Asymptotic Results for Multiplexing Subexponential On-Off Processes

Predrag R. Jelenković

Bell Laboratories
Lucent Technologies
600 Mountain Avenue
Murray Hill, NJ 07974

Aurel A. Lazar

Department of Electrical Engineering
and CTR
Columbia University
New York, NY 10027

July 15, 1996
Revised: April 15, 1998

Abstract

Consider an aggregate arrival process A^N obtained by multiplexing N On-Off processes with exponential Off periods of rate λ and subexponential On periods τ^{on}. As N goes to infinity, with $\lambda N \to \Lambda$, A^N approaches an $M/G/\infty$ type process. Both for finite and infinite N, we obtain the asymptotic characterization of the arrival process activity period.

Using these results we investigate a fluid queue with the limiting M/G/∞ arrival process A_t^∞ and capacity c. When On periods are regularly varying (with non-integer exponent), we derive a *precise asymptotic behavior* of the queue length random variable Q_t^P observed at the beginning of the arrival process activity periods

$$\mathbb{P}[Q_t^P > x] \sim \Lambda \frac{r+\rho-c}{c-\rho} \int_{x/(r+\rho-c)}^{\infty} \mathbb{P}[\tau^{on} > u] du \quad x \to \infty,$$

where $\rho = \mathbb{E}A_t^\infty < c$; r ($c \leq r$) is the rate at which the fluid is arriving during an On period. The asymptotic (time average) queue distribution lower bound is obtained under more general assumptions on On periods than regular variation.

In addition, we analyze a queueing system in which one On-Off process, whose On period belongs to a subclass of subexponential distributions, is multiplexed with independent exponential processes with aggregate expected rate $\mathbb{E}e_t$. This system is shown to be asymptotically equivalent to the same queueing system with the exponential arrival processes being replaced by their total mean value $\mathbb{E}e_t$.

Keywords: Non-Cramér type conditions; Subexponential distributions; Long-tailed distributions; Long-range dependency; Network multiplexer; Fluid flow queue; M/G/∞ queue.

Submitted for publication to Advances in Applied Probability. Appeared in part in: INFOCOM'97, Kobe, Japan, April 1997; ITC 15, Washington, D.C., USA, June 1997 (The Best Student Paper Award); INFORMS Applied Probability Conference, June 30-July 2, 1997, Boston (invited talk).

1. Introduction

The problem of multiplexing On-Off sources arises frequently as the basic model of contention in multimedia communication systems, as well as in some storage systems. More specifically, in modern multimedia communication networks, such as ATM, various calls are simultaneously established among the different source-destination pairs. These calls are usually discretized (packetized). An individual call/source can be either active (in which case it transmits a packet) or silent (no transmission). Therefore, each source can be viewed as an On-Off process. Along their routes, sources that are simultaneously active share common network resources: bandwidth, buffer space, computational power, etc. The fundamental building blocks for sharing bandwidth and buffer space are network multiplexers (MUX). Sharing of common network resources may lead to Quality Of Service (QOS) degradation for individual flows. Therefore, it is important to have computationally efficient algorithms for evaluating QOS measures under all possible traffic loads. A first step towards a satisfactory solution to this problem is understanding network MUX units in isolation. The most fundamental mathematical model of a network MUX is an infinite buffer queue loaded with (multiplexed) On-Off arrival processes (sources); the main QOS parameter (performance measure) for this queueing system is the buffer occupancy probability distribution.

The problem of multiplexing dates back to [RUB73, COH74]. In [COH74], Cohen obtained a complete Laplace transform solution to this problem! (More recently, he revisited this problem in [COH94].) However, inverting the Laplace transform is usually a very tedious process. Hence, investigating computationally tractable exact and approximate solution techniques are needed. For Markovian (fluid) On-Off processes a thorough investigation of this problem was done in [AMS82]. Many other results for multiplexing Markovian On-Off processes followed. These led to the Equivalent Bandwidth theory for Markovian, or, in general, exponentially bounded arrival processes; extensive references can be found for example in [EHL95, GLW94, DUC95].

Recently statistical analysis has increasingly shown that the traffic streams in modern broad-band networks exhibit long-tailed (subexponential) characteristics. For the case of Ethernet traffic such results were examined in [LTW93]. These statistical results have stimulated research in queueing analysis under the heavy-tailed (non Cramér) assumptions. Queueing analysis with self-similar long-range dependent arrival processes appears in [NOR94, DUC95, LTG95, PAM96, RYL96, RES96]. Recently, long-tailed characteristics of the scene length distribution of MPEG video streams were explored in [JLS95b, HEL96, JLS96, JLA96h].

Parallel to the modeling approach through self-similar long-range dependent processes, a more analytically tractable approach using fluid renewal type models in which renewal times are long-tailed has been explored in [ANA95, HRS96]. Queueing results in these two papers rely on the classical result by Pakes [PAK75] on the subexponential (long-tailed) asymptotics of the waiting time distribution in a GI/GI/1 queue or in earlier work of Cohen [COH73] which considered a regularly varying GI/GI/1 queue.

The result of Pakes has been generalized to a Markov modulated setting [AHK94, JLA95g]. In [AHK94] the subexponential asymptotics of a Markov modulated M/G/1 queue was investigated. Work in [JLA95g] further generalized these results to Markov modulated G/G/1 queues. In the same paper it was shown that a subexponential GI/GI/1 queue is invariant under Markov modulation. In other words, a subexponential Markov modulated G/G/1 queue has the same asymptotics as the corresponding GI/GI/1 queue. These results made possible the analysis of a subexponential semi-Markov fluid queue ([JLA95g]). Further generalizations of the result in [JLA95g] to arrival processes with a more complex dependency structure were investigated in [ASC97]. Asymptotic expansion refinements of Pakes' result can be found in [WIT92, ACW94].

