Subexponential Asymptotics of a Network Multiplexer

Predrag R. Jelenković and Aurel A. Lazar
Department of Electrical Engineering
and
Center for Telecommunications Research
Columbia University, New York, NY 10027-6699
e-mail: {aurel, predrag}@ctr.columbia.edu

Abstract

For a Markov-modulated random walk with negative drift and long-tailed right tail we prove that the ascending ladder height matrix distribution is asymptotically proportional to a long-tailed distribution. This result enable us to generalize a recent result on subexponential asymptotics of a Markov-modulated M/G/1 queue to subexponential asymptotics of a Markov-modulated G/G/1 queue.

For a class of processes constructed by embedding a Markov chain into a subexponential renewal process we prove that the autocorrelation function has a subexponential tail. Furthermore, we prove that when a fluid flow queue is fed by these processes the queue length distribution is asymptotically proportional to its autocorrelation function.

1 Introduction

Under the variety of assumptions of Cramér type (exponentially bounded marginals and autocorrelation function) many published results have shown that the queue length distribution of a network multiplexer has exponential asymptotics, i.e., $\mathbb{P}[\mathrm{Q} > \mathrm{x}] \sim \alpha \mathrm{e}^{-\theta^* \mathrm{x}}$ as $\mathrm{x} \to \infty$. Some authors have argued that an even simpler approximation holds $\mathbb{P}[\mathrm{Q} > \mathrm{x}] \sim \mathrm{e}^{-\theta^* \mathrm{x}}$. This has led to the development of the so called effective bandwidth based admission control (see [6, 10, 13, 12, 17]). However, this approach may often lead to poor approximations [8, 14], which may result in a significant underutilization of network resources. The shortcoming of the single exponential approximation often manifests itself with arrival processes that span over multiple time scales. This has been independently shown in [14, 18]. A recursive numerical algorithm for computing the queue length distribution with multiple time scale arrivals can be found in [16].

The main motivation for this work are the results presented in [11] which show that the (marginal) distribution function and the autocorrelation function of the arrival processes that appear in communications networks may have a long (subexponential) tail. For such processes the Cramér type conditions are not satisfied.

To get some feeling about the behavior of the queue when the arrival process has a long tailed distribution let us examine the following example. (All the examples in this paper are calculated using the z-transform technique and Mathematica 2.2.)

Example 1 Consider a discrete time queue with a service rate of one packet per slot ($c = 1$) and an arrival process characterized by a sequence A_n of i.i.d. random variables distributed as $\mathbb{P}[\mathrm{A}_0 = 0] = 0.2, \mathbb{P}[\mathrm{A}_0 = \mathrm{i}] = \mathrm{d}/\mathrm{i}^6, 1 \leq \mathrm{i} \leq 150$, $d = 0.77151$. Thus, this source (arrival process) has a truncated heavy tail with peak rate of 150 packets. Since this process is bounded from above its cumulant function $\mathbb{E}\mathrm{e}^{\theta \mathrm{A}_0}$ exists for all $\theta > 0$ and, therefore, the queue tail is asymptotically exponential. However, the range of the exponential asymptotic may be far out of the relevant range of probabilities. On the right-hand side of figure 1 we can see that the exponential asymptotics starts to work for very small probabilities (roughly smaller than 10^{-40}). However, in the relevant range of probabilities ($10^{-4} - 10^{-10}$) we see, on the left-hand side of figure 1, that the exponential approximation fails. In this region, queue length probabilities have a functional form approximately proportional to the integrated tail of A_n ($1/i^5$).

In Proceedings of the Thirty-Third Annual Allerton Conference on Communication, Control, and Computing, pp. 746-755, Urbana-Champaign, Illinois, October 4-6, 1995.

Thus, we can see that *even in the case of bounded heavy tailed arrivals*, for which the queue length asymptotics is eventually exponential, the relevant part of the queue length distribution may be subexponential.

In this paper we examine the queue length distribution when the Cramér type conditions are replaced by subexponential assumptions.

The paper is organized as follows. In section 2 we give the basic definitions and results on subexponential distributions. At the end of that section (subsection 2.2) the classical result on the subexponential $GI/GI/1$ queue asymptotics is presented in Theorem 1. This result has been recently generalized to the Markov-modulated $M/G/1$ queue [4]. Our further generalization of this result to the Markov-modulated $G/G/1$ queue is presented in section 3. In the rest of the paper (section 4) we consider a class of processes that have subexponentially correlated arrivals. This processes are obtained by embedding Markov chains in a stationary subexponential renewal process. When this processes are fed into a fluid flow queue the queue length distribution is *asymptotically proportional to the autocovariance (autocorrelation) function of the arrival process.* The paper is concluded in section 5 with a brief application of these results to broadband network admission control.

2 Subexponential Distributions

Definition 1 *A distribution function F with $F(0-) = 0$ is called* long-tailed *($F \in \mathcal{L}$) if*

$$\lim_{x\to\infty} \frac{1-F(x-y)}{1-F(x)} = 1, \quad y \in \mathbb{R}. \tag{1}$$

Definition 2 *A distribution function F with $F(0-) = 0$ is called* subexponential *($F \in \mathcal{S}$) if*

$$\lim_{x\to\infty} \frac{1-F^{*2}(x)}{1-F(x)} = 2, \tag{2}$$

where F^{*2} denotes the 2-nd convolution of F with itself.

