailed — distribution. The arrival process can be used to model bursty traffic and/or traffic exhibiting long-range dependence, a situation which is commonly encountered in networking. The closed-form results lend themselves for numerical evaluation of performance measures, in particular the mean queue-length.

Keywords

Queueing; Stochastic modeling; Communication networks; Heavy-tailed distribution; Riemann-Hilbert boundary value problem; Long-range dependence; Bursty traffic.

Categories and Subject Descriptors

C4 [**Performance of Systems**]: Modeling techniques; G.3 [**Mathematics of Computing**]: Probability and Statistics—*Queueing Theory*; I6 [**Simulation and modeling**]: General

1. INTRODUCTION

We consider the M/M/1 queue in which the arrival and service rates depend on the state of an underlying alternating renewal process. High-traffic periods which are distributed according to a distribution function $F_1(x)$, alternate with low-traffic periods having a distribution function $F_0(x)$ (we assume $F_0(0+) = F_1(0+) = 0$). During high-traffic (resp., low-traffic) periods, we say that the underlying process is in state 1 (resp., 0), customers arrive according to a Poisson process with rate λ_1 (resp., λ_0) and are served at rate μ_1 (resp., μ_0). Let $\rho_i := \lambda_i/\mu_i$ and assume that $0 \leq \rho_0 \leq \rho_1 < \infty$ (hence, the terminology high-traffic and low-traffic periods). The case $\mu_1 = 0$, i.e., $\rho_1 = \infty$, can be analyzed similarly, but for conciseness of presentation we assume $\mu_1 > 0$. Also, the condition $\rho_0 \leq \rho_1$ is not essential to the analysis and it will be convenient to allow $\rho_0 > \rho_1$, in which case the low-traffic periods are indexed by 1 and the high-traffic ones by 0. When $F_1(x)$ and $F_0(x)$ are phase-type distribution functions, the arrival process is a MMPP (Markov Modulated Poisson Process), see [14]. Related models under different assumptions were studied in [23] and [31]. An asymptotic analysis of the present model was given in [5]. The aim of this paper is to compute the z-transform of the stationary queue-length and, in particular, moments of the queue-length distribution.

This work is motivated by the need to evaluate the performance of queueing models fed by bursty processes. As it is well-established by now, the burstiness of traffic in today's networks in many cases rules out the use of Poisson traffic models [10, 19, 29] and triggers the need for new models. Several models have already been proposed in the literature, including the fractional Brownian motion [24, 25], on/off sources with heavy-tailed distributions for the on and/or off periods [1, 2, 3, 4, 11, 18, 34] and the $M/G/\infty$ input process [17, 20, 27, 28, 32] (these lists of references are not exhaustive as the activity in this domain is very dense). More generally, studies of queues in presence of heavy-tailed distributions, initiated with the works of Cohen [7], Pakes [26] and Veraverbeke [33], can be found in a recent special issue of QUESTA devoted to this topic [30].

The traffic process in our model exhibits burstiness when ρ_0 and ρ_1 are significantly different and the changes in the semi-Markovian environment occur on a comparable or larger time-scale than the arrival and departure processes. In particular, if $\rho_1 > 1$ then the traffic intensity will exceed the server capacity (i.e., the queue is temporarily unstable) during high traffic periods, a situation which is likely to create congestion, particularly if the duration of high traffic pe-

*Also affiliated with Eindhoven University of Technology. Part of this work was carried out while visiting INRIA during Spring 2000, funded by the Netherlands Organization for Scientific Research (NWO) under grant R 61-451.

ACM SIGMETRICS 2001 / Performance 2001 Cambridge, MA, June 17-20, 2001

riods has a heavy-tailed distribution. In the extreme case when $\rho_1 > 1$ and the variance of the high traffic periods is infinite, the traffic process is long-range dependent.

In the above cited references the emphasis was on asymptotic analysis assuming heavy-tailed input processes. One merit of this paper is that an explicit expression for the generating function of the stationary queue-length distribution is derived in the case that the duration of the high-traffic periods has a general (possibly heavy-tailed) distribution and the distribution of the low periods has a rational Laplace-Stieltjes transform (this includes phase-type distributions). We do so by first establishing two functional equations (11) in Section 2 to be satisfied by two joint Laplace-Stieltjes and z-transforms related to the queue-length distribution. Solving for these functional equations is then the objective of Sections 3, 4 and 5. In a first step, we show (Lemma 4.1) that the unknown function $L_i(0,s)$ appearing in (11) is rational if $\phi_i(s)$ — the Laplace-Stieltjes transform of $F_i(x)$ — is itself rational. This result allows us to formulate and solve a Riemann-Hilbert boundary value problem on the unit circle in Section 5. The use of boundary value problems to solve queueing problems is not new and can be traced back to the seminal work by Fayolle and Iasnogorodski [12] (see also the monograph [8] and the more recent [13], as well as the references therein). Solving the boundary value problem allows us to compute the z-transform of the queue-length, as shown in Section 6.

A word on the notation used in this paper: $\mathbf{C}$ will denote the set of all complex numbers, $\mathrm{Re}\,(z)$ (resp. $\mathrm{Im}\,(z)$, $|z|$) the real part (resp. imaginary part and modulus) of any complex number z. The closure of any set $\mathcal{A}$ will be denoted by $\overline{\mathcal{A}}$. $\mathbf{1}(E)$ will denote the indicator function with $\mathbf{1}(E) = 1$ if condition E is satisfied and $\mathbf{1}(E) = 0$ otherwise. For $a > 0$, let $C_a = \{z \in \mathbf{C} : |z| = a\}$ be the circle centered at $z = 0$ with radius a and let $C_a^+ = \{z \in \mathbf{C} : |z| < a\}$ and $C_a^- = \{z \in \mathbf{C} : |z| > a\}$ be the domain inside and the domain outside the circle C_a, respectively.

2. THE FUNCTIONAL EQUATIONS

We now return to the queueing system at hand. We assume that the queue is stable, i.e.,

$$\lambda_0 \frac{a_0}{a_0 + a_1} + \lambda_1 \frac{a_1}{a_0 + a_1} < \mu_0 \frac{a_0}{a_0 + a_1} + \mu_1 \frac{a_1}{a_0 + a_1}, \qquad (1)$$

where $a_0 \in (0, \infty)$ and $a_1 \in (0, \infty)$ are the means of the low-traffic and high-traffic periods, respectively.

Assume that the system is in equilibrium and denote by $P_i(n, x)$ the probability that the underlying process is in state $i \in \{0, 1\}$, there being $n \in \{0, 1, 2, \dots\}$ customers in the system and less than $x \in [0, \infty)$ time units remaining until the next switch of the underlying process to state $i' := 1 - i$, $i \in \{0, 1\}$ (we shall use this notation throughout the paper). Obviously,

$$\sum_{n=0}^{\infty} P_i(n, x) = \frac{a_i}{a_0 + a_1} \int_{y=0}^{x} \frac{1 - F_i(y)}{a_i} \, \mathrm{d}y. \qquad (2)$$

It can be shown that $P_i(n, x)$, $i \in \{0, 1\}$, is concave in $x \geq 0$ and, hence, $DP_i(n) := \lim_{x \downarrow 0} \frac{P_i(n,x)}{x}$ exists, is strictly positive and, because of (2), it is finite. From the dynamics of the system it follows that $P_i(n, x)$, satisfies

$$\begin{aligned} -\frac{\mathrm{d}}{\mathrm{d}x} P_i(n, x) &= -(\lambda_i + \mu_i \mathbf{1}(n \geq 1)) P_i(n, x) \\ &\quad + \lambda_i P_i(n-1, x) + \mu_i P_i(n+1, x) \\ &\quad - DP_i(n) + DP_{i'}(n) F_i(x), \end{aligned} \qquad (3)$$

where $P_i(-1, x) := 0$. From (3) it follows that without loss of generality we can restrict ourselves to the case where the service rate always equals 1 (independent of the state of the underlying renewal process) and the arrival rates are ρ_0 and ρ_1, respectively. Informally, we can 'speed up' time by a factor $1/\mu_i$ if the underlying alternating process is in state i. Note that, after the change of time, the sojourn time of the underlying process in state i is distributed according to the distribution function $\hat{F}_i(x) := F_i(x/\mu_i)$ with mean $\hat{a}_i := \mu_i a_i$. Formally, we note that $P_i(n, x)$ is a solution to (3) if and only if, for any constant c, $\hat{P}_i(n, x) := c\mu_i P_i(n, x/\mu_i)$ is a solution to

$$\begin{aligned} -\frac{\mathrm{d}}{\mathrm{d}x} \hat{P}_i(n, x) &= -(\rho_i + \mathbf{1}(n \geq 1)) \hat{P}_i(n, x) \\ &\quad + \rho_i \hat{P}_i(n-1, x) + \hat{P}_i(n+1, x) \\ &\quad - D\hat{P}_i(n) + D\hat{P}_{i'}(n) \hat{F}_i(x). \end{aligned} \qquad (4)$$

If $P_i(n, x)$ satisfies (2) and $\hat{P}_i(n, x)$ is a distribution function in the sense that

$$\sum_{n=0}^{\infty} \left(\hat{P}_0(n, \infty) + \hat{P}_1(n, \infty) \right) = 1,$$

then the constant c is given by

$$c = \frac{a_0 + a_1}{\hat{a}_0 + \hat{a}_1},$$

and, consequently,

$$\sum_{n=0}^{\infty} \hat{P}_i(n, x) = \int_{y=0}^{x} \frac{1 - \hat{F}_i(y)}{\hat{a}_0 + \hat{a}_1} \mathrm{d}y. \qquad (5)$$