The analysis of a fluid queue in which more than one long-tailed process is multiplexed appears to be a very difficult problem. This is due to the fact that the renewal structure of an aggregate arrival process may be very complex, although the appearance of each individual process may be truly innocuous (like an On-Off process). The complex autocorrelation structure of the aggregate process obtained by multiplexing long-tailed On-Off processes has been examined in [HRS96]. General bounds for multiplexing long-tailed fluid processes have been derived in [CHW95]. In [BOX96] a limiting case of an infinite number of On-Off processes with regularly varying On distribution has been investigated. In the same paper (see also [BOX97]) a case of two processes, one of which had regularly varying On periods and the other had exponential On periods, has been solved. Similar scenario with intermediately regularly varying On periods has been examined in [RSS97]. The literature does not explicitly give precise asymptotic results for the case of multiplexing two or more long-tailed processes.

From a mathematical point of view the new results in this paper were achieved through the combination of the Theory of Subexponential Distributions, Renewal Theory, Karamata's Theory and the utilization of sample path arguments. From an engineering standpoint this paper advances two important results. The first result intuitively states that when a process with subexponential characteristics (e.g., MPEG video) is multiplexed with a process that has exponential characteristics (e.g., voice), the contribution to the large buffer asymptotics of the exponential processes is reflected only through their mean values. This result suggests that, under appropriate conditions, for admission control of both VBR video and voice streams, the voice streams need to be characterized only by their mean values. The second result is an accurate approximation with low computational complexity of the large buffer probabilities of finitely many subexponential On-Off processes. Besides accuracy, it is of special importance for engineering the MUX that this approximation has basically *negligible computational complexity.* To the best of our knowledge, this is the only result in literature of comparable computational complexity that is both proven theoretically and demonstrated experimentally as a good approximation for the buffer overflow probabilities with multiplexed long-tailed arrival streams.

The rest of the paper is organized as follows. Section 2 contains necessary definitions and examples of long-tailed and subexponential distributions. In Section 3 we examine the aggregate arrival process obtained by multiplexing N independent and identically distributed On-Off processes. For this process we derive the asymptotic relation between the distribution of its activity period and the distribution of On periods of individual processes. For the case when On periods are regularly varying (with noninteger exponent) using Karamata's theory we obtain a precise asymptotic behavior of the server overflow distribution during the arrival process activity period. Using these asymptotic relations, in Section 4, we obtain several results for the fluid queue asymptotics of multiplexed long-tailed processes. In the same section, based on queueing theoretic results, we suggest a computationally efficient approximation for multiplexing a finite number of subexponential On-Off processes. Accuracy of the approximation is tested using extensive simulation experiments. The paper is concluded in Section 5.

2. Long-Tailed and Subexponential Distributions

This section contains necessary definitions of long-tailed and subexponential distributions. For convenience, we give some basic results on these distributions in Appendix A.

Definition 1 *A distribution function F on $[0,\infty)$ is called* long-tailed *($F \in \mathcal{L}$) if*

$$\lim_{x\to\infty} \frac{1-F(x-y)}{1-F(x)} = 1, \quad \forall y \in \mathbb{R}. \tag{2.1}$$

Definition 2 *A distribution function F on $[0,\infty)$ is called* subexponential *($F \in \mathcal{S}$) if*

$$\lim_{x\to\infty} \frac{1-F^{*2}(x)}{1-F(x)} = 2, \tag{2.2}$$

where F^{*2} denotes the 2-nd convolution of F with itself, i.e., $F^{*2}(x) = \int_{[0,\infty)} F(x-y)F(dy)$.

The class of subexponential distributions was first introduced by Chistyakov [CHI64]. The definition is motivated by the simplification of the asymptotic analysis of convolution tails. One of the best known examples of distribution functions in $\mathcal{S}$ (and $\mathcal{L}$) are functions of Regular Variation $\mathcal{R}_{-\alpha}$ (in particular Pareto family); $F \in \mathcal{R}_{-\alpha}$ if it is given by

$$F(x) = 1 - \frac{l(x)}{x^\alpha} \quad \alpha \geq 0,$$

where $l(x) : \mathbb{R}_+ \to \mathbb{R}_+$ is a function of slow variation, i.e., $\lim_{x\to\infty} l(\delta x)/l(x) = 1, \delta > 1$. These functions were invented by Karamata [KAR30] (the main reference book is [BGT87]). The other examples include lognormal and some Weibull distributions (see [KLU88, JLA95g]).

3. Analysis of the Aggregate Arrival Process

This section consists of two parts. The first part is contained in Section 3.1, where we asymptotically relate the tail of the activity period distribution of an aggregate arrival process, obtained by multiplexing independent and identically distributed On-Off processes, and the tail of the On period distribution of the individual On-Off processes. The main results are given in Theorem 3 and Theorem 5. In the second part, Section 3.2, using Karamata's theory we derive the asymptotic behavior of the distribution of the queue increment during the arrival process activity period. In Section 4, these results will be used to obtain asymptotic queueing results.