The class of subexponential distributions was first introduced by Chistakov [7]. The definition is motivated by the simplification of the asymptotic analysis of the convolution tails. Some examples of distribution functions in $\mathcal{S}$ are:

(I) the Pareto family

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

$x > \beta > 0, \alpha > 0$.

(II) the lognormal distribution

$$F(x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right), \quad \mu \in \mathbb{R}, \sigma > 0,$$

where Φ is the standard normal distribution.

(III) Weibull distribution

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

for $0 < \beta < 1, \alpha > 0$.

(IV)

$$F(x) = e^{-x(\log x)^{-a}},$$

for $a > 0$.

(V) Benktander Type I distribution

$$F(x) = 1 - cx^{-a-1}x^{-b\log x}(a + 2b\log x),$$

$a > 0, b > 0$, and c appropriately chosen.

(V) Benktander Type II distribution

$$F(x) = 1 - cax^{-(1-b)}\exp\{-(a/b)x^{b}\},$$

$a > 0, 0 < b < 1$, and c appropriately chosen.

2.1 Basic Results

In what follows we will state an important result from the literature on subexponential distributions. The general relation between $\mathcal{S}$ and $\mathcal{L}$ is the following.

Lemma 1 *(Athrey and Ney, [5]) $\mathcal{S} \subset \mathcal{L}$.*

Lemma 2 *If $F \in \mathcal{L}$ then $(1 - F(x))e^{\alpha x} \to \infty$ as $x \to \infty$, for all $\alpha > 0$.*

Note: Lemma 2 clearly shows that for long tailed distributions Cramér type conditions are not satisfied.

To simplify the notation let us denote by $\bar{F}(x) \stackrel{def}{=} 1 - F(x)$. The following lemma, used in the proof of theorem 4, is from [4].

Lemma 3 *Let $G = \{G_{ij}\}$ be a matrix of nonnegative measures such that $\|G\| \stackrel{def}{=} G(0,\infty)$ is substochastic (the spectral radius is < 1). If there exists some probability distribution $G \in \mathcal{S}$ such that $\bar{G}_{ij}(x) \sim l_{ij}\bar{G}(x)$ as $x \to \infty$ for some matrix $L = \{l_{ij}\}, 0 < l_{ij} < \infty$, then*

$$\sum_{n=0}^{\infty} \overline{G^{*n}}(x)(I - \|G\|) \sim (I - \|G\|)^{-1}L\bar{G}(x).$$

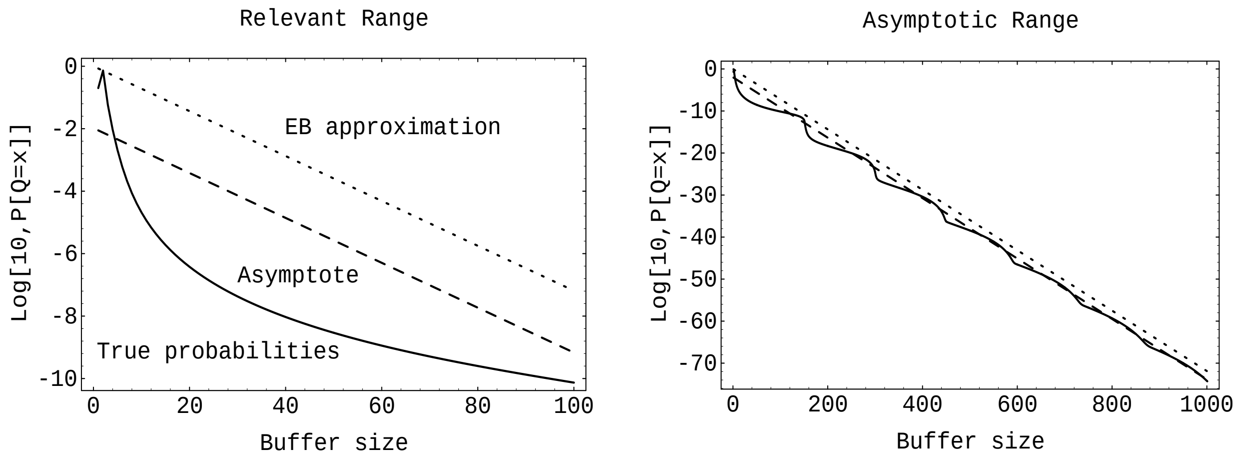

Figure 1: Illustration for Example 1.

2.2 The GI/GI/1 queue with subexponential arrivals

Let $\{A_t, C_t, t \geq 0\}$ be a sequence of random variables (on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$). Then, for any initial random variable Q_0 the following (Lindley's) equation

$$Q_{t+1} = (Q_t + A_t - C_t)^+ \tag{3}$$

defines the queue length process $\{Q_t\}$; throughout we deal with a stationary version of this process. The dynamics of a broadband network multiplexer is defined by the previous recursion. Q_t represents the workload at the end of the time slot t, A_t, represent the amount of traffic (packets) that arrives at the multiplexer, and C_t represents the amount of traffic that is served during the slot t (we do not assume that A_t, C_t are integer valued). Note that the recursion (3) represents the workload of the $G/G/1$ queue with C_t being interpreted as the customer interarrival time, A_t as the customer service requirements, and Q_t as the queue workload.