In the remainder of the paper we shall work with (4) and (5) instead of (2) and (3). The ergodicity condition (1) can be written as

$$\hat{a}_0 (\rho_0 - 1) + \hat{a}_1 (\rho_1 - 1) < 0. \qquad (6)$$

For $i \in \{0, 1\}$, $|z| \leq 1$ and $x \geq 0$ let $G_i(z, x)$ and $D_i(z)$ be the generating functions of $\hat{P}_i(n, x)$ and $D\hat{P}_i(n)$, respectively,

$$G_i(z, x) := \sum_{n=0}^{\infty} z^n \hat{P}_i(n, x), \qquad (7)$$

$$D_i(z) := \sum_{n=0}^{\infty} z^n D\hat{P}_i(n). \qquad (8)$$

Clearly, $G_i(1, x)$ is given by (5) and

$$D_i(1) = \frac{1 - \hat{F}_i(0+)}{\hat{a}_0 + \hat{a}_1} = \frac{1}{\hat{a}_0 + \hat{a}_1}. \qquad (9)$$

The generating function $D_i(z)$, $i = 0, 1$, is analytic in $|z| < 1$ and continuous in $|z| \leq 1$. This is because $D_i(1) < \infty$ and $D\hat{P}_i(n) \geq 0$ for all n (see below (2)), and therefore

$$|D_i(z)| \leq \sum_n |z^n| D\hat{P}_i(n) \leq \sum_n D\hat{P}_i(n) < \infty, \qquad |z| \leq 1.$$

Next, we define the LST (Laplace-Stieltjes Transform) of $G_i(z,x)$

$$L_i(z,s) := \int_{x=0}^{\infty} \mathrm{e}^{-sx} \mathrm{d}G_i(z,x), \qquad (10)$$

for $i \in \{0,1\}$, $|z| \le 1$, $\mathrm{Re}\,(s) \ge 0$,

Let us return to (4): Multiplying both sides of that equation by $z^n e^{-sx}$, summing over all $n \ge 0$ and then integrating over all $x \ge 0$ yields after routine manipulations

$$\begin{aligned}\left(\rho_i z^2 - (1+\rho_i - s)z + 1\right) L_i(z,s) &= (1-z)L_i(0,s) \\ &+ z\left(D_i(z) - \phi_i(s) D_{i'}(z)\right), \quad i = 0,1, \qquad (11)\end{aligned}$$

where $\phi_i(s) := \int_{x=0}^{\infty} \mathrm{e}^{-sx} \mathrm{d}\hat{F}_i(x)$ is the LST of $\hat{F}_i(x)$.

Also note from definitions (7) and (10) that the normalizing condition

$$L_0(1,0) + L_1(1,0) = 1 \qquad (12)$$

holds.

REMARK 2.1. *Above we derived (5) and (9) from physical considerations of the model. These properties also follow from (11) and (12). If we set $z = 1$ in (11) and let $s \to 0$ we get $D_0(1) = D_1(1)$. Substituting this into (11) and using (12) gives (9). Moreover, substitution into (11), for $z = 1$, gives*

$$L_i(1,s) = \frac{1-\phi_i(s)}{(\hat{a}_0 + \hat{a}_1)s},$$

which is in agreement with (5). In particular,

$$L_i(1,0) = \frac{\hat{a}_i}{\hat{a}_0 + \hat{a}_1}, \quad i = 0,1. \qquad (13)$$

$\diamond$

The first factor on the left-hand side of (11), called the *kernel* of $L_i(z,s)$, will play a central role in our analysis. From its very definition it is seen that $L_i(z,s)$ is analytic for $|z| < 1$, $\mathrm{Re}\,(s) > 0$ and continuous for $|z| \le 1$, $\mathrm{Re}\,(s) \ge 0$. Hence, if the kernel of $L_i(z,s)$ vanishes for some $|z| \le 1$ and $\mathrm{Re}\,(s) \ge 0$, then the right-hand side of (11) must also vanish. This key observation will allow us in the forthcoming sections to identify $L_i(0,s)$ and $D_i(z)$ and, subsequently, to determine $L_i(z,s)$ for $i = 0,1$.

Once the functions $L_0(z,s)$ and $L_1(z,s)$ have been determined for $|z| \le 1$ and $\mathrm{Re}\,(s) \ge 0$, we can find the z-transform of the stationary queue-length for the original model governed by the Kolmogorov equations (3). If we denote that z-transform by $N(z)$, we have, cf. (7),(10),

$$\begin{aligned} N(z) &= \sum_{n=0}^{\infty} z^n \lim_{x\to\infty} \left(P_0(n,x) + P_1(n,x)\right) \\ &= \frac{1}{c\mu_0} L_0(z,0) + \frac{1}{c\mu_1} L_1(z,0), \quad |z| \le 1. \qquad (14)\end{aligned}$$

In particular, the average queue-length is given by $\overline{N} := N^{(1)}(1)$, where $N^{(1)}(z)$ denotes the first order derivative of $N(z)$. The ultimate objective in this paper is to find closed-form expressions for $N(z)$ (see Section 6). We shall do so by first expressing $L_i(z,0)$ in terms of $L_i(0,s)$ and $D_i(z)$ (Corollary 3.1) which can then be plugged into (14). Deriving $L_i(0,s)$ and $D_i(z)$ is the subject of Sections 4 and 5.

In the subsequent analysis it turns out that, even when the LST $\phi_i(s)$, $i = 0,1$, are rational functions, an algebraic proof of the uniqueness of $L_0(z,s)$ and $L_1(z,s)$ satisfying (11) and (12) is quite involved. Following [9], we circumvent this technicality using the following theorem:

THEOREM 2.1. *If (6) is satisfied then there are unique functions $L_0(z,s)$ and $L_1(z,s)$ which are analytic for $|z| < 1$ and $\mathrm{Re}\,(s) > 0$, continuous for $|z| \le 1$ and $\mathrm{Re}\,(s) \ge 0$, and satisfy (11) and (12).* $\diamond$

PROOF. The triple queue-length, state of the alternating environment and remaining time until the next switch in the environment is a Markov process with irreducible state space $\{0,1,2,\dots\} \times \{0,1\} \times [0,\infty)$. Under (6) this process is positive recurrent and, hence, up to multiplication by a constant, there is exactly one bounded solution to the Kolmogorov equations (4), see also [9]. $\square$

3. ANALYSIS OF THE KERNEL

For $i \in \{0,1\}$ we denote the kernel of $L_i(z,s)$ by

$$K_i(z,s) := \rho_i z^2 - (1+\rho_i - s)z + 1, \quad z, s \in \mathbf{C}. \qquad (15)$$

Solving $K_i(z,s) = 0$ for z, we find

$$z_i(s) = \frac{1+\rho_i - s \pm \sqrt{(1+\rho_i - s)^2 - 4\rho_i}}{2\rho_i}.$$

The algebraic function $z_i(s)$ has 2 branch points s_i^1 and s_i^2 (the zeros of $(1+\rho_i - s)^2 - 4\rho_i$), given by

$$s_i^1 = (1-\sqrt{\rho_i})^2 \quad \text{and} \quad s_i^2 = (1+\sqrt{\rho_i})^2. \qquad (16)$$

LEMMA 3.1. *We assume that $\rho_i > 0$. The following statements hold*

(1) The equation $K_i(z,s) = 0$ has two roots (in z): one, denoted as $z_i^+(s)$, is an analytic function of s in $\mathbf{C} - [s_i^1, s_i^2]$; the second one, $z_i^-(s)$, is given by

$$z_i^-(s) = 1/(\rho_i z_i^+(s)) \qquad (17)$$

(2) for $s \in [s_i^1, s_i^2]$, $z_i^+(s)$ and $z_i^-(s)$ are each others complex conjugates with common modulus $1/\sqrt{\rho_i}$;

(3) $z_i^+(s) \in C^+_{1/\sqrt{\rho_i}}$ and $z_i^-(s) \in C^-_{1/\sqrt{\rho_i}}$ for all $s \in \mathbf{C} - [s_i^1, s_i^2]$;

(4) when s moves along the "contour" $[s_i^1, s_i^2]$, denoted as $[s_i^1, s_i^2] \underset{\leftarrow}{\overset{s}{\rightarrow}}$ (i.e., s goes from s_i^1 to s_i^2 and returns to s_i^1) then both $z_i^+(s)$ and $z_i^-(s)$ describe the circle $C_{1/\sqrt{\rho_i}}$.

$\diamond$

PROOF. The first part of statement (1) results from the general theory of polynomials of two complex variables [15];

(17) follows from $K_i(z,s) = \rho_i z^2 K_i(1/(\rho_i z), s)$ for all $s \in \mathbf{C}$, $z \in \mathbf{C}$.