More formally, consider two independent sequences of i.i.d. random variables $\{\tau_n^{off}, n \geq 0\}$, $\{\tau_n^{on}, n \geq 0\}$, $\tau_0^{off} = \tau_0^{on} = 0$. Define a point process $T_n^{off} \stackrel{\text{def}}{=} \sum_{i=0}^{n}(\tau_i^{off} + \tau_i^{on}), n \geq 0$;

this process will be interpreted as representing the beginnings of Off periods in an On-Off process. Further, define an On-Off process a_t with rate r, as

$$a_t = r \text{ if } T_n^{off} - \tau_n^{on} \leq t < T_n^{off}, \quad n \geq 1,$$

and $a_t = 0$, otherwise. For the rest of the paper, unless otherwise specified, we will assume that τ_n^{off} is exponentially distributed with parameter λ, i.e., $\mathbb{P}[\tau_n^{off} > t] = e^{-\lambda t}, t \geq 0$. Also, τ_n^{on} is assumed to have a finite mean. Steady state probabilities of this process are given as $\pi_0 = \lim_{t\to\infty} \mathbb{P}[a_t = 0] = 1/(1 + \lambda \mathbb{E}\tau^{on}) = 1 - \pi_1$, where $\pi_1 = \lim_{t\to\infty} \mathbb{P}[a_t = r]$. Let $A^N = \sum_1^N a^i$, be an aggregate arrival process obtained by multiplexing N independent and identically distributed On-Off processes $a^i, 1 \leq i \leq N$.

3.1 Asymptotic Behavior of the Aggregate Process Activity Period

The *central idea* of this section is to relate the rate of convergence of $\mathbb{P}[a_t = 0]$ to its steady state and the tail of its On period (Section 3.1.1). Then, the rate of convergence for the aggregate process is easily computable from $\mathbb{P}[A_t^N = 0] = (\mathbb{P}[a_t = 0])^N$; from here one can refer back and relate $\mathbb{P}[A_t^N = 0]$ to the tail of its activity period distribution (Sections 3.1.2 and 3.1.3).

3.1.1 Single On-Off Process: Convergence to Steady State

Let us now investigate the speed of convergence of $\mathbb{P}[a_t = 0]$ to its steady state. For that reason we define a transient function $F_{tr}(t) \stackrel{\text{def}}{=} \mathbb{P}[a_t = 1]/\pi_1$; notice that $F_{tr}(0) = 0$ and $\lim_{t\to\infty} F_{tr}(t) = 1$. Therefore, if $F_{tr}(t)$ is monotonic, it will represent a proper probability distribution function, in which case we call it the transient probability measure. To simplify the notation, throughout the rest of the paper, for any distribution F we define its tail function as $\bar{F}(x) \stackrel{\text{def}}{=} 1 - F(x)$; in addition, if F has a finite mean $m \stackrel{\text{def}}{=} \int_0^\infty \bar{F}(x)dx$, we denote its integrated tail distribution as $F_1(x) \stackrel{\text{def}}{=} m^{-1} \int_0^x \bar{F}(u)du$. In the following theorem we use a class of subexponential distributions named $\mathcal{S}_d$. For all practical purposes $F \in \mathcal{S}_d$ is virtually the same as $F \in \mathcal{S}$. A precise definition of $\mathcal{S}_d$ is given in Definition 5 (A) (henceforth, symbol (A) will be used as a shortcut for Appendix A).

Theorem 1 *Let F be a distribution of τ^{on} with $\mathbb{E}\tau^{on} < \infty$ and let τ^{off} be exponentially distributed with parameter λ. Then, the Laplace-Stieltjes transform of the transient function is given as*

$$\tilde{F}_{tr}(s) = s \int_0^\infty e^{-st} F_{tr}(t)dt = \frac{(1 + \lambda \mathbb{E}\tau^{on})F_1(s)}{1 + \lambda \mathbb{E}\tau^{on} F_1(s)}. \tag{3.1}$$

If $F_1 \in \mathcal{S}$ and $\lambda \mathbb{E}\tau^{on} < 1$, then

$$\bar{F}_{tr}(t) \sim \frac{1}{1 + \lambda \mathbb{E}\tau^{on}} \bar{F}_1(t) \quad \text{as} \quad t \to \infty. \tag{3.2}$$

If $F \in \mathcal{S}_d$ and $\lambda \mathbb{E}\tau^{on} < 1$, then (3.2) holds, and the density function $f_{tr}(t) \stackrel{\text{def}}{=} dF_{tr}(t)/dt$ satisfies

$$f_{tr}(t) \sim \frac{1}{1+\lambda \mathbb{E}\tau^{on}} \frac{\bar{F}(t)}{\mathbb{E}\tau^{on}} \quad as \ t \to \infty. \tag{3.3}$$

Proof: Equation (3.1) follows directly from equation (2.1.3) in [COH74], and the observation that $\tilde{F}_1(s) = (1-\tilde{F}(s))/(s\mathbb{E}\tau^{on})$. The asymptotic relation in (3.2) is a direct consequence of Theorem 14 (A), and the fact that for all $\lambda \mathbb{E}\tau^{on} < 1$ equation (3.1) implies

$$\bar{F}_{tr}(t) = (1+\lambda \mathbb{E}\tau^{on}) \sum_{n=1}^{\infty} (-\lambda \mathbb{E}\tau^{on})^{n-1} \overline{F_1^{*n}}(t). \tag{3.4}$$

Similarly, equation (3.3) follows by differentiation of the equation above and Theorem 15 (A). This finishes the proof of the theorem. $\diamond$

For the remainder of this paper, it is of special interest to find sufficient conditions under which the transient function $F_{tr}(t)$ is a proper distribution function (i.e., monotonic). One set of sufficient conditions, for the case when Off periods are large, is given in the following theorem (this is typically satisfied when there are a large number of processes with a small average arrival rate).