Let $X_t = A_t - C_t, t \geq 0$, and let X_t be a sequence of i.i.d. random variables with a distribution F, and A_t independent of C_t. Further, denote its *integrated tail* as $\hat{F}(x) \stackrel{def}{=} \int_x^\infty [1 - F(t)]dt$. Let's define $F_1(x) = m^{-1}(1 - \hat{F}(x))$, where $m = \hat{F}(0)$. Similarly, in the rest of the paper for any d.f. G, we define its corresponding $\hat{G}(x)$ and $G_1(x)$. Then the following result on the $GI/GI/1$ queue was proven by Veraverbeke [19]. Let B be a d.f. of A_t.

Theorem 1 **(i)** $F_1 \in \mathcal{S} \iff B_1 \in \mathcal{S}$ *and* $\lim_{x\to\infty} \frac{\hat{F}(x)}{\hat{B}(x)} = 1$

(ii) $B_1 \in \mathcal{S} \iff \mathbb{P}[\mathrm{Q_t \leq x}] \in \mathcal{S}$ *and each implies*

$$\lim_{x\to\infty} \frac{\mathbb{P}[\mathrm{Q_t > x}]}{\hat{B}(x)} = \frac{1}{\alpha - \beta}.$$

Some recent results on long tailed asymptotics of a $GI/GI/1$ are given in [1, 20]. (Also, in [1] further motivation is given for the application of long tailed distributions in communication networks.)

The subexponential asymptotic behavior of the queue length distribution of the Markov-modulated $M/G/1$ queue was derived in [4]. Our extension of this result to the Markov-modulated $G/G/1$ queue is presented in the next section.

3 Markov-modulated $G/G/1$ Queue with Subexponential Arrivals

Consider a Markov-modulated random walk with a negative drift and long-tailed right tail. We prove that the ascending ladder height matrix distribution is asymptotically proportional to a long-tailed distribution. Using this result we show that the queue length distribution of a Markov-modulated $G/G/1$ queue with subexponential arrivals is proportional to the integrated tail of the arrival distribution. Preliminaries on the generalization of the Ladder heights approach to the Markov-modulated random walk recently obtained in [2, 3] are presented in the following subsection (we borrow the notation from the same papers).

3.1 Markov-modulated Random Walk and Ladder Heights

Let $\{J_n\}$ be an irreducible aperiodic Markov chain with a finite state space E (say with N elements) and let $\{X_n\}$ be a sequence of real valued random variables. A Markov process $\{(J_n, X_n)\}$ on $E \times \mathbb{R}$ whose transition distribution depends only on the first coordinate is called a Markov-modulated random walk (MMRW). Let $\{(\tilde{J}_n, \tilde{X}_n)\}$ denote the associated reversed process. This process is determined by a set of transition measures $F_{ij}(A) = \mathbb{P}[\mathrm{J_0 = j, X_1 \in A | J_1 = i}]$, and let $F = \{F_{ij}\}$ be the corresponding transition matrix measure.

Further, define $\tilde{S}_0 = 0, \tilde{S}_n = \sum_{i=1}^n \tilde{X}_i$,

$$\tau_+ = \inf\{n > 0 : \tilde{S}_n > 0\},$$
$$G_+(i,j;A) = \mathrm{I\!P}_{\mathrm{i}}[\tilde{\mathrm{J}}_{\tau_+} = \mathrm{j}, \tilde{\mathrm{S}}_{\tau_+} \in \mathrm{A}, \tau_+ < \infty],$$
$$\|G_+(i,j)\| = G_+(i,j;(0,\infty)),$$
$$G_+(A) = \{G_+(i,j;A)\}_{i,j\in E},$$
$$\|G_+\| = \{\|G_+\|\}_{i,j\in E}.$$

The convolution of the matrix measure G_+ is naturally extended to

$$G_+^{*2}(i,j) = \sum_{k\in E} G_+(i,j) * G_+(k,j),$$
$$G_+^{*2} = \{G_+^{*2}(i,j)\}_{i,j\in E};$$

higher convolution powers are similarly defined.

Then in [3] the following extension of the Pollaczek-Khinchine identity is provided for $\tilde{M} = \sup_{n\geq 0} \tilde{S}_n$.

Theorem 2 $\mathrm{I\!P}_{\mathrm{i}}[\tilde{\mathrm{M}} \in \mathrm{A}]$ *is the ith component of the vector*

$$\sum_{n=0}^{\infty} G_+^{*n}(A)(I - \|G_+\|)e,$$

where e is the column vector of ones and I is the identity matrix.

Note: A well known relation between the supremum of the reversed random walk and the stationary queue length distribution is $\mathrm{I\!P}_{\mathrm{i}}[\mathrm{Q} \in \mathrm{A}] = \mathrm{I\!P}_{\mathrm{i}}[\tilde{\mathrm{M}} \in \mathrm{A}]$.

Let $\tau_- = \inf\{n \geq 1 : S_n \leq 0\}$, and define

$$G_-(i,j;A) = \mathrm{I\!P}_{\mathrm{i}}[\mathrm{S}_{\tau_-} \in \mathrm{A}, \mathrm{J}_{\tau_-} = \mathrm{j}, \tau_- < \infty],$$
$${}^{\#}G_-(i,j) = \frac{\pi_j}{\pi_i} G_-(j,i);$$

(note that in this definition S is not the reversed process). Then, the following Wiener-Hopf identity holds

$$I - F(A) = (I - {}^{\#}G_-) * (I - G_+)(A), \qquad (4)$$

where F is a matrix measure induced by the random walk, and A is any real Borel set. Note that this equation may be difficult to solve. Discussion and references to the computational aspects of the Wiener-Hopf factorization are given in [15].