Define $s_\theta := 1 + \rho_i - 2\sqrt{\rho_i}\cos(\theta)$ for $0 \le \theta \le 2\pi$. Observe that s_θ describes the contour $[s_i^1, s_i^2]_{\overset{s}{\rightleftarrows}}$ as θ increases from 0 to 2π. Substituting s_θ for s in the equation $K_i(z,s) = 0$ yields

$$z_i(s_\theta) = \begin{cases} \pm \dfrac{\mathrm{e}^{i\theta}}{\sqrt{\rho_i}}, & \text{if } 0 \le \theta \le \pi; \\[2ex] \pm \dfrac{\mathrm{e}^{-i\theta}}{\sqrt{\rho_i}}, & \text{if } \pi \le \theta \le 2\pi, \end{cases}$$

which proves both statements (2) and (4)

The mapping $s \to z_i^+(s)$ being analytic in $\mathbf{C} - [s_i^1, s_i^2]$, we know from the maximum modulus principle [15, pp. 201-203] that the maximum modulus of $z_i^+(s)$ cannot be reached inside the domain $\mathbf{C} - [s_i^1, s_i^2]$. Since $z_i^+(\infty) = 0$ we conclude from the above that the maximum modulus is reached on the segment $[s_i^1, s_i^2]$. Since we have already shown that $|z_i^+(s)| = 1/\sqrt{\rho_i}$ for $s \in [s_i^1, s_i^2]$, we see that necessarily $|z_i^+(s)| \in C^+_{1/\sqrt{\rho_i}}$, thereby implying that $|z_i^-(s)| \in C^-_{1/\sqrt{\rho_i}}$ because of (17). $\square$

Remark 3.1. *In Lemma 3.1 we assumed that $\rho_i > 0$. When $\rho_i = 0$ the kernel $K_i(z,s)$ has a unique root with respect to z given by $z(s) := 1/(1-s)$ for all $s \in \mathbf{C} - \{1\}$. In the sequel, whenever $\rho_i = 0$ we may read $z_i(s)$ instead of $z_i^+(s)$ (and, in that case, $z_i^-(s)$ is not defined).* $\diamond$

Remark 3.2. *If $\rho_i < 1$ then, for $\mathrm{Re}\,(\omega) > -s_i^1$, $\pi(\omega) := z_i^+(-\omega)$ is the LST of the busy period of the M/M/1 queue with arrival rate ρ_i and service rate 1.* $\diamond$

It will also be convenient to study the solution of $K_i(z,s) = 0$ in the variable s. This is done in the following lemma whose proof follows directly from the definition of $K_i(z,s)$.

Lemma 3.2. *For $z \in \mathbf{C} - \{0\}$,*

$$s_i(z) := \rho_i(1-z) - (1/z - 1),$$

is the unique root of $K_i(z,s)$ in the variable s. We have $s_i(z_i^+(s)) = s$ for all $s \in \mathbf{C}$, and $s_i(z_i^-(s)) = s$ for all s wherever $z_i^-(s)$ is defined (i.e., everywhere except when $s = \infty$).

For $r > 0$, $r \neq 1/\sqrt{\rho_i}$, $s_i(z)$ maps C_r onto an ellipse centered at $1+\rho_i$, symmetric with respect to the real axis and the line $\mathrm{Re}\,(s) = 1+\rho_i$, with extremal points $1+\rho_i \pm r\left(\rho_i + \frac{1}{r^2}\right)$ and $1+\rho_i \pm \mathrm{i}\, r\left(\rho_i - \frac{1}{r^2}\right)$. The point $z = r$ (the right extremal point on the circle) corresponds to the left extremal point of the ellipse, $1 + \rho_i - r\left(\rho_i + \frac{1}{r^2}\right)$. As z traverses C_r in the positive direction (i.e., counter clock wise), $s_i(z)$ traverses the ellipse in the positive direction if $r > 1/\sqrt{\rho_i}$ and in the negative direction if $r < 1/\sqrt{\rho_i}$.

Moreover, $s_i(z)$ maps both $C^+_{1/\sqrt{\rho_i}} - \{0\}$ and $C^-_{1/\sqrt{\rho_i}}$ onto $\mathbf{C} - [s_i^1, s_i^2]$ and it maps onto the cut $[s_i^1, s_i^2]$. $\diamond$

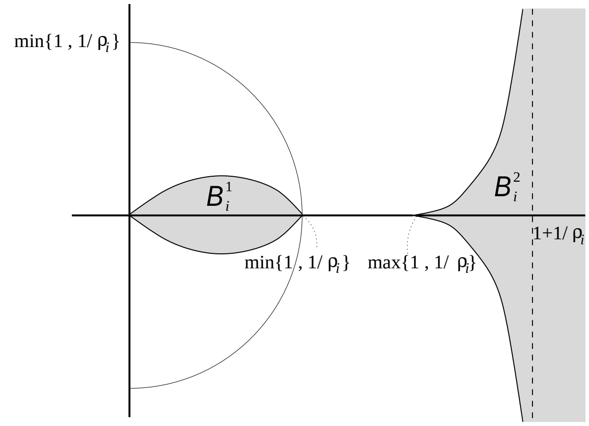

Figure 1: The non-shaded areas form the set $\mathcal{A}_i$

We shall denote the set in the z-plane for which $\mathrm{Re}\,(s_i(z)) > 0$ by (we write $z = x + \mathrm{i}y$)

$$\mathcal{A}_i \quad := \quad \left\{x + \mathrm{i}y : y^2\,(1 + \rho_i(1-x)) > x(1-x)(1-\rho_i x)\right\}.$$

Typically $\mathcal{A}_i$ looks like the non-shaded parts in Figure 1. It will be convenient to define the following subsets of $\mathbf{C} - \mathcal{A}_i$

$$\mathcal{B}_i^1 \quad := \quad \left\{0 < x < \min\left\{1, \frac{1}{\rho_i}\right\} y^2 < \frac{x(1-x)(1-\rho_i x)}{1+\rho_i(1-x)}\right\},$$

$$\mathcal{B}_i^2 \quad := \quad \left\{\max\left\{1, \frac{1}{\rho_i}\right\} < x < 1 + \frac{1}{\rho_i}\right\}$$

$$\bigcup \left\{y^2 < \frac{x(1-x)(1-\rho_i x)}{1+\rho_i(1-x)}, x \ge 1 + \frac{1}{\rho_i}\right\}.$$

In particular we have $\mathcal{B}_i^1 \subset C^+_{\min\{1,1/\rho_i\}}$. In Figure 1 the sets $\mathcal{B}_i^1$ and $\mathcal{B}_i^2$ correspond to the two shaded areas. Note that the sets $\overline{\mathcal{A}_i}$, $\mathcal{B}_i^1$ and $\mathcal{B}_i^2$ form a disjoint partition of the complex plane. When $\rho_i = 0$ the set $\mathcal{B}_i^2$ is empty.

We conclude this section with a lemma that will play a key role in the subsequent analysis.

Lemma 3.3. *For $z \in \overline{\mathcal{A}_i} - \{0\}$, we have*

$$L_i(0, s_i(z)) = \frac{z\,[D_i(z) - \phi_i(s_i(z))D_{i'}(z)]}{z-1}, \quad i = 0, 1, \quad (18)$$

and, hence, for $z \in \overline{\mathcal{A}_0 \cap \mathcal{A}_1} - \{0\}$

$$L_i(0, s_i(z)) + \phi_i(s_i(z))L_{i'}(0, s_{i'}(z)) \quad (19)$$
$$= z\,\frac{1 - \phi_0(s_0(z))\phi_1(s_1(z))}{z-1}\,D_i(z), \quad i = 0, 1.$$

Both in (18) and (19) the right-hand sides are defined by their respective analytic continuations in the given domains when $|z| > 1$. $\diamond$

Proof. Since $L_i(z,s)$ is analytic for $|z| < 1$, $\mathrm{Re}\,(s) > 0$ and continuous for $|z| \le 1$, $\mathrm{Re}\,(s) \ge 0$, the right-hand side of (11) must vanish when $s = s_i(z)$ for all $z \in \overline{\mathcal{A}_i \cap C_1^+} - \{0\}$. This gives (18) for $z \in \overline{\mathcal{A}_i \cap C_1^+} - \{0\}$. Note that the right-hand side of (18) is well defined when $z = 1$ since the numerator also vanishes at this point thanks to the identities $D_i(1) = D_i(1)$ (see (9)) and $\phi_i(s_i(1)) = \phi_i(0) = 1$. By the principle of analytic continuation, the left-hand side of (18) defines the analytic continuation of the right-hand side in $z \in \overline{\mathcal{A}_i} \cap C_1^- - \{0\}$, so that (18) holds for all $z \in \overline{\mathcal{A}_i} - \{0\}$.

Interchanging i and i' in (18), then multiplying both sides of the equation by $\phi_i(s_i(z))$ yields

$$\begin{aligned} &\phi_i(s_i(z))\, L_{i'}(0, s_{i'}(z)) \\ &\quad = \frac{z\left[\phi_i(s_i(z)) D_{i'}(z) - \phi_0(s_0(z))\phi_1(s_0(z)) D_i(z)\right]}{z-1}, \end{aligned} \tag{20}$$

for $z \in \overline{\mathcal{A}_{i'} \cap C_1^+} - \{0\}$. Summing up both sides of equations (18) and (20) gives (19) for $z \in \overline{\mathcal{A}_0 \cap \mathcal{A}_1 \cap C_1^+} - \{0\}$. By definition of the sets $\mathcal{A}_0$ and $\mathcal{A}_1$ we observe that the left-hand side of (19) is analytic for $z \in \mathcal{A}_0 \cap \mathcal{A}_1$ (since $\mathrm{Re}\,(s_i(z)) > 0$ for $i = 0, 1$) and continuous for $z \in \overline{\mathcal{A}_0 \cap \mathcal{A}_1} - \{0\}$ (since $\mathrm{Re}\,((s_i(z)) \geq 0$ for $i = 0, 1$). $\square$

The next corollary gives $L_i(z, 0)$, $i = 0, 1$, in terms of $D_i(z)$ and $L_{i'}(0, s_{i'}(z))$. Therefore, once these functions are found, we can compute $N(z)$ (and therefore $\overline{N}$) from (14), as shown in Section 6.