Theorem 2 *Let F be a distribution of τ^{on} with $\mathbb{E}\tau^{on} < \infty$ and let τ^{off} be exponentially distributed with parameter λ. For any fixed $F \in \mathcal{S}_d$, there exists a $\lambda_0 > 0$, such that for all $\lambda < \lambda_0$, $F_{tr}(t) \equiv F_{tr}^{\lambda}(t)$ is a probability distribution function, i.e., $f_{tr}(t) \geq 0, t \geq 0$.*

Proof: Differentiation of (3.4) gives

$$\begin{aligned} f_{tr}(t) &= (1+\lambda \mathbb{E}\tau^{on}) \sum_{n=1}^{\infty} (-\lambda \mathbb{E}\tau^{on})^{n-1} f^{\otimes n}(t) \\ &= (1+\lambda \mathbb{E}\tau^{on}) \left(f(t) - \lambda \mathbb{E}\tau^{on} \sum_{n=2}^{\infty} (-\lambda \mathbb{E}\tau^{on})^{n-2} f^{\otimes n}(t) \right) \\ &\geq (1+\lambda \mathbb{E}\tau^{on}) \left(f(t) - \lambda \mathbb{E}\tau^{on} \sum_{n=2}^{\infty} (\lambda \mathbb{E}\tau^{on})^{n-2} f^{\otimes n}(t) \right), \end{aligned}$$

where $f(t) = \bar{F}(t)/\mathbb{E}\tau^{on}$. Now, by applying Lemma 7 (i) (A), for $\epsilon > 0$,

$$\mathbb{E}\tau^{on} \sum_{n=2}^{\infty} (\lambda \mathbb{E}\tau^{on})^{n-2} f^{\otimes n}(t) \leq \frac{\mathbb{E}\tau^{on} C_\epsilon (1+\epsilon)^2 f(t)}{1-(1+\epsilon)\lambda \mathbb{E}\tau^{on}} \stackrel{\text{def}}{=} \frac{C'_\epsilon f(t)}{1-(1+\epsilon)\lambda \mathbb{E}\tau^{on}},$$

for some $C_\epsilon, C'_\epsilon > 0$, and all $\lambda < 1/(\mathbb{E}\tau^{on}(1+\epsilon))$. Therefore,

$$f_{tr}(t) \geq (1+\lambda \mathbb{E}\tau^{on}) f(t) \left(1 - \frac{\lambda C'_\epsilon}{1-(1+\epsilon)\lambda \mathbb{E}\tau^{on}}\right) \geq 0,$$

for all $\lambda \leq 1/(\mathbb{E}\tau^{on}(1+\epsilon) + C'_\epsilon)$. This concludes the proof of the theorem. $\diamond$

3.1.2 Finite number of processes

In this section we consider an aggregate process A_t^N that is obtained by multiplexing N independent On-Off processes, i.e., $A_t^N = \sum_{i=1}^N a_t^i$, where a_t^i are independent and identically distributed On-Off processes. Note that the indicator process $1(A_t^N = 0)$ is an On-Off process with exponentially distributed Off periods with parameter $N\lambda$. Let $\{I_n^{N,off}, I_n^{N,on}, n \geq 1\}$ be the lengths of the n-th Off and On periods in the indicator process $1(A_t^N = 0)$, respectively. In the following theorem we characterize asymptotically the tail of the distribution function of $I_n^{N,on}$. Observe that the steady state probability of the aggregate process being in state 0 is given by $\Pi_0 = \pi_0^N$.

Theorem 3 *Assume that $\lambda < \lambda_0$, where λ_0 is the same as in Theorem 2 (or that $\mathbb{P}[a_t = 0]$ is monotonic, and $\lambda \mathbb{E}\tau^{on} < 1$). If $F \in \mathcal{S}_d$, then*

$$\int_t^\infty \mathbb{P}[I^{N,on} > u]du \sim (1 + \lambda \mathbb{E}\tau^{on})^{N-1} \int_t^\infty \mathbb{P}[\tau^{on} > u]du \quad \textit{as} \quad t \to \infty, \tag{3.5}$$

and

$$\mathbb{P}[I^{N,on} > t] \sim (1 + \lambda \mathbb{E}\tau^{on})^{N-1} \mathbb{P}[\tau^{on} > t] \quad \textit{as} \quad t \to \infty. \tag{3.6}$$

Proof: From the definition of $F_{tr}(t)$ and Theorem 1 it follows that

$$\mathbb{P}[a_t^i = 0] = \pi_0(1 + \pi_1 \bar{F}_1(t) + o(\bar{F}_1(t))), \quad 1 \leq i \leq N,$$

where $\pi_0 = 1 - \pi_1 = 1/(1 + \lambda \mathbb{E}\tau^{on})$, and F_1 is the integrated tail distribution of F. This implies that

$$\mathbb{P}[A_t = 0] = \mathbb{P}[a_t^i = 0]^N = \pi_0^N(1 + N\pi_1 \bar{F}_1(t) + o(\bar{F}_1(t))) \quad \text{as} \quad t \to \infty.$$

Therefore, the transient function $F_{tr,N}(t)$ satisfies the following asymptotics

$$\bar{F}_{tr,N}(t) \stackrel{\text{def}}{=} 1 - \frac{\mathbb{P}[A_t^N = 1]}{1 - \pi_0^N} = \frac{\mathbb{P}[A_t^N = 0] - \pi_0^N}{1 - \pi_0^N} \sim \frac{\pi_0^N N \pi_1}{1 - \pi_0^N} \bar{F}_1(t) \quad \text{as} \quad t \to \infty. \tag{3.7}$$