3.2 Long Tailed Asymptotics of Signed Measures

In this section we prove a few general results on the long tailed asymptotics of signed measures. Combination of these results essentially will give a proof of our main results presented in the following subsection. (These results may also be of independent interest.) Let $(\mathrm{I\!R}, \mathcal{B}(\mathrm{I\!R}))$ be a Borel σ-algebra on $\mathrm{I\!R}$.

Lemma 4 *Let μ, μ_- be two* finite *(signed) measures on $(\mathrm{I\!R}, \mathcal{B}(\mathrm{I\!R}))$, such that $\lim_{x\to\infty} \mu([x,\infty))/\bar{Y}(x) = c$, $Y(x) \in \mathcal{L}$, $|c| < \infty$, and μ_- has support on $(-\infty, 0]$. Then, $\nu \overset{def}{=} \mu_- * \mu$ satisfies*

$$\lim_{x\to\infty} \frac{\nu([x,\infty))}{\bar{Y}(x)} = c\mu_-((-\infty,0]).$$

Proof: Given in [15].

Lemma 5 *Let μ, μ_-, μ_+, be* finite *(possibly signed) measures on $(\mathrm{I\!R}, \mathcal{B}(\mathrm{I\!R}))$; with μ_- having a support on $(-\infty, 0]$, $|\mu_-((-\infty,0])| > 0$; μ_+ has a support on $[0,\infty)$ and is strictly positive on $[K,\infty)$, $K > 0$, and $\lim_{x\to\infty} \mu([x,\infty))/\bar{Y}(x) = c$, $Y(x) \in \mathcal{L}$, $|c| < \infty$. If $\mu = \mu_- * \mu_+$, then*

$$\lim_{x\to\infty} \frac{\mu_+([x,\infty))}{\bar{Y}(x)} = \frac{c}{\mu_-((-\infty,0])}. \qquad (5)$$

Proof: Given in [15]. ◇

Proposition 1 *Let $\mu = \mu_- * \mu_+$, where measures μ, μ_-, μ_+ satisfy the conditions from the previous lemma, and, in addition, $\mu_-((-\infty,0]) = 0$, and $0 < \left|\int_{(-\infty,0]} u d\mu_-(u)\right| < \infty$. Then,*

$$\lim_{x\to\infty} \frac{\mu_+([x,\infty))}{\hat{Y}(x)} = \frac{c}{\int_{(-\infty,0]} u d\mu_-(u)},$$

(recall $\hat{Y}(x) = \int_{[x,\infty)} \bar{Y}(u) du$).

Proof: Let $\mu^1([x,\infty)) \overset{def}{=} \int_x^\infty \mu([u,\infty))du$, and $\mu_-^1([z,0]) \overset{def}{=} \int_{[z,0]} \mu_-([u,0])du$. Observe that $\mu^1([x,\infty)) \sim c\hat{Y}(x)$ as $x \to \infty$. Also, from the assumptions it follow that μ_-^1 defines a finite measure on $(-\infty,0]$, since $\mu_-^1((-\infty,0]) = \int_{-\infty}^0 -\mu_-((-\infty,u))du = \int_{(-\infty,0]} u\mu_-(du)$. Then, by applying Fubini's theorem (see [21], pp. 180), we get

$$\begin{aligned}
&\mu^1([y,\infty)) \\
&\quad = \int_y^\infty du \int_{[0,\infty)} \mu_-([u-x,0])\mu_+(dx) \\
&\quad = \int_{[0,\infty)} \mu_+(dx) \int_y^x \mu_-([u-x,0])du \\
&\quad = \int_{[0,\infty)} \mu_-^1([y-x,0])\mu_+(dx).
\end{aligned}$$

So, $\mu_1([y,\infty))$ is obtained by convolution of finite measures μ_+, μ_-^1. Therefore, by applying lemma 5, the conclusion of the proposition follows (recall that $\mu_-^1((-\infty,0]) = \int_{(-\infty,0]} u\mu_-(du)$).

◇

3.3 Subexponential Queue Asymptotics

We will now proceed to prove our main result of this section. In order to state the result we need to introduce some additional notation. Let $H(x) = \{H_{ij}\}$ be a matrix composed of distribution functions, and its Fourier transform defined as $\tilde{H}(\omega) = \{\tilde{H}_{ij}(\omega)\}$, $\tilde{H}_{ij}(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} dH_{ij}(x)$. We will use the symbol $\mathcal{F}^{-1}(\cdot)$ to denote the operation of taking the inverse Fourier transform. Note that there is a one-to-one correspondence between the distribution functions on $\mathbb{R}$ and its Fourier transforms (see [21], section 8.3). The Wiener-Hopf factorization can be written as

$$(I - \tilde{F}(\omega)) = (I - {}^{\#}\tilde{G}_-(\omega))(I - \tilde{G}_+(\omega)). \quad (6)$$

Observe also that $\tilde{H}_{ij}(0) = H((-\infty, \infty))$, and that $-i\tilde{H}'_{ij}(0) = \int_{-\infty}^{\infty} x dH_{ij}(x)$; assume that the first moments of H_{ij} are finite. For any matrix A let us define the *adjoint* matrix $adj(A)$, $adj(A)_{ij} = (-1)^{i+j}\det(A^{ij})$, where A^{ij} denotes the matrix obtained by deleting the ith row and jth column from A. If A is invertible then $A^{-1} = (\det(A))^{-1} adj(A)$. Assume $\mathbb{E}\mathrm{X_n} < 0$ (negative drift) and $\int_{-\infty}^{\infty} |x| F_{ij}(dx) < \infty$.