Corollary 3.1. *For $|z| \leq 1$,*

$$\begin{aligned} L_i(z,0) &= \frac{L_i(0,0) + L_{i'}(0, s_{i'}(z))}{1 - \rho_i z} \\ &\quad + z\,\frac{1 - \phi_{i'}(s_{i'}(z))}{(1-z)\,(1 - \rho_i z)}\, D_i(z), \end{aligned} \tag{21}$$

$$\begin{aligned} L_{i'}(z,0) &= \frac{L_{i'}(0,0) - L_{i'}(0, s_{i'}(z))}{1 - \rho_{i'} z} \\ &\quad - z\,\frac{1 - \phi_{i'}(s_{i'}(z))}{(1-z)\,(1 - \rho_{i'} z)}\, D_i(z). \end{aligned} \tag{22}$$

In both equations, the right-hand sides are given by their analytic continuations for $z \in C_1^+ - \mathcal{A}_0 \cap \mathcal{A}_1 - \{0\}$.

Moreover, the constant $L_i(0,0)$ in (21) is given by

$$L_i(0,0) = \frac{(1-\rho_i)\,\hat{a}_i + (1 - \rho_{i'})\,\hat{a}_{i'}}{\hat{a}_0 + \hat{a}_1} - L_{i'}(0,0). \tag{23}$$

$\diamond$

Proof. From (18), with i and i' interchanged, we obtain

$$z\, D_{i'}(z) = z\, \phi_{i'}(s_{i'}(z))\, D_i(z) - (1-z)\, L_{i'}(0, s_{i'}(z)), \tag{24}$$

for $z \in \overline{\mathcal{A}_0 \cap \mathcal{A}_1 \cap C_1^+} - \{0\}$.

Now (21) is found by setting $s = 0$ in (11) and replacing $zD_{i'}(z)$ by the right-hand side of (24). Equation (21) holds, a priori, for $z \in \overline{\mathcal{A}_0 \cap \mathcal{A}_1 \cap C_1^+} - \{0\}$, with $z \notin \{1, 1/\rho_i\}$ (note that the right-hand side of (21) is well defined if $z = 1$ when $\rho_i \neq 1$).

In the same way, we now determine $L_{i'}(z, 0)$. Interchanging i and i' in (11) and letting $s = 0$, gives

$$L_{i'}(z,0) = \frac{L_{i'}(0,0)}{1 - \rho_{i'} z} + z\,\frac{D_{i'}(z) - D_i(z)}{(1-z)\,(1 - \rho_{i'} z)}$$

for $|z| \leq 1$, $z \notin \{1, 1/\rho_{i'}\}$. Substituting (24) into the above equation, gives (22) for $z \in \overline{\mathcal{A}_0 \cap \mathcal{A}_1 \cap C_1^+} - \{0\}$, with $z \notin \{1, 1/\rho_{i'}\}$.

From their definitions, $L_0(z, 0)$ and $L_1(z, 0)$ are analytic for $|z| < 1$ and continuous for $|z| \leq 1$. We may therefore invoke the principle of analytic continuation to define (21) and (22) in the entire domain $\{|z| \leq 1\}$.

Finally, expression (23) is obtained by letting $z = 1$ in (21) and by using (9) and (13) together with the identity $\lim_{z \to 1}(1 - \phi_{i'}(s_{i'}(z))/(1-z) = -\hat{a}_{i'}\,(1 - \rho_{i'})$. $\square$

4. ONE RATIONAL LAPLACE-STIELTJES TRANSFORM

In this section we assume that $\phi_1(s)$ is a rational function. This occurs when $F_1(x)$ has a phase-type distribution. Lemma 4.1 shows that in this case the function $L_1(0, s)$ is a rational function too. This result will be used in the next section to compute $L_i(z, s)$ for $i = 0, 1$.

Lemma 4.1. *Assume $\phi_1(s)$ is a rational function and write*

$$\phi_1(s) = \frac{\eta_1(s)}{\delta_1(s)}, \tag{25}$$

where $\eta_1(s)$ and $\delta_1(s)$ are polynomials of degree n_1 and d_1, respectively, that have no common zeros. (Since $\phi_1(s)$ is the LST of the distribution of a random variable with support on the positive real line, it must be that $n_1 < d_1$.) Then $L_1(0, s)$ is a rational function too, and can be written as

$$L_1(0,s) = \frac{\zeta_1(s)}{\delta_1(s)}, \tag{26}$$

where $\zeta_1(s)$ is a polynomial of degree $d_1 - 1$. $\diamond$

Proof. For $\mathrm{Re}\,(s) \geq 0$ define,

$$\zeta_1(s) := \delta_1(s) L_1(0, s).$$

Clearly, the function $\zeta_1(s_1(z))$ is analytic for all $z \in \mathcal{A}_1$. Using (18) we have

$$\zeta_1(s_1(z)) = \frac{z\left[\delta_1(s_1(z)) D_1(z) - \eta_1(s_1(z)) D_0(z)\right]}{z-1}, \tag{27}$$

for $z \in \mathcal{A}_1$ (where for $|z| \geq 1$ the right-hand side must be understood to be its analytic continuation). Note that from the right-hand side it follows that we can analytically continue the function $\zeta_1(s_1(z))$ to all $0 < |z| \leq 1$. Hence, $\zeta_1(s_1(z))$ is analytic for $z \in \mathcal{A}_1 \cup \overline{\mathcal{B}_1^1} - \{0\}$. Since $\min\{1, 1/\rho_1\} \leq \sqrt{1/\rho_1} \leq \max\{1, 1/\rho_1\}$ it follows that $C^+_{\sqrt{1/\rho_1}} - \{0\} \subset \mathcal{A}_1 \cup \overline{\mathcal{B}_1^1} - \{0\}$ (see also Figure 1). From Lemma 3.1 and Lemma 3.2 it then follows that $\zeta_1(s) = \zeta_1(s_1(z_1^+(s)))$ is analytic in $s \in \mathbf{C} - \left[s_1^1, s_1^2\right]$. From its definition, $\zeta_1(s)$ is also analytic for $\mathrm{Re}\,(s) \geq 0$. Hence $\zeta_1(s)$ is analytic in the entire plane.

From Liouville's theorem (cf. [16, p. 90]) we now have that $\zeta_1(s)$ is a polynomial of degree at most $d_1 - 1$, since the multiplicity of the (possible) singularity at infinity is not more than $d_1 - 1$:

$$\begin{aligned}\lim_{|s|\to\infty} \frac{\zeta_1(s)}{s^{d_1-1}} &= \lim_{z\to 0} \frac{\zeta_1(s_1(z))}{(s_1(z))^{d_1-1}} \\ &= \lim_{z\to 0} \frac{-\delta_1(s_1(z))D_1(z) + \eta_1(s_1(z))D_0(z)}{(s_1(z))^{d_1}\left(\frac{1}{z}-1\right)/s_1(z)} \\ &= D_1(0) \lim_{z\to 0} \frac{\delta_1(s_1(z))}{(s_1(z))^{d_1}}.\end{aligned}$$

Since $D_1(0) = D\hat{P}_1(0) > 0$ — see above (3) — the degree is exactly $d_1 - 1$. □

REMARK 4.1. *By symmetry, Lemma 4.1 also applies to* $\phi_0(s)$ *and* $L_0(0,s)$ *when* $\phi_0(s)$ *is a rational function. When both LST's* $\phi_0(s)$ *and* $\phi_1(s)$ *are rational, then Lemma 4.1 leads to the solution of (11). The* $d_0 + d_1$ *unknown coefficients in the polynomials* $\zeta_0(s)$ *and* $\zeta_1(s)$ *in (26) are then determined by (12) and the equations resulting from the zeros of* $1 - \phi_0(s_0(z))\phi_1(s_1(z))$ *inside the unit disk. This derivation, which is not the main objective of the present study, will be found in an extended version of this work.* ⋄

5. REDUCTION TO A BOUNDARY VALUE PROBLEM

We show that $D_0(z)$ can be obtained as the solution of a Riemann-Hilbert boundary value problem. Due to space constraints, we only investigate the case when $0 < \rho_0 < 1$. The general case will be addressed in an extended version of this paper.

In addition, the following assumptions will be enforced from now on:

Assumptions:

A1 $\phi_1(s) = \eta_1(s)/\delta_1 s)$ is rational with $\eta_1(s)$ and $\delta_1(s)$ polynomials of degree n_1 and d_1, respectively, with $n_1 < d_1$, and $\phi_0(s)$ is the LST of a general distribution function;

A2 $\delta_1(s_1(z))$ does not vanish for $z \in C_{1/\sqrt{\rho_0}}$ (see Remark 5.1);

A3 The stability condition (6) holds.

Recall that Assumption **A1** implies (Lemma 4.1)

$$L_1(0,s) = \frac{\zeta_1(s)}{\delta_1(s)}. \tag{28}$$

By $z_1, \ldots, z_M$ we shall denote the distinct zeros (with multiplicity $m_1, \ldots, m_M$, respectively), if any, of the function $1 - \phi_0(s_0(z))\phi_1(s_1(z))$ in

$$\mathcal{D} := \{1 < |z| \leq 1/\sqrt{\rho_0}\}. \tag{29}$$

The function $1 - \phi_0(s_0(z))\phi_1(s_1(z))$ being meromorphic in the (bounded) domain $\mathcal{D}$, it has a finite number of zeros/poles in this domain. Hence, $M < \infty$.