By Theorem 2, for all $\lambda < \lambda_0$, $F_{tr}(t)$ is a distribution function, implying that $\mathbb{P}[a_t^i = 0]$ and $\mathbb{P}[A_t^N = 0]$ are monotonic, which further implies that $F_{tr,N}(t)$ is a probability distribution function. Here, let $F_{N,1}$ be the integrated tail distribution of $I^{N,on}$ and $\tilde{F}_{N,1}$ be its Laplace-Stieltjes transform. Then, similarly as in (3.1), we obtain

$$\tilde{F}_{tr,N}(s) = \frac{(1 + N\lambda \mathbb{E} I^{N,on})\tilde{F}_{N,1}(s)}{1 + N\lambda \mathbb{E} I^{N,on} \tilde{F}_{N,1}(s)}.$$

After simple algebra, it follows that

$$\tilde{F}_{N,1}(s) = \frac{\pi_0^N \tilde{F}_{tr,N}(s)}{1 - (1 - \pi_0^N)\tilde{F}_{tr,N}(s)},$$

or, in the time domain

$$\bar{F}_{N,1}(t) = \pi_0^N \sum_{n=1}^{\infty} (1-\pi_0^N)^{n-1} \overline{F_{tr,N}^{*n}}(t). \tag{3.8}$$

Now, $F \in \mathcal{S}_d$ implies $F_1 \in \mathcal{S}$ (Theorem 1.1, [KLU89]), which by (3.7), and Lemma 5 (ii) (A), yields $F_{tr,N} \in \mathcal{S}$. Hence, by Theorem 14 (A) and (3.8) it follows that

$$\bar{F}_{N,1}(t) \sim \pi_0^{-N} \bar{F}_{tr,N}(t) \sim \frac{N\pi_1}{1-\pi_0^N} \bar{F}_1(t) \text{ as } t \to \infty, \tag{3.9}$$

where the second asymptotic relation follows from (3.7). By replacing $\mathbb{E}I^{N,on} = (1/\pi_0^N - 1)/(N\lambda)$ in (3.9) we obtain

$$\int_t^{\infty} \mathbb{P}[I^{N,on} > u] du \sim (1+\lambda \mathbb{E}\tau^{on})^{N-1} \int_t^{\infty} \mathbb{P}[\tau^{on} > u] du \text{ as } t \to \infty,$$

which proves (3.5).

In order to prove (3.6) we first determine the asymptotic behavior of $f_{tr,N}$

$$\begin{aligned} f_{tr,N}(t) &\stackrel{\text{def}}{=} \frac{d}{dt} \frac{\mathbb{P}[A_t^N = 1]}{1-\pi_0^N} \\ &= \frac{d}{dt} \frac{1-\mathbb{P}[a_t = 0]^N}{1-\pi_0^N} \\ &= \frac{d}{dt} \frac{1-(1-\pi_1 F_{tr}(t))^N}{1-\pi_0^N} \\ &= \frac{N}{1-\pi_0^N}(1-\pi_1 F_{tr}(t))^{N-1} \pi_1 f_{tr}(t) \\ &\sim \frac{\pi_0^N N\pi_1}{1-\pi_0^N} \frac{\bar{F}(t)}{\mathbb{E}\tau^{on}} \quad \text{as} \quad t \to \infty, \end{aligned} \tag{3.10}$$

where the last asymptotic relation follows from (3.3). Next, by taking a derivative in (3.8) we obtain

$$\bar{F}_N(t) = \mathbb{E}I^{N,on} \pi_0^N \sum_{n=1}^{\infty} (1-\pi_0^N)^{n-1} f_{tr,N}{}^{\otimes n}(t), \tag{3.11}$$

where $F_N(t)$ is the distribution function of $I^{N,on}$. Finally, by applying Theorem 15 (A) we get

$$\bar{F}_N(t) \sim \mathbb{E}I^{N,on} \pi_0^{-N} f_{tr,N}(t) \text{ as } t \to \infty,$$

which together with (3.10) yields (3.6). $\diamond$

3.1.3 Infinite number of On-Off processes

In this subsection we analyze the limiting case of an infinite number of On-Off sources. First, we show that the aggregate process A_t^N converges in distribution to an M/G/∞ process A_t^∞ which we define as follows. Let $T_n, n \geq 0, T_0 = 0$, be a Poisson process with rate Λ. Define $A_t^\infty = \sum_{n=1}^{\infty} r 1(T_n \leq t < T_n + \tau_n^{on}), r > 0$. Then the following theorem holds.

Theorem 4 *If $\mathbb{E}\tau_n^{on} < \infty$, and $\lambda N \to \Lambda$ as $N \to \infty$, then*

$$A_t^N \stackrel{d}{\Rightarrow} A_t^\infty \text{ as } N \to \infty, \tag{3.12}$$

where $\stackrel{d}{\Rightarrow}$ symbolizes convergence in distribution.

Proof: It is enough to prove that the beginnings of the On periods in the process A_t^N converge to a Poisson process with rate Λ. This follows from a classical result on multiplexing a large number of renewal processes [COS54, CIN72]. $\diamond$

Lemma 1 *The transient probability of the arrival process A_t^∞ being silent is given by*

$$\mathbb{P}[A_t^\infty = 0] = e^{-\Lambda \int_0^t \mathbb{P}[\tau^{on} > u]du}. \tag{3.13}$$

Furthermore, if $\mathbb{E}\tau^{on} < \infty$, then $\lim_{t\to\infty} \mathbb{P}[A_t^\infty = 0] = e^{-\Lambda \mathbb{E}\tau^{on}}$.