Theorem 3 *Let $\bar{F}(x)/\bar{Y}(x) \to W$, $W = \{W_{ij}\}$, $W_{ij} < \infty$, $Y(x) \in \mathcal{L}$. Then,*

$$\frac{\bar{G}_+(x)}{\hat{Y}(x)} \to \frac{adj(I - {}^{\#}\tilde{G}_-)(0)W}{-i\det(I - {}^{\#}\tilde{G}_-)'(0)}, \quad (7)$$

as $x \to \infty$.

Proof: Fist let us observe that $\det(I - \tilde{F})(\omega)$ has a zero of order one for $\omega = 0$. $\omega = 0$ is a zero since $\det(I - \tilde{F})(0) = \det(I - P) = 0$, as P is a stochastic transition matrix; that this zero is of order one follows from $\mathbb{E}\mathrm{X_n} < 0$. Furthermore, since $\|G_+\|$ is substochastic (see [2], proposition 4.2), we have that $|\det(I - \tilde{G}_+)(0)| > 0$, which implies (by equation (6)) that $\omega = 0$ is also a zero of order one for the $\det(I - {}^{\#}\tilde{G}_-)(\omega)$, implying $0 < |\det(I - {}^{\#}\tilde{G}_-)'(0)| < \infty$; finiteness follows from $\det(I - \tilde{F})'(0) = \det(I - \tilde{G}_+)(0)\det(I - {}^{\#}\tilde{G}_-)'(0)$.

Let us define the measure $\mu_- \stackrel{def}{=} \mathcal{F}^{-1}(\det(I - {}^{\#}\tilde{G}_-(\omega)))$. Note that this measure has support on $(-\infty, 0]$ and $\int_{(-\infty,0]} u d\mu_-(u) = -i\det(I - {}^{\#}\tilde{G}_-)'(0) \neq 0$ (finite). Also, equation (6) can be written as

$$\begin{aligned} & adj(I - {}^{\#}\tilde{G}_-(\omega))(\tilde{F}(\omega) - I) \\ & \quad = (\tilde{G}_+(\omega) - I)\det(I - {}^{\#}\tilde{G}_-(\omega)), \end{aligned}$$

or componentwise

$$\begin{aligned} & (\tilde{G}_+(\omega) - I)_{ij}\det(I - {}^{\#}\tilde{G}_-(\omega)) \\ & \quad = \sum_{k=1}^{n} adj(I - {}^{\#}\tilde{G}_-(\omega))_{ik}(\tilde{F}(\omega) - I)_{kj}. \end{aligned}$$

If $\mu_{ij} \stackrel{def}{=} \mathcal{F}(adj(I - {}^{\#}\tilde{G}_-(\omega))_{ik}(\tilde{F}(\omega) - I)_{kj})$, then, by Lemma 4, $\mu([x, \infty)) \sim \sum_{k=1}^{n} adj(I - {}^{\#}\tilde{G}_-(0))_{ik} W_{kj}\bar{Y}(x)$ as $x \to \infty$. If $\mu_{+ij} \stackrel{def}{=} \mathcal{F}^{-1}(\tilde{G}_+(\omega) - I)_{ij}$, then $\mu_{ij} = \mu_{+ij} * \mu_-$, where $\mu_{ij}, \mu_{+ij}, \mu_-$, satisfy the conditions of proposition 1. Thus, by the same proposition,

$$\frac{\bar{G}_{+ij}(x)}{\hat{Y}(x)} \to \frac{\sum_{k=1}^{n} adj(I - {}^{\#}\tilde{G}_-(0))_{ik} W_{kj}}{-i\det(I - {}^{\#}\tilde{G}_-)'(0)},$$

as $x \to \infty$, which is tantamount to equation (7). This finishes the proof of the theorem. ◇

Now by combining the previous theorem and the extension of lemma 3 from [4] we will prove the following theorem on subexponential asymptotics for the Markov-modulated $G/G/1$ queue. Let (J_n, A_n) and (J_n, C_n) be two MMRW such that (J_n, A_n) and (J_n, C_n) are conditionally independent given J_{n-1}, J_n. $\{A_n\}$ and $\{C_n\}$ are arrival and service processes, respectively. Let B be a reverse transition matrix measure of the MMRW (J_n, A_n), such that for each $i, j \in E$ $\bar{B}_{ij}(x) \sim W_{ij}\bar{Y}(x)$ as $x \to \infty$ and $W_{ij} > 0$.

Theorem 4 *If $Y \in \mathcal{L}, Y_1 \in \mathcal{S}$ then there exists a constant $c > 0$ such that.*

$$\frac{\mathbb{P}[\mathrm{Q_t > x}]}{\int_x^{\infty} \mathbb{P}[\mathrm{A_t > u}]\mathrm{du}} \to c \text{ as } x \to \infty.$$

Furthermore, when the Wiener-Hopf factorization is explicitly solvable the constant c is explicitly computable.

Proof: Componentwise the asymptotic proportionality of the matrix distributions $\hat{F}(x)$ and $\hat{B}(x)$ follows from Theorem 1 (i) as $x \to \infty$. Then, combining Theorems 2, 3, and Lemma 3 the conclusion of the theorem follows. ◇

Remark: The assumption $W_{ij} > 0$ for all i, j can be removed. The queue distribution is subexponential as long as at least one measure B_{ij} is subexponential. For a precise statement of this result see [15].

Illustration of this remark and the preceding theorem is given in the following numerical example.