With i the imaginary unit, define the function

$$G(z) := \frac{\mathrm{i}\,(1-z)\,R(z)}{z\,(1 - \phi_0(s_0(z))\phi_1(s_1(z)))}, \tag{30}$$

where

$$R(z) := \prod_{k=1}^{M} (z - z_k)^{m_k}, \tag{31}$$

if $M \geq 1$, and $R(z) \equiv 1$ if $M = 0$. Also define

$$\psi(z) := \frac{G(z)}{|G(z)|}, \tag{32}$$

and

$$c(z) = \phi_0(s_0(z))\,|G(z)|\,\mathrm{Re}\left(\mathrm{i}\,\frac{\zeta_1(s_1(z))}{\delta_1(s_1(z))}\right). \tag{33}$$

Observe that $c(z)$ is *real-valued* when $z \in C_{1/\sqrt{\rho_0}}$ (since $s_0(z)$ is real for z lying on this contour by Lemma 3.2).

For later use, some properties of the functions $\psi(z)$ and $c(z)$ are collected in Lemma 5.1. We recall that a function $f(z)$ satisfies a Hölder condition on some smooth contour L if there exist constants $A > 0$ and $0 < \nu \leq 1$ such that $|f(z_2) - f(z_1)| \leq A|z_2 - z_1|^\nu$ for all $z_1, z_2 \in L$ [22, pp. 11-12]. The following properties are direct consequences of this definition [22, pp. 13-21]:

P1 If $f(z)$ satisfies a Hölder condition on a contour L then so do the functions $|f(z)|$, $\mathrm{Re}\,(f(z))$, $\mathrm{Im}\,(f(z))$ and — if $f(z)$ does not vanish on L — $1/f(z)$.

P2 If $f(z)$ and $g(z)$ both satisfy a Hölder condition on the same contour then so do the functions $f(z)g(z)$ and $f(z) + g(z)$.

P3 If $f(z)$ is differentiable on some contour L with a uniformly bounded derivative on L, then $f(z)$ satisfies a Hölder condition on L.

LEMMA 5.1. *Under assumptions* **A1**-**A3** *both functions* $\psi(z)$ *and* $c(z)$ *satisfy a Hölder condition on the circle* $C_{1/\sqrt{\rho_0}}$*; moreover* $\psi(z)$ *is non-vanishing on* $C_{1/\sqrt{\rho_0}}$ *and* $|\psi(z)| = 1$ *for all* $z \in C_{1/\sqrt{\rho_0}}$. ⋄

The proof of Lemma 5.1 is given in the Appendix.

Since, by Lemma 5.1, $\psi(z)$ is well-defined and non-vanishing on $C_{1/\sqrt{\rho_0}}$, we can define the index of this function on the contour $C_{1/\sqrt{\rho_0}}$,

$$\chi := \frac{1}{2\pi}\,[\arg\,\psi(z)]_{C_{1/\sqrt{\rho_0}}}, \tag{34}$$

that is the variation of the argument of $\psi(z)$ as z moves counter clock wise along the contour $C_{1/\sqrt{\rho_0}}$, divided by 2π [16, 22] .

The index χ, which will play an important role in the sequel, is determined in the following lemma.

We introduce some additional notation: Let N_Z be the number of poles (counting multiplicities) of $1-\phi_0(s_0(z))\,\phi_1(s_1(z))$ in $\mathcal{M}$, where

$$\mathcal{M} := \{1 < |z| < 1/\sqrt{\rho_0}\}. \tag{35}$$

Under the enforced assumption **A1**, N_Z is the number of zeros (hence, the subscript Z in N_Z) of $\delta_1(s_1(z))/\phi_0(s_0(z))$ in $\mathcal{M}$ (counting multiplicities).

LEMMA 5.2. *Under assumptions* **A1**-**A3**, *the index* χ *is given by*

$$\chi = N_Z.$$

In particular, $\chi = 0$ *when* $0 \le \rho_1 \le \sqrt{\rho_0} < 1$. $\diamond$

PROOF. Recall the definition of $z_1, \ldots, z_M$, the distinct zeros of $1 - \phi_0(s_0(z))\,\phi_1(s_1(z))$ in $\mathcal{D}$. Without loss of generality, let us assume that $z_1, \ldots, z_{M_0}$ and $z_{M_0+1}, \ldots, z_M$ $(0 \le M_0 \le M)$ lie in the domain $\mathcal{M}$ and on the contour $C_{1\sqrt{\rho_0}}$, respectively. Define $R_1(z) = \prod_{k=1}^{M_0}(z - z_k)^{m_k}$ and $R_2(z) = \prod_{k=M_0+1}^{M}(z - z_k)^{m_k}$, so that $R(z)$ in (31) is given by $R(z) = R_1(z)\,R_2(z)$. With these definitions, $G(z)$ in (30) rewrites $G(z) = \mathrm{i}R_1(z)/zW(z)$, with

$$W(z) := \frac{1 - \phi_0(s_0(z))\,\phi_1(s_1(z))}{(1 - z)\,R_2(z)}. \tag{36}$$

The function $W(z)$ is meromorphic in $\mathcal{M}$, continuous and non-vanishing on $C_{1/\sqrt{\rho_0}}$. It has M_0 zeros $z_1, \ldots, z_{M_0}$ and N_Z poles. Define its index κ on $C_{1/\sqrt{\rho_0}}$, namely,

$$\kappa := \frac{1}{2\pi}\,[\arg W(z)]_{C_{1/\sqrt{\rho_0}}}. \tag{37}$$

By the argument principle [21, pp. 48_2-52_2], we have

$$\begin{aligned}\chi &= \frac{1}{2\pi}\,[\arg R_1(z)]_{C_{1/\sqrt{\rho_0}}} - \frac{1}{2\pi}\,[\arg z]_{C_{1/\sqrt{\rho_0}}} \\ &\quad - \frac{1}{2\pi}\,[\arg W(z)]_{C_{1/\sqrt{\rho_0}}} \\ &= \sum_{m=1}^{M_0} m_k - 1 - \kappa. \end{aligned} \tag{38}$$

It remains to compute κ. To this end, consider the closed contour $\Gamma_{A,B,C,D}$ depicted in Figure 2: It is composed of the circle $\Gamma_{A,C}$ with center 0 and radius $1/\sqrt{\rho_0}$ (resp. the circle $\Gamma_{B,D}$ with center 0 and radius 1) from which we have removed the arc (AC) (resp. (BD)), and of the segments $[A, B]$ and $[D, C]$. The points A, B, C and D are chosen in such a way that all zeros and poles of $W(z)$ in the domain $\mathcal{M}$ also lie inside the contour $\Gamma_{A,B,C,D}$ (this is possible since there are a finite number of such zeros and poles and also because none of them lie on the segment $[-1/\sqrt{\rho_0}, -1]$).

As already observed, $W(z)$ is a meromorphic function inside the contour $\Gamma_{A,B,C,D}$, and it is well-defined on $\Gamma_{A,B,C,D}$. Hence,

$$\frac{1}{2\pi\mathrm{i}}\int_{\Gamma_{A,B,C,D}} \frac{W'(z)}{W(z)}\,dz = \sum_{m=1}^{M_0} m_k - N_Z. \tag{39}$$

This result is a well-known consequence of the theorem of residues applied to the logarithmic derivative

$$w(z) := W'(z)/W(z)$$

of the function $W(z)$ [21, pp. 48_2-52_2]. Since $\Gamma_{A,B,C,D} = \Gamma_{A,C} \cup [A, B] \cup \Gamma_{B,D} \cup [D, C]$, (39) rewrites as

$$\begin{aligned} &\frac{1}{2\mathrm{i}\pi}\int_{\Gamma_{A,C}} w(z)\,dz + \frac{1}{2\mathrm{i}\pi}\int_A^B w(z)\,dz \\ &+ \frac{1}{2\mathrm{i}\pi}\int_D^C w(z)\,dz + \frac{1}{2\mathrm{i}\pi}\int_{\Gamma_{B,D}} w(z)\,dz \\ &= \sum_{m=M_0+1}^{M} m_k - N_Z. \end{aligned}$$

Letting now A and C tend to $-1/\sqrt{\rho_0}$ (resp. B and D tend to -1) in the latter equation, we see that the 1st integral is nothing but the index κ, that the 2nd and the 3rd integral cancel each other, and that the 4th integral is equal to $-\frac{1}{2\pi}\,[\arg W(z)]_{C_1}$ (since z moves in the negative direction on C_1). Consequently,

$$\kappa = \sum_{m=1}^{M_0} m_k - N_Z + \frac{1}{2\pi}\,[\arg W(z)]_{C_1}. \tag{40}$$

We are left with computing $\kappa_0 := \frac{1}{2\pi}\,[\arg W(z)]_{C_1}$. Since $R_2(z)$ is a polynomial with no zeros in $\{|z| \le 1\}$, we may invoke again the argument principle, to obtain that $\kappa_0 = \frac{1}{2\pi}\,[\arg f(z)]_{C_1}$, where we have set

$$f(z) := (1 - \phi_0(s_0(z))\,\phi_1(s_1(z)))/(1 - z).$$

For $z \in C_1 - \{1\}$, we know by Lemma 3.2, that $\mathrm{Re}\,(s_0(z)) > 0$ and $\mathrm{Re}\,(s_1(z)) > 0$, thereby implying that

$$|\phi_0(s_0(z))\,\phi_1(s_1(z)| < 1,$$

which in turn implies that $\mathrm{Re}\,(1 - \phi_0(s_0(z))\,\phi_1(s_1(z))) > 0$. Since $\mathrm{Re}\,(1 - z) > 0$ for $z \in C_1 - \{1\}$, this implies that $f(z)$ does not cross the negative real axis when $z \in C_1 - \{1\}$ [Hint: if $\mathrm{Arg}(z_i) \in (-\pi/2, \pi/2)$ for $i = 1, 2$, then $\mathrm{Arg}(z_1/z_2) = \mathrm{Arg}(z_1) - \mathrm{Arg}(z_2) \in (-\pi, \pi)$]. On the other hand, an application of L'Hopital's rule shows that

$$\lim_{z \to 1} f(z) = \hat{a}_0(\rho_0 - 1) + \hat{a}_1(\rho_1 - 1) < 0, \tag{41}$$

where the latter inequality follows from assumption **A3**. Since $f(z)$ is a continuous function of $z \in C_1$, the above shows that, as z describes the unit circle, $f(z)$ describes once a closed contour, say $\mathcal{C}$, around $z = 0$, crossing the negative real axis only at $z = 1$ (with $f(1) < 0$).