Proof: Follows from Theorem 2.2 in [COH74]. $\diamond$

Remark: Observe that A_t^∞ represents the number of customers in service in an M/G/∞ queue in which the customer service requirement has the same distribution as τ^{on} and the arrival rate is Λ. (For recent asymptotic results on M/G/∞ processes see [PAM97].)

Note that Theorem 3 implies that

$$\lim_{\substack{N\to\infty \\ \lambda N \to \Lambda}} \lim_{t\to\infty} \frac{\mathbb{P}[I^{N,on} > t]}{\mathbb{P}[\tau^{on} > t]} = e^{\Lambda \tau^{on}}. \tag{3.14}$$

However, this does not necessarily imply that we can interchange the limit and derive the asymptotics of the activity period $I^{\infty,on}$ in the A_t^∞ process. The following result gives the asymptotic characterization of $I^{\infty,on}$ and indeed shows that the limits in (3.14) can be interchanged.

Theorem 5 *The asymptotics of the distribution of $I^{\infty,on}$ and its integrated tail are related as follows:*

(i) If $F_1 \in \mathcal{S}$, then

$$\int_t^\infty \mathbb{P}[I^{\infty,on} > u]du \sim e^{\Lambda \mathbb{E}\tau^{on}} \int_t^\infty \mathbb{P}[\tau^{on} > u]du \quad \text{as } t \to \infty.$$

(ii) If in addition $F \in \mathcal{S}_d$, then

$$\mathbb{P}[I^{\infty,on} > t] \sim e^{\Lambda \mathbb{E}\tau^{on}} \mathbb{P}[\tau^{on} > t] \quad \text{as } t \to \infty.$$

Remark: For the case of τ^{on} being regularly varying $\mathbb{P}[\tau^{on} > t] = l(t)/t^{\alpha}, 1 < \alpha < 2$, this result was obtained in [BOX96] where Karamata's Tauberian/Abelian theorems were used to asymptotically relate $I^{\infty,on}$ and τ^{on}.

Proof: Similarly as before, we define a transient probability distribution function

$$F_{tr,\infty}(t) \stackrel{\text{def}}{=} \frac{\mathbb{P}[A_t^{\infty} > 0]}{1 - e^{-\Lambda \mathbb{E}\tau^{on}}}.$$

Observe that this is a proper probability distribution function, i.e., it is monotonically increasing from $F_{tr,\infty}(0) = 0$ to $\lim_{t\to\infty} F_{tr,\infty}(t) = 1$. Next, by using Lemma 1 we derive

$$\begin{aligned} \bar{F}_{tr,\infty}(t) &= 1 - \frac{1 - e^{-\Lambda \int_0^t \mathbb{P}[\tau^{on} > u] du}}{1 - e^{-\Lambda \mathbb{E}\tau^{on}}} \\ &= \frac{e^{-\Lambda \int_0^t \mathbb{P}[\tau^{on} > u] du} \left(1 - e^{-\Lambda \int_t^\infty \mathbb{P}[\tau^{on} > u] du}\right)}{1 - e^{-\Lambda \mathbb{E}\tau^{on}}} \\ &\sim \frac{\Lambda e^{-\Lambda \mathbb{E}\tau^{on}}}{1 - e^{-\Lambda \mathbb{E}\tau^{on}}} \int_t^\infty \mathbb{P}[\tau^{on} > u] du \end{aligned} \tag{3.15}$$

as $t \to \infty$. Now, the Laplace transform of $F_{tr,\infty}$ for any $s \in \mathbb{R}^+$ is given by

$$\begin{aligned} \tilde{F}_{tr,\infty}(s) &= \int_0^\infty e^{-st} dF_{tr,\infty}(t) \\ &= s \int_0^\infty e^{-st} F_{tr,\infty}(t) dt \\ &= \frac{s}{1 - e^{-\Lambda \mathbb{E}\tau^{on}}} \mathbb{E} \int_0^\infty e^{-st} 1(A_t^\infty > 0) dt \\ &= \frac{s}{1 - e^{-\Lambda \mathbb{E}\tau^{on}}} \mathbb{E} \sum_{n=0}^\infty \int_{I_{n+1}^{\infty,off} + \sum_{i=0}^n (I_i^{\infty,on} + I_i^{\infty,off})}^{\sum_{i=0}^{n+1} (I_i^{\infty,on} + I_i^{\infty,off})} e^{-st} dt \\ &= \frac{s}{1 - e^{-\Lambda \mathbb{E}\tau^{on}}} \mathbb{E} \sum_{n=0}^\infty e^{-s\left(I_{n+1}^{\infty,off} + \sum_{i=0}^n (I_i^{\infty,on} + I_i^{\infty,off})\right)} \frac{\left(1 - e^{-sI_{n+1}^{\infty,on}}\right)}{s} \\ &= \frac{\Lambda}{1 - e^{-\Lambda \mathbb{E}\tau^{on}}} \frac{\left(1 - \mathbb{E}e^{-sI^{\infty,on}}\right)}{s + \Lambda - \Lambda \mathbb{E}e^{-sI^{\infty,on}}}. \end{aligned}$$

By solving the last equation in $\mathbb{E}e^{-sI^{\infty,on}}$ and putting $\gamma = 1 - e^{-\Lambda \mathbb{E}\tau^{on}} (< 1)$, we obtain

$$\mathbb{E}e^{-sI^{\infty,on}} = 1 - \frac{s\Lambda^{-1}\gamma \tilde{F}_{tr,\infty}(s)}{1 - \gamma \tilde{F}_{tr,\infty}(s)},$$

or equivalently

$$\frac{1 - \mathbb{E}e^{-sI^{\infty,on}}}{s} = \frac{\Lambda^{-1}\gamma \tilde{F}_{tr,\infty}(s)}{1 - \gamma \tilde{F}_{tr,\infty}(s)}.$$