Example 2 Consider a constant server queue with $C_t = 1$ and two state (say $\{0, 1\}$) Markov-modulated arrivals (source). The transition probabilities for the modulating Markov chain are $p_{01} = 1/3, p_{10} = 3/4$. When in state 0 the source is producing zero arrivals, and when in state 1 the source is producing (independent of the previous state) arrivals according to the distribution $\mathbb{P}[\mathrm{A_t = 0|J_t = 1}] = 0.327144, \mathbb{P}[\mathrm{A_t = 1|J_t = 1}] = 0$, and $\mathbb{P}[\mathrm{A_t = i|J_t = 1}] = \mathrm{w/i^5}, \mathrm{w} = 18.220859, 2 \leq \mathrm{i} \leq 350$. (Note that these are bounded arrivals.) Thus, according to the previous theorem (and the remark after it) the queue length distribution is proportional to $1/i^4$. The comparison between the true probabilities and the approximation $c/i^4, c = 2.617872$ is given in figure 2.

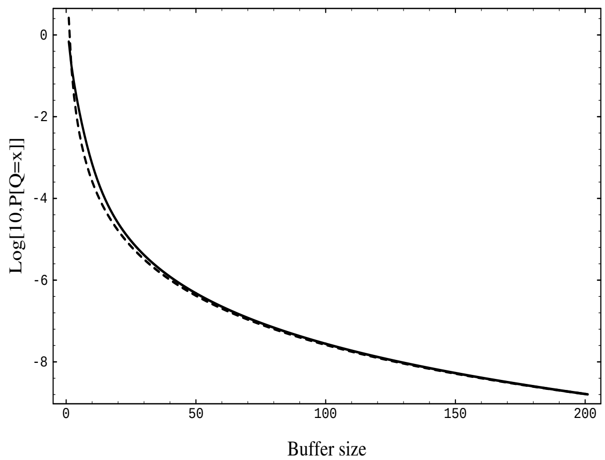

Figure 2: Graph of $\log_{10} \mathbb{P}[\mathrm{Q} = \mathrm{i}]$ versus buffer size i from Example 2; solid line represents the true probabilities, and dashed line represents the approximation $2.617872/i^{-4}$.

4 Asymptotics of a Fluid Flow Queue with Subexponentially Correlated Arrivals

In this section we construct a class of processes for which we show that its autocorrelation function (ac.f.) is subexponential. Furthermore, when these processes are fed to a fluid flow queue, we prove the asymptotic proportionality of the queue length distribution with the arrival process ac.f.. Throughout this section we assume a continuous time model (of course all the results are valid for discrete time also).

4.1 Stationary Subexponentially Correlated Arrivals

Consider a point process $T = \{T_0 \leq 0, T_n, n \geq 1\}$ such that $T_n - T_{n-1}, n \geq 1$ are i.i.d. with subexponential distribution function F. Further, let $J_n, n \geq 0$ be an irreducible aperiodic Markov Chain with finite state space $\{1, \ldots, K\}$, transition matrix $\{P_{ij}\}$, and stationary probability distribution $\pi_i, 1 \leq i \leq K$. In order to make this process stationary (see [9], section 9.3), we choose the residual time at zero until the first jump to be distributed as an integrated tail of F, i.e., $F_1(t) = \mathbb{P}[\mathrm{T}_1 \leq \mathrm{t}] = \mathrm{m}^{-1} \int_{0,\mathrm{t}} \bar{\mathrm{F}}(\mathrm{u})\mathrm{du}$.

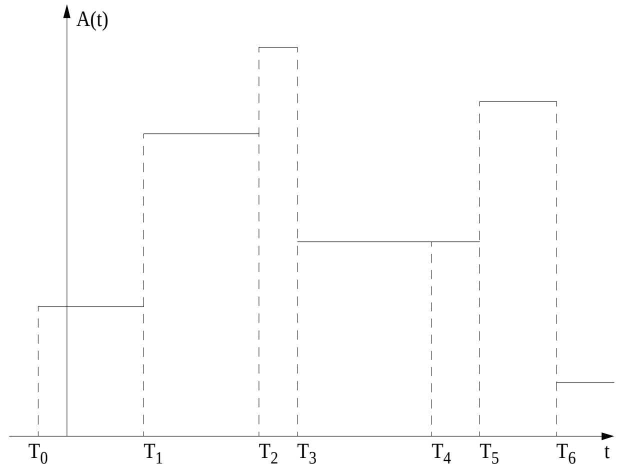

Figure 3: A possible realization of a Markov chain embedded into a renewal process.

Now we construct the following process:

$$A_t = J_n \quad \text{for} \quad T_{n-1} \leq t < T_n. \tag{8}$$

called a Markov Chain Embedded in a Stationary Subexponential Renewal Process (MCESSR). A typical sample path of this process is given in figure 3. It is well known that under fairly general conditions a Markov chain converges exponentially fast to the steady state distribution. However, the process that we have constructed, because of the subexponentially distributed sojourn times, converges with subexponential speed to its steady state. This is stated in the following lemma.

Proposition 2 *If $F, F_1 \in \mathcal{S}$, then*

$$(\mathbb{P}_{\mathrm{i}}[\mathrm{A_t} = \mathrm{j}] - \pi_{\mathrm{j}})\bar{\mathrm{F}}_1(\mathrm{t})^{-1} \to (\delta_{\mathrm{ij}} - \pi_{\mathrm{j}}),$$

as $t \to \infty$, and $\delta_{ij} = 1$ if $i = j$ and zero otherwise.