It remains to determine the direction in which $f(z)$ moves along $\mathcal{C}$ as z moves along C_1 in the positive direction. Take $z = e^{\mathrm{i}\theta} \in C_1$, with $\theta < 0$, close enough to the point 1; then $0 < \mathrm{Arg}(1 - z) < \pi/2$. On the other hand, $\mathrm{Re}\,(1 - \phi_0(s_0(z))\phi_1(s_1(z))) > 0$ for $z \in C_1 - \{1\}$ as already observed, so that $-\pi/2 < \mathrm{Arg}(1 - \phi_0(s_0(z))\phi_1(s_1(z))) < \pi/2$. Therefore, $-\pi < \mathrm{Arg}(f(z)) < 0$. When $\theta \to 0$ with $\theta < 0$ (i.e. in the positive direction) then $f(z) \to \hat{a}_0(\rho_0 - 1) + \hat{a}_1(\rho_1 - 1) < 0$ from below (since $-\pi < \mathrm{Arg}(f(z)) < 0$), i.e, in the negative direction. Hence, $\kappa_0 = -1$. The latter result, combined together with (38) and (40), proves that $\chi = N_Z$.

When $0 \le \rho_1 \le \sqrt{\rho_0} < 1$ we know by Lemma 3.2 that $\mathrm{Re}\,(s_1(z)) \ge 0$ (in particular) for $z \in \mathcal{D}$, which implies that $\phi_1(s_1(z))$ has no pole in $\mathcal{D}$. Therefore, $N_Z = 0$ and $\chi = 0$. $\square$

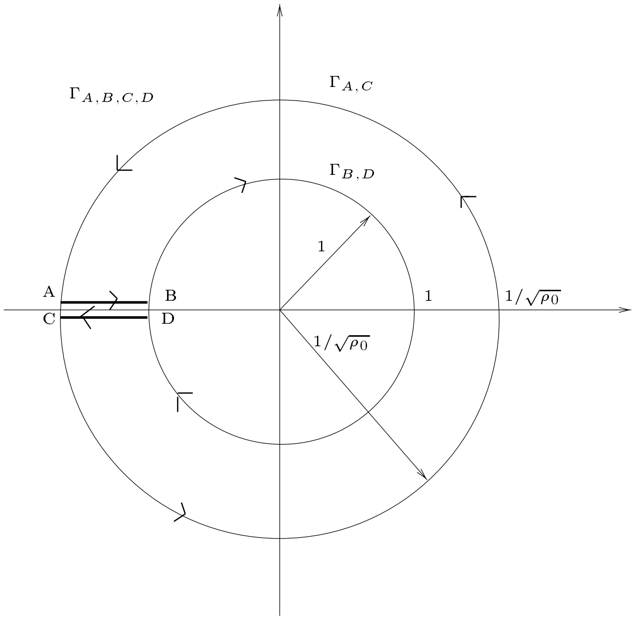

Figure 2: The contour $\Gamma_{A,B,C,D}$

The following proposition gives $D_0(z)$ for all $|z| < 1/\sqrt{\rho_0}$:

PROPOSITION 5.1. *Under assumptions* **A1**-**A3** *the following holds:*

$$D_0(z) = \frac{\left(z\sqrt{\rho_0}\right)^{\chi}\, e^{i\,\gamma(z)}}{R(z)}\,\left(\Theta(z) + \mathbf{1}(\chi \geq 0)\, Q(z)\right), \quad (42)$$

for all $|z| < 1/\sqrt{\rho_0}$, where

$$\gamma(z) := \frac{1}{2\pi}\int_0^{2\pi}\left(\operatorname{arctg}\left(\frac{b(e^{i\,\sigma}/\sqrt{\rho_0})}{a(e^{i\,\sigma}/\sqrt{\rho_0})}\right) - \chi\sigma\right) \times h(\sigma, z)\, d\sigma, \quad (43)$$

$$\Theta(z) := \frac{1}{2\pi}\int_0^{2\pi} e^{\omega_1(e^{i\,\sigma}/\sqrt{\rho_0})}\, c(e^{i\,\sigma}/\sqrt{\rho_0}) h(\sigma, z)\, d\sigma, \quad (44)$$

$$\omega_1(z) := \operatorname{Im}\left(\gamma(z)\right), \quad (45)$$

$$Q(z) := i\,\beta_0 + \sum_{k=1}^{\chi}\left(c_k\left(z\sqrt{\rho_0}\right)^k - \overline{c}_k\frac{1}{\left(z\sqrt{\rho_0}\right)^k}\right), \quad (46)$$

with

$$h(\sigma, z) := \frac{e^{i\,\sigma} + z\sqrt{\rho_0}}{e^{i\,\sigma} - z\sqrt{\rho_0}},$$

where the sum in (46) equals 0 when $\chi = 0$, and β_0 and the c_k's are constants to be determined. In (43) $a(z)$ and $b(z)$ are the two real functions defined by $a(z) := \operatorname{Re}(\psi(z))$ and $b(z) := \operatorname{Im}(\psi(z))$. $\diamond$

PROOF. With (28), equation (19) becomes

$$\delta_1(s_1(z))\, L_0(0, s_0(z)) + \phi_0(s_0(z))\zeta_1(s_1(z)) \quad (47)$$
$$= z\,\delta_1(s_1(z))\,\frac{1 - \phi_0(s_0(z))\phi_1(s_1(z))}{(z-1)R(z)}\, R(z)D_0(z),$$

for $z \in \overline{\mathcal{A}_0 \cap \mathcal{A}_1} - \{0\}$, where $G(z)$ and $R(z)$ are defined in (30) and (31), respectively.

The left-hand side of (47) being analytic in $\mathcal{A}_0$ and continuous in $\overline{\mathcal{A}_0} - \{0\}$, it defines the analytic continuation of the right-hand side in $\overline{\mathcal{A}_0} - \{0\}$. In particular, the right-hand side of (47) is analytic in $\{1 < |z| < 1/\sqrt{1/\rho_0}\}$ and continuous in $\{1 \leq |z| \leq 1/\sqrt{1/\rho_0}\}$, since $\{1 \leq |z| \leq 1/\sqrt{1/\rho_0}\} \subset \overline{\mathcal{A}_0}$ (Lemma 3.2). Now, since the function $z\delta_1(s_1(z))\,(1 - \phi_0(s_0(z))\phi_1(s_1(z)))/((z-1)R(z))$ is analytic in $\{1 < |z| < 1/\sqrt{1/\rho_0}\}$, continuous and non-vanishing in $\{1 \leq |z| \leq 1/\sqrt{1/\rho_0}\}$, we may conclude that $R(z)D_0(z)$ is also analytic in $\{1 < |z| < 1/\sqrt{1/\rho_0}\}$ and continuous in $\{1 \leq |z| \leq 1/\sqrt{1/\rho_0}\}$. In summary, we have shown that $R(z)D_0(z)$ is analytic in $\{|z| < 1/\sqrt{\rho_0}\}$ and continuous in $\{|z| \leq 1/\sqrt{\rho_0}\}$ when $\rho_0 < 1$.

Letting $z \in C_{1/\sqrt{\rho_0}}$ in (47), multiplying both sides of the equation by $i/\delta_1(s_1(z))$, then taking the real part on both sides, and finally noting that $s_0(z)$ is real (Lemma 3.2), yields

$$\operatorname{Re}\left(i\,z\frac{1 - \phi_0(s_0(z))\phi_1(s_1(z))}{z-1}\, D_0(z)\right)$$
$$= \phi_0(s_0(z))\operatorname{Re}\left(i\, L_1(0, s_1(z))\right), \qquad z \in C_{1/\sqrt{\rho_0}}. (48)$$

With (32)-(33) we may rewrite (48) as

$$\operatorname{Re}\left(\frac{R(z)D_0(z)}{\psi(z)}\right) = c(z), \qquad z \in C_{1/\sqrt{\rho_0}}, \qquad (49)$$

where $\psi(z)$ is non-vanishing on $C_{1/\sqrt{\rho_0}}$; more precisely, we know that $|\psi(z)| = 1$ everywhere on $C_{1/\sqrt{\rho_0}}$ by Lemma 5.1.

Equation (49) defines a Riemann-Hilbert boundary value problem on the contour $C_{1/\sqrt{\rho_0}}$, namely, it is required to find a function $R(z)D_0(z)$ that is analytic in $C^+_{1/\sqrt{\rho_0}}$, continuous in $\overline{C^+_{1/\sqrt{\rho_0}}}$ and which satisfies a condition of the type (49) on the contour $C_{1/\sqrt{\rho_0}}$.