Observing that $\frac{1-\mathbb{E}e^{-sI^{\infty,on}}}{s}$ is the Laplace-Stieltjes transform of $\int_0^t \mathbb{P}[I^{\infty,on} > u]du$ we arrive at

$$\int_0^t \mathbb{P}[I^{\infty,on} > u]du = \Lambda^{-1} \sum_{n=1}^{\infty} \gamma^n \left(F_{tr,,\infty}\right)^{*n}(t), \tag{3.16}$$

which in combination with (3.15) and the same arguments as in the proof of Theorem 3 yields the conclusion (i) of this theorem.

The proof of (ii) can be obtained in a similar manner by first deriving the asymptotic behavior of $f_{tr}(t)$ and combining it with the derivative of (3.16); we omit the details. $\diamond$

3.2 Total Server Overflow During the Activity Period

Let $B_n, n \geq 1$, be a sequence of random variables representing the total amount of fluid that is brought to the system during the nth activity period, i.e., $B_n = \int_{t_n^b}^{t_n^e} A_t^{\infty} dt$, where t_n^b, t_n^e, represent the beginning and end of the nth activity period, respectively. Further, define $D_{c,n} \stackrel{\text{def}}{=} B_n - cI_n^{on}, 0 < c \leq r$; note that $D_n \equiv D_{c,n}$ is a non-negative random variable. If we imagine that A_t^{∞} represents the rate at which the fluid is arriving to a fluid queue, and that c is the constant rate at which the queue drains, then D_n represents the queue increment during the nth activity period. In order to derive the queueing asymptotics, we first have to understand the asymptotic behavior of D_n. Unfortunately, this is a much more difficult task than the investigation of the asymptotic behavior of the activity period that we have done so far. For that reason we are forced to work under much more restrictive assumptions with distribution functions of regular variation. The method of proof for the following result will be through Karamata's Tauberian/Abelian theorems.

Theorem 6 *Consider an $M/G/\infty$ arrival process with On periods being regularly varying $\mathbb{P}[\tau^{on} > x] = l(x)/x^{\alpha}, \alpha > 1$, where α is noninteger. If $0 < c \leq r$, then*

$$\mathbb{P}[D_{c,n} > x] \sim e^{\Lambda \mathbb{E}\tau^{on}} \mathbb{P}\left[\tau^{on} > \frac{x}{r + r\Lambda\mathbb{E}\tau^{on} - c}\right] \quad as \;\; x \to \infty. \tag{3.17}$$

Proof: Given in Appendix B. $\diamond$

Next, consider a stationary version of the arrival process $A_t^{\infty,s} = \sum_{-\infty<n<\infty} r1(T_n \leq t < T_n + \tau_n^{on})$, where T_n is a stationary Poisson process with rate Λ. Given that at time $t = 0$, the arrival process is active ($A_t > 0$), denote with $D_{c(0)}$ the total queue increment since the beginning of the last activity period until time zero, i.e., $D_{c(0)} = \int_{t_0^b}^0 (A_t^{\infty} - c)dt, 0 < c \leq r$, where t_0^b represents the beginning of the activity period that is still active at $t = 0$.

Now, by Theorem 4.3, pp. 64, [ASM87], it follows that the process $\{T_n + \tau_n^{on}, -\infty < n < \infty\}$ is also a stationary Poisson process with the same rate Λ. Therefore, the process $A_t^{\infty,s}$ is reversible. This implies that $D_{c(0)}$ is equal in distribution to $\int_0^{t_0^e} (A_t^{\infty} - c)dt$, where t_0^e represents end of the activity period that is active at $t = 0$. For simplicity reasons, we will refer to both of these variables with $D_{c(0)}$.

Conjecture 1 *Consider an $M/G/\infty$ arrival process with On periods being regularly varying $\mathbb{P}[\tau^{on} > x] = l(x)/x^{\alpha}, \alpha > 1$, where α is noninteger. If $0 < c \le r$, then*

$$\mathbb{P}[D_{c(0)} > x] \sim \frac{\Lambda e^{\Lambda \mathbb{E}\tau^{on}}}{e^{\Lambda \mathbb{E}\tau^{on}} - 1} \int_{x/(r + r\Lambda\mathbb{E}\tau^{on} - c)}^{\infty} \mathbb{P}\left[\tau^{on} > u\right] du \quad as \quad x \to \infty. \tag{3.18}$$

Heuristics: We believe that the proof of this theorem can be done in the same spirit as the proof of Theorem 6. Unfortunately, this seems to be very technical and for that reason we do not attempt to provide a rigorous proof. Instead we give the following heuristics. From Theorem 6 it follows that

$$\mathbb{P}[D_c > x] \sim \mathbb{P}[I^{on}(r + r\Lambda\mathbb{E}\tau^{on} - c) > x] \quad \text{as} \quad x \to \infty. \tag{3.19}$$