Proof: Given in [15] ◇

We will illustrate this lemma by the following example.

Example 3 Let F be a discrete distribution function with support $[1, 1000]$, $\mathbb{P}[\mathrm{T}_2 - \mathrm{T}_1 = 1] = 0.186532$, and $\mathbb{P}[\mathrm{T}_2 - \mathrm{T}_1 = \mathrm{i}] = \mathrm{w}/\mathrm{i}^5, \mathrm{w} = 22.028625$, $2 \leq i \leq 1000$; chose a two state Markov chain with transition probabilities $p_{01} = 1/3$ and $p_{10} = 3/4$. Then, the functions $(d_{i,1}(t) \stackrel{def}{=} (\mathbb{P}_{\mathrm{i}}[\mathrm{A_t} = 1] - \pi_1)(\bar{\mathrm{F}}_1(\mathrm{t})(\delta_{\mathrm{i}1} - \pi_1))^{-1}$, $i = 0, 1$, converge to one as $t \to \infty$, with subexponential rate. This can be clearly seen in figure 4.

Now it is easy to prove that the covariance function $\rho(t) \stackrel{def}{=} (\mathbb{E}\mathrm{A}_0\mathrm{A_t} - (\mathbb{E}\mathrm{A}_0))^2$ of the MCESSR process satisfies the following asymptotic relation. Let $\mathrm{Var}(J_0) > 0$ be the variance of J_0.

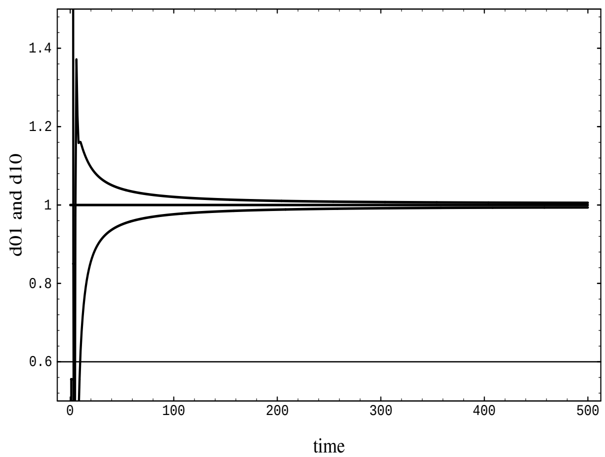

Figure 4: Functions $d_{i,1}(t) \stackrel{def}{=} (\mathrm{IP_i[A_t = 1]} - \pi_1)(\mathrm{\bar{F}_1(t)}(\delta_{\mathrm{i}1} - \pi_1))^{-1}$, $i = 0, 1$. The graph shows that $d_{i,1}(t) \to 1$ as $t \to 1$.

Theorem 5 *If $F, F_1 \in \mathcal{S}$, then*

$$\rho(t) \ \to \ \mathrm{Var}(J_0)\bar{F}_1(t),$$

as $t \to \infty$.

Remark: A nontrivial extension of this theorem and the previous lemma for the case of subexponential semi-Markov processes is given in [15].
Proof: By applying the previous lemma and after some simple algebraic manipulations we get

$$\begin{aligned}
&\left(\mathrm{IEA_0A_t} - (\mathrm{IEA_0})^2\right)\bar{F}_1(t)^{-1} \\
&\quad = \sum_{i,j} a_i a_j (\pi_i \mathrm{IP_i[A_t = j]} - \pi_\mathrm{i}\pi_\mathrm{j})\mathrm{\bar{F}_1(t)}^{-1} \\
&\quad \sim \sum_{i,j} a_i a_j \pi_i (\delta_{ij} - \pi_j) \\
&\quad = \sum_i \pi_i a_i^2 - \sum_{i,j} \pi_i \pi_j a_i a_j = \mathrm{Var}(J_0),
\end{aligned}$$

as $t \to \infty$. This completes the proof of the theorem. ◇

4.2 Subexponential Asymptotics of a Fluid Flow Queue

Now, we investigate the queue length distribution of a fluid queue fed with a MCESSR process. We assume that both the arrival process A_t and the server process C_t are MCESSR processes embedded into the same renewal process $\{T_n\}$, such that when the Markov chain J_n is equal to i, $A_t = a_i \geq 0$ and $C_t = c_i \geq 0$, $T_n \leq t < T_{n+1}$. Intuitively the pair A_t, C_t represents a fluid queueing model in which $A_t = a_i$ means that the flow is arriving to the queue with a rate a_i and $C_t = c_i$ means that the flow is coming from the queue with a rate c_i. We will calculate the queue length distribution at the jump times (Palm probability); $Q_n \equiv Q(T_n)$ satisfies the recursion

$$Q_{n+1} = (Q_n + x_{J_n}\Delta T_n)^+,$$

where $x_i = a_i - c_i$, and $\Delta T_n = T_{n+1} - T_n$.

We are now ready to state the following result on the asymptotic proportionality of the queue length distribution and its autocorrelation function. To avoid trivialities we assume that at least for one i, $x_i > 0$; also $Var(J_0) > 0$.

Theorem 6 *Let the stability condition $\mathrm{IEx_{J_n}} < 0$ be satisfied, and for all $x_i > 0$, $\mathrm{IP[\Delta T_n > t/x_i]/\bar{Y}(t) \to w_i}$, as $t \to \infty$, with at least one $w_i > 0$, and $Y, Y_1 \in \mathcal{S}$. Then, there exist a positive constant r such that*

$$\mathrm{IP[Q > t] \to r}\ \rho(\mathrm{t}),$$

as $t \to \infty$.