Under assumptions **A1**-**A3** we know, by Lemma 5.1, that $\psi(z)$ and $c(z)$ satisfy a Hölder condition on $C_{1/\sqrt{\rho_0}}$. Therefore, the solution to the boundary value problem (49) is given in [16, Sect. 29.3] (see also [22, pp. 100-107]) which proves (42). $\square$

Let us pause for a while and summarize the results we have obtained so far. Lemma 4.1 determines $L_1(0, s)$ up to the d_1 unknown coefficients of the polynomial $\zeta_1(s)$ (of degree $d_1 - 1$). Proposition 5.1 then gives $D_0(z)$ in the entire unit disk, up to the d_1 coefficients of $\zeta_1(s)$ — which appear in $c(z)$ — and the unknown $\chi+1$ additional constants $\beta_0, c_1, \ldots, c_\chi$ involved in $Q(z)$. Note that both $L_1(0, s)$ and $D_0(z)$ are linear in these unknown coefficients. Moreover, for any choice of these coefficients, $L_1(0, s)$ is meromorphic in the entire complex plane (its poles being the zeros of $\delta_1(s)$) and $D_0(z)$ is meromorphic in $\{|z| \leq 1/\sqrt{1/\rho_0}\}$ (its poles being the zeros of $R(z)$. From this we can determine the transform $D_1(z)$ (up to the $d_1 + \chi + 1$ unknown constants) for $|z| \leq 1$, and $\delta_1(s_1(z)) \neq 0$, using (18) with $i = 1$:

$$D_1(z) = \frac{z-1}{z} L_1(0, s_1(z)) + \phi_1(s_1(z))D_0(z). \qquad (50)$$

Similarly we find, using (19) with $i=0$, for $z \in \overline{C}_{1/\sqrt{\rho_0}} \bigcap \mathcal{A}_0$ and $\delta_1(s_1(z)) \neq 0$,

$$L_0(0, s_0(z)) = -\phi_0(s_0(z)) L_1(0, s_1(z)) + z\, \frac{1-\phi_0(s_0(z))\phi_1(s_1(z))}{z-1}\, D_0(z). \tag{51}$$

The following two lemmas give $d_1 + \chi$ linear relations to be satisfied by the unknown coefficients. By s_k, $k = 1, \ldots, d_1$, we shall denote the zeros of the polynomial $\delta_1(s)$.

LEMMA 5.3. *For $k = 1, 2, \ldots, d_1$,*

$$\zeta_1(s_k) = \frac{z_1^+(s_k)\eta_1(s_k) D_0(z_1^+(s_k))}{1 - z_1^+(s_k)}. \tag{52}$$

If s_k ($k = 1, 2, \ldots, d_1$) is a zero of multiplicity $t_k > 1$, then we have the additional set of linear equations

$$\frac{\partial^n}{\partial s_k^n} \zeta_1(s_k) = \frac{\partial^n}{\partial s_k^n} \left(\frac{z_1^+(s_k)\eta_1(s_k) D_0(z_1^+(s_k))}{1 - z_1^+(s_k)} \right), \tag{53}$$

for $n = 1, \ldots, t_k - 1$. ◇

PROOF. To see (52) note that if $\delta_1(s_k) = 0$ then $\mathrm{Re}\,(s_k) < 0$. Therefore $|z_1^+(s_k)| < 1$ (in fact $|z_1^+(s_k)| < \min\{1, 1/\rho_1\}$). Since $D_1(z)$ is analytic in C_1^+, we can substitute $z = z_1^+(s)$ into (50) and let $s \to s_k$. A similar reasoning leads to the additional equations in case the zero s_k of $\delta_1(s)$ has multiplicity larger than 1. □

The linear equations (52) and (53) give us d_1 equations to which we can add the normalizing condition (12). However, when $\chi > 0$ then χ extra conditions are needed in order to compute the $d_1 + \chi + 1$ unknown constants. These χ additional (linear) equations are provided in the next lemma. Recall (cf. Lemma 5.2) that $\chi = N_Z$, i.e., χ equals the number of zeros of $\delta_1(s_1(z))/\phi_0(s_0(z))$ in $\mathcal{M}$. This number is not greater than the number of zeros of $\delta_1(s_1(z))$ in $\mathcal{M}$. Note that the zeros of $\delta_1(s_1(z))$ in $\mathcal{M} \subset C_1^-$ must be in $\mathcal{B}_1^2$ (see Figure 1). There are exactly d_1 zeros in $\mathcal{B}$, since these are the images of the s_k, $k = 1, \ldots, d_1$, $\mathcal{M}$ under the mapping $z_1^-(s)$. As a result we have $\chi = N_Z \leq d_1$.

We partition the zeros of $\delta_1(s)$ into two subsets: For $k = 1, \ldots, \chi$, s_k will belong to a zero of $\delta_1(s)/\phi_0(s_0(z_1^-(s)))$ (counting multiplicities) with $z_1^-(s_k) \in \mathcal{M}$; for $k = \chi + 1, \ldots, d_1$, either $z_1^-(s_k) \notin \mathcal{M}$ or the zero is cancelled by a zero of $\phi_0(s_0(z_1^-(s)))$. In the above, we have implicitly used the relation $s_1(z_1^-(s)) = s$ (cf. Lemma 3.2).

LEMMA 5.4. *For $k = 1, \ldots, \chi$,*

$$\zeta_1(s_k) = \frac{z_1^-(s_k)\eta_1(s_k) D_0(z_1^-(s_k))}{1 - z_1^-(s_k)}. \tag{54}$$

If the zero s_k of $\delta_1(s)/\phi_0(s_0(z_1^-(s)))$ has multiplicity $t_k > 1$ ($k = 1, \ldots, \chi$), then

$$\frac{\partial^n}{\partial s_k^n} \zeta_1(s_k) = \frac{\partial^n}{\partial s_k^n} \left(\frac{z_1^-(s_k)\eta_1(s_k) D_0(z_1^-(s_k))}{1 - z_1^-(s_k)} \right), \tag{55}$$

for $n = 1, \ldots, t_k - 1$. ◇

PROOF. The proof is analogous to that of Lemma 5.3. If $\delta_1(s_k)/\phi_0(s_0(z_1^-(s_k)) = 0$ and $z_1^-(s_k) \in \mathcal{M}$ then we can substitute $z = z_1^-(s)$ into (51) and let $s \to s_k$ (because $L_0(s_0(z))$ is analytic in $\mathcal{A}_0$). The additional equations in case the zero s_k of $\delta_1(s)/\phi_0(s_0(z_1^-(s_k)))$ has multiplicity larger than 1 follow similarly. □

The following proposition summarizes how the functions on the right-hand sides of the functional equations (11) and, hence, the functions $L_0(z,s)$ and $L_1(z,s)$ can be determined.

PROPOSITION 5.2. *Suppose $0 < \rho_0 < 1$ and that Assumptions* **A1-A3** *hold. Then $L_1(0,s)$, $D_0(z)$, $D_1(z)$ and $L_0(0,s)$ are given by (28), (42), (50) and (51) respectively. The $d_1 + \chi + 1$ unknown coefficients introduced in the functions $\zeta_1(s)$ and $Q(z)$ are determined by (12), (52), (53), (54) and (55). These $d_1 + \chi + 1$ equations are linear in the unknowns.* ◇

PROOF. For any choice of the $d_1 + \chi + 1$ unknowns, $L_1(0,s)$ and $D_0(z)$ are analytic in the right-half plane and inside the unit disc, respectively. Because of (52), (53), (54) and (55), $D_1(z)$ and $L_0(0,s)$ are also analytic inside the unit disc and in the right-half plane, respectively. Using Lemma 3.3, $L_0(z,s)$ and $L_1(z,s)$ given by (11) are analytic for $|z| < 1$, $\mathrm{Re}\,(s) > 0$ and continuous for $|z| \leq 1$ and $\mathrm{Re}\,(s) \geq 0$. Equations (52), (53), (54) and (55) form a homogeneous linear system in the unknowns, which is invariant with respect to multiplication by a constant. Equation (12) determines this constant. Theorem 2.1 ensures this solution is unique. The equations are linear in the unknowns because $\zeta_1(s)$ and $Q(z)$ are linear in the unknowns. □

REMARK 5.1. *If $\delta_1(s_1(z))$ vanishes on the circle $1/\sqrt{\rho_0}$ for some values of the model parameters, then this can only occur at a finite number of points on this contour since $\delta_1(s_1(z))$ is a meromorphic function. When this occurs, we can usually slightly perturb one of the parameters so that* **A2** *and* **A3** *are satisfied, and use the analysis developed in this section. As far as numerical results are concerned, this approach will give accurate results since the performance metrics under consideration (e.g. average queue-length) are continuous in the model parameters.*

6. THE z-TRANSFORM

We now return to the computation of the z-transform $N(z)$. We have already shown in (14) that

$$N(z) = \frac{1}{c\mu_0}\, L_0(z,0) + \frac{1}{c\mu_1}\, L_1(z,0), \quad |z| \leq 1.$$

Introducing in this formula the values for $L_i(z,0)$, $i = 0,1$, found in Corollary 3.1 yields [with $i = 0$, $i' = 1$]

$$N(z) = \frac{1}{c}\left(\frac{1}{\mu_0\,(1-\rho_0 z)} - \frac{1}{\mu_1\,(1-\rho_1 z)} \right) \times \left(L_1(0, s_1(z)) + z\, \frac{1-\phi_1(s_1(z))}{1-z}\, D_0(z) \right) + \frac{L_0(0,0)}{c\mu_0\,(1-\rho_0 z)} + \frac{L_1(0,0)}{c\mu_1\,(1-\rho_1 z)}. \tag{56}$$

On the other hand, we have established in (23) that

$$L_0(0,0) = \frac{\hat{a}_0\,(1-\rho_0) + \hat{a}_1\,(1-\rho_1)}{\hat{a}_0 + \hat{a}_1} - L_1(0,0),$$

thereby showing that $N(z)$ in (56) is entirely determined for $|z| \leq 1$ when one knows $L_1(0,s)$ for $\mathrm{Re}\,(s) \geq 0$ and $D_0(z)$ for $|z| \leq 1$. Proposition 5.2 determines these functions when $F_1(x)$ has a rational LST, for $0 < \rho_0 < 1$ (and under the technical assumption **A2**). The situation where $\rho_0 \geq 1$ will be addressed in an extended version of this paper.

7. CONCLUDING REMARKS

The generating function for the queue-length distribution in an M/M/1 queue evolving in a semi-Markov environment has been obtained via techniques pertaining to complex analysis. An interesting feature of this queueing model is that the traffic process can exhibit burstiness and long-range dependence depending on the choice of the model parameters. We emphasize that the obtained results allow for numerical computation of performance measures, particularly the mean queue-length. This allows us to evaluate the impact of such correlated processes on the performance measures of interest. This evaluation which has not been carried out in the present paper is the subject of ongoing, mostly numerical, studies.

Acknowledgements: We thank Mor Harchol-Balter and Nikhil Bansal (Carnegie Mellon University) for useful discussions at the beginning of this work.

APPENDIX: Proof of Lemma 5.1

PROOF. Under assumptions **A1-A3** and the definition of the function $R(z)$ it is easily seen that $G(z)$ is (i) non-vanishing on $C_{1/\sqrt{\rho_0}}$ and (ii) has a bounded derivative on this contour (when $\rho_0 = 1$ then $1 - \phi_0(s_0(z))\phi_1(s_1(z))$ has a zero of multiplicity 1 at $z = 1$ and does not vanish for all $|z| = 1$ with $z \neq 1$). We conclude from (i) and (32) that $|\psi(z)| = 1$ on $C_{1/\sqrt{\rho_0}}$ and from (ii) that $G(z)$ satisfies a Hölder condition on $C_{1/\sqrt{\rho_0}}$ (apply **P3**) which in turn implies that $\psi(z)$ satisfies a Hölder condition on $C_{1/\sqrt{\rho_0}}$ (apply **P1**-**P2**).

Consider now the function

$$c_1(z) := \mathrm{i}\,\phi_0(s_0(z))\,\zeta_1(s_1(z))/\delta_1(s_1(z)).$$

Under assumptions **A1**-**A2** the function $c_1(z)$ has a bounded derivative on the contour $C_{1/\sqrt{\rho_0}}$. Therefore, $c_1(z)$ satisfies a Hölder condition on $C_{1/\sqrt{\rho_0}}$ (apply **P3**) and so does the function $|G(z)|\,c_1(z)$ as the product of two functions that satisfy a Hölder condition on $C_{1/\sqrt{\rho_0}}$ (apply **P2**). From the identity $c(z) = \mathrm{Re}\,(c_1(z))$ for all $z \in C_{1/\sqrt{\rho_0}}$, we conclude that $c(z)$ satisfies a Hölder condition on $C_{1/\sqrt{\rho_0}}$ (apply **P1**). $\square$

8. REFERENCES

[1] R. Agrawal, A. M. Makowski and P. Nain. On a reduced load equivalence for fluid queues under subexponentiality. *QUESTA*, Vol. 33, No. 1-3, pp. 5-41, 1999.

[2] S. C. Borst, O. J. Boxma and P. R. Jelenković. Asymptotic behavior of Generalized Processor Sharing with Long-Tailed Traffic Sources. In: *Proc. INFOCOM 2000*, Tel-Aviv, Israel, 912-921, 2000.

[3] O. J. Boxma. Fluid queues and regular variation. *Performance Evaluation*, Vol. 27&28, pp. 699-712, 1996.

[4] O. J. Boxma and V. Dumas. Fluid queues with long-tailed activity period distributions. *Computer Communications*, Vol. 21, pp. 1509-1529, 1998.

[5] O. J. Boxma and I. A. Kurkova. The M/M/1 queue in a heavy-tailed random environment. Statistica Neerlandica 54 (2000), 221–236.

[6] F. Brichet, J. W. Roberts, A. Simonian and D. Veitch, Heavy traffic analysis of a storage model with long range dependent on/off sources. *QUESTA*, Vol. 23, pp. 197-215, 1996.

[7] J. W. Cohen. Some results of regular variation for distributions in queueing and fluctuation theory. *J. Appl. Prob.*, Vol. 10, pp. 343-353, 1973.

[8] J. W. Cohen and O. J. Boxma. *Boundary Value Problems in Queueing System Analysis.* North-Holland, Amsterdam-New York, 1983.

[9] J. W. Cohen and D. G. Down. On the role of Rouché's theorem in queueing analysis. *QUESTA* 23 (1996), 281–291.

[10] M. Crovella and A. Bestavros. Self-similarity in world wide Web traffic: evidence and possible causes. *Proc. 1996 ACM Sigmetrics Int. Conf. on Measurements and Modeling of Comput. Syst.*, May 1996.

[11] N. G. Duffield. On the relevance of long-tailed durations for the statistical multiplexing of large aggregations. *Proc. of the 34th Annual Allerton Conf. on Communication, Control and Computing*, Oct. 2-4, 1996.

[12] G. Fayolle and R. Iasnogorodski. Two coupled processors: the reduction to a Riemann-Hilbert problem. Z. Wahrsch. Verw. Gebiete 47 (1979), 325–351.

[13] G. Fayolle, R. Iasnogorodski and V. Malyshev. *Random Walks in the Quarter-Plane. Algebraic Methods, Boundary Value Problems and Applications.* Applications of Mathematics, Springer-Verlag, Vol. 40, 1999.

[14] W. Fischer and K. Meier-Hellstern. The Markov-modulated Poisson process (MMPP) cookbook. *Performance Evaluation* 18 (1993), 149–171.

[15] B. A. Fuchs and B. V. Shabat. *Functions of a Complex Variable and Some of Their Applications.* Vol. 1. Pergamon Press, 1964.

[16] F. D. Gakhov. *Boundary Value Problems.* Pergamon Press, Oxford, 1966.

[17] D. Heath, S. Resnick and G. Samorodnitsky. Patterns of buffer overflow in a class of queues with long memory in the input stream. *Ann. Appl. Prob.*, Vol. 7, pp. 1021-1057, 1997.

[18] P. R. Jelenkovic and A. A. Lazar. Asymptotic results for multiplexing on-off sources with subexponential on periods. *Adv. Appl. Prob.*, Vol. 31, No. 2, Jun. 1999.

[19] W. Leland, M. Taqqu, W. Willinger and D. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans. on Networking*, Vol. 2, pp. 1-15, 1994.

[20] Z. Liu, P. Nain, D. Towsley and Z.-L. Zhang. Asymptotic behavior of a multiplexer fed by a long-range dependent process. *J. Appl. Prob.*, Vol. 36, No. 1, pp. 105-118, Mar. 1999.

[21] A. I. Markushevich, *Theory of Functions of a Complex Variable*. Chelsea Publishing Company, New York, 2nd Ed., 1985.

[22] N. I. Muskhelishvili. *Singular Integral Equations*. P. Noordhoff, 1953.

[23] M. F. Neuts. The single-server queue with Poisson input and semi-Markov service times. *J. Appl. Prob.*, Vol. 3, pp. 202-230, 1966.

[24] I. Norros. A storage model with self-similar input. *QUESTA*, Vol. 16, pp. 387-396, 1994.

[25] I. Norros. On the use of fractional Brownian motion in the theory of connectionless networks. *IEEE J. on Sel. Areas in Comm.*, Vol 13, No. 6, pp. 953-962, Aug. 1995.

[26] A. G. Pakes. On the tails distribution of waiting time distributions. *J. Appl. Prob*, Vol. 12, pp. 555-564, 1975.

[27] M. Parulekar. *Buffer Engineering for $M|G|\infty$ Traffic Models: Analysis and Simulations*, Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, Dec. 1999.

[28] M. Parulekar and A. M. Makowski. Tail probabilities for M/G/∞ input processes (I): Preliminary asymptotics. *QUESTA*, Vol. 27, pp. 271-296, 1997.

[29] V. Paxson and S. Floyd. Wide-area traffic: the failure of Poisson modeling. *Proc. of the ACM Sigcomm'94 Conf.*, London, UK, pp. 257-268, 1994.

[30] Queues with Heavy-Tailed Distributions. Special Issue of *QUESTA*, Vol. 33, No. 1-3, 1999.

[31] G. J. K. Regterschot and J. H. A. de Smit. The queue M/G/1 with Markov modulated arrivals and services. *Math. Op. Res.*, Vol. 11, No. 3, pp. 465-483, 1986.

[32] B. Tsybakov and N. D. Georganas, On self-similar traffic in ATM queues: Definitions, overflow probability and cell delay distribution. *IEEE/ACM Trans. on Networking*, Vol. 5, No.3, pp. 397-409, Jun. 1997.

[33] N. Veraverbeke. Asymptotic behavior of Wiener-Hopf factors of a random walk. *Stoch. Proc. and their Applications*, Vol. 5, pp. 27-37, 1977.

[34] A. P. Zwart. A fluid queue with a finite buffer and subexponential input. *Adv. Appl. Prob.*, Vol. 32, pp. 221–243, 2000.