Proof: Follows by straightforward combination of theorems 4, 5; (more precisely an extension of theorem 4 mentioned in the remark after it). ◇

Remarks: **(i)** If the distribution function of ΔT_n belongs to the Pareto family the assumption $\mathrm{IP[\Delta T_n > t/x_i] \sim w_i\bar{Y}(t), w_i > 0}$ will be satisfied for all $x_i > 0$. **(ii)** Taking into account the remark after the theorem 5 we see that this relationship holds for the general class of *subexponential semi-Markov* arrivals. To the best of our knowledge this is the first rigorous result of this kind and generality.

5 Conclusion

For a Markov-modulated random walk with a negative drift and long-tailed right tail we have shown that the ascending ladder height matrix distribution is asymptotically proportional to a long-tailed distribution. This result enabled us to generalize a recent result on subexponential asymptotics of a Markov-modulated M/G/1 queue to subexponential asymptotics of a Markov-modulated G/G/1 queue. If the matrix Wiener-Hopf factorization is explicitly solvable then the asymptotic constant of proportionality is explicitly computable.

We also constructed a general class of processes, termed MCESSR, for which the covariance (autocorrelation) function has a subexponential tail. Furthermore, when this processes are fed into a fluid flow queue, the queue length distribution was proven to be asymptotically *proportional to its autocorrelation function.* In short

- (subexp. marginal d.f. + exp. ac.f.) ⇒ (queue distribution is determined by marginal d.f.),

- (bounded (exp) marginal d.f. + subexp. ac.f.) $\Rightarrow$ (queue distribution is determined by ac. f.).

When these type of conditions are met in practice, the above results may lead to an efficient admission control policy at network multiplexers. Admission controllers may decide its admission control policy based on either the marginal distributions or the autocorrelation functions of the arrival streams, depending which conditions are satisfied.

References

[1] J. Abate, G.L. Choudhury, and W. Whitt. Waiting-time tail probabilities in queues with long-tail service-time distributions. *Queueing systems.*, 1994.

[2] S. Asmussen. Aspects of matrix wiener-hopf factorization in applied probability. *Math. Scientist*, 14:101–116, 1989.

[3] S. Asmussen. Ladder heights and the markov-modulated M/G/1 queue. *Stochastic Processes and their Applications*, 37:313–326, 1991.

[4] S. Asmussen, L. F. Henriksen, and C. Klüppelberg. Large claims approximations for risk processes in a markovian environment. *Stochastic Processes and their Applications*, 54:29–43, 1994.

[5] K. B. Athreya and P. E. Ney. *Branching Processes.* Springer-Verlag, 1972.

[6] C. S. Chang. Stability, queue lenght and delay of deterministic and stochastic queueing networks. *IEEE Transactions on Automatic Control*, 39:913–931, 1994.

[7] V. P. Chistakov. A theorem on sums on independent positive random variables and its application to branching random processes. *Theor. Probab. Appl.*, 9:640–648, 1964.

[8] Gagan L. Choudury, David M. Lucantoni, and Ward Whitt. Squeezing the most of atm. *to appear in IEEE Trans. on Communications*, 1995.

[9] E. Cinlar. *Introduction to Stochastic Processes.* Prentice-Hall, 1975.

[10] A. I. Elwalid and D. Mitra. Effective bandwidth of general markovian traffic sources and admission control of high speed networks. *IEEE/ACM Trans. on Networking*, 1(3):329–343, June 1993.

[11] M. W. Garett and W. Willinger. Analysis, modeling and generation of self-similar vbr video traffic. In *SIGCOMM'94*, pages 269–280, 1994.

[12] P. V. Glynn and W. Whitt. Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. *Studies in Appl. Prob.*, 1994.

[13] R. Guerin, H. Ahmadi, and M. Nagshineh. Equivalent capacity and its application to bandwidth allocation in high-speed networks. *IEEE J. Select. Areas Commun.*, 9:968–981, 1991.

[14] P. R. Jelenković and A. A. Lazar. On the dependence of the queue tail distribution on multiple time scales of atm multiplexers. In *Conference on Information Sciences and Systems*, pages 435–440, Baltimore, MD, March 1995.

[15] P. R. Jelenković and A. A. Lazar. Subexponential asumptotics of a network multiplexer. CTR Technical Report CU/CTR/TR, Columbia University, September 1995.

[16] P. R. Jelenković and A. A. Lazar. Evaluating the queue length distribution of an atm multiplexer with multiple time scale arrivals. In *Proceedings of INFOCOM'96*, San Francisco, California, March 1996, *to appear*.

[17] F. P. Kelly. Effective bandwidths at multi-class queues. *Queueing Systems*, 9:5–16, 1991.

[18] D. Tse, R. Gallager, and J. Tsitsiklis. Statistical multiplexing of multiple time-scale markov streams. *IEEE, Selected Areas in Communications*, August 1995.

[19] N. Veraverbeke. Asymptotic behavior of wiener-hopf factors of a random walk. *Stochastic Proc. and Appl.*, 5:27–37, 1977.

[20] E. Willekens and J. L. Teugels. Asymptotic expansion for waiting time probabilities in an M/G/1 queue with long-tailed service time. *Queueing Systems*, 10:295–312, 1992.

[21] Chow Y. S. and Teicher H. *Probability Theory.* Springer-Verlag, 1988.