Based on this, one can expect that

$$\mathbb{P}[D_{c(0)} > x] \sim \mathbb{P}[I^{on}_{(0)}(r + r\Lambda\mathbb{E}\tau^{on} - c) > x] \quad \text{as} \quad x \to \infty,$$

where $I^{on}_{(0)}$ is the residual activity time at time zero, which satisfies

$$\begin{aligned} \mathbb{P}[I^{on}_{(0)} > x] &= 1/\mathbb{E}I^{on} \int_x^{\infty} \mathbb{P}[I^{on} > u] du \\ &\sim 1/\mathbb{E}I^{on} e^{\Lambda\mathbb{E}\tau} \int_x^{\infty} \mathbb{P}[\tau^{on} > u] du \quad \text{as} \quad x \to \infty, \end{aligned} \tag{3.20}$$

where (3.20) follows from Theorem 5. Finally, (3.18) follows by combining (3.20), (3.19), and $\mathbb{E}I^{on} = (e^{\Lambda\mathbb{E}\tau^{on}} - 1)/\Lambda$. $\diamond$

4. Queueing Analysis

We begin this section with a classical result on subexponential asymptotics of a GI/GI/1 queue. The result was obtained by Pakes 1975 (see also Veraverbeke for the random walk approach to this problem). For extensions of this result to Markov-modulated M/G/1 queues see [AHK94], and to Markov-modulated G/G/1 queues (equivalently random walks) see [JLA95g]. Further exstension of these results to more general arrival processes was obtained in [ASC97].

Let $X_n, n \ge 0$, be a sequence of i.i.d. random variables that are driving a queueing process (Lindley's recursion)

$$Q_{n+1} = (Q_n + X_n)^+, \quad n \ge 0, \tag{4.1}$$

where $q^+ = \max(0, q)$. According to the classical result of Loynes' [LOY62] under the stability condition $\mathbb{E}X_n < 0$ this recursion admits a unique stationary solution, and for all initial conditions $\mathbb{P}[Q_n \le x]$ converges to the stationary distribution $\mathbb{P}[Q \le x]$. For the rest of this paper we will assume that all queueing systems under consideration are in their stationary regimes. Let G and G_1 represent the distribution and its integrated tail distribution for X_n, respectively.

Theorem 7 (Pakes) *If $G \in \mathcal{L}$, $G_1 \in \mathcal{S}$ (or $G \in \mathcal{S}_d$), and $\mathbb{E}X_n < 0$, then*

$$\mathbb{P}[Q_n > x] \sim \frac{1}{-\mathbb{E}X_n} \int_x^\infty \mathbb{P}[X_n > u]du \ \ as \ \ t \to \infty.$$

4.1 Fluid Queue: Preliminaries

The physical interpretation for a fluid queue is that at any moment of time t, fluid is arriving to the system with rate a_t and is leaving the system with rate c_t. We term a_t and c_t to be the arrival and the service process, respectively. Then, the evolution of the amount of fluid Q_t (also called queue length) evolves according to

$$dQ_t = (a_t - c_t)dt \ \text{ if } \ Q_t > 0, \ \text{ or } \ a_t > c_t, \tag{4.2}$$

and $dQ_t = 0$, otherwise. It is not very difficult to see that, starting from $Q_0 = 0$, the solution $Q_t, t \geq 0$, to (4.2) is given by

$$Q_t = \sup_{0 \leq u \leq t} \int_u^t (a_v - c_v)dv. \tag{4.3}$$

And if a_t and c_t are stationary, Q_t is equal in distribution to

$$\mathbb{P}[Q_t \leq x] = \mathbb{P}[\sup_{0 \leq u \leq t} W_u \leq x],$$

where $W_t \stackrel{\text{def}}{=} \int_{-t}^0 (a_u - c_u)du, t \geq 0$. Now, whenever the stability condition $\mathbb{E}a_t < \mathbb{E}c_t$ is satisfied (by Birkhoff's Strong Law of Large Numbers), $\mathbb{P}[Q_t \leq x]$ converges to a proper probability distribution, i.e.,

$$\mathbb{P}[Q \leq x] \stackrel{\text{def}}{=} \lim_{t \to \infty} \mathbb{P}[Q_t \leq x] = \mathbb{P}[\sup_{0 \leq u < \infty} W_u \leq x].$$

Further, when the difference process $x_t \stackrel{\text{def}}{=} a_t - c_t$ is driven by a stationary and ergodic point process $\{T_n, -\infty < n < \infty\}$, i.e.,

$$x_t = x_{T_n}, \quad t \in [T_n, T_{n+1}),$$

then the fluid queue process evolves as

$$Q_t = (Q_{T_n-} + (t - T_n)x_{T_n})^+, \quad t \in [T_n, T_{n+1}), \tag{4.4}$$

where $q^+ = \max(q, 0)$. From the recursion above, it is clear that the process Q_t is essentially the same as the $G/G/1$ workload process. Hence, by the fundamental stability theorem of Loynes (see Chapter 2 in [BAB94]) there exists a unique stationary process $\{Q_t^s, -\infty < t < \infty\}$ ($\mathbb{P}[Q_t^s \leq x] = \mathbb{P}[Q \leq x]$) that satisfies (4.4) (or equivalently (4.2)). In the rest of the paper, whenever we refer to Q_t, we will actually mean Q_t^s. (The existence and uniqueness of this stationary solution will be important for establishing the relation between the Palm queue probabilities and the time average probabilities $\mathbb{P}[Q_t \le x]$.)

4.2 Fluid Queue with a Single On-Off Process

Consider a fluid queue with capacity c and an On-Off arrival process with On arrival rate r. In this subsection we assume that Off periods are also general (not necessarily exponential). (A general storage model in a two state random environment was investigated in [KEW92].) Then, if we observe the queue at the beginning of On periods, the queue length Q_n^P evolves as follows (P stands for Palm probability [BAB94]).

$$Q_{n+1}^P = (Q_n^P + (r-c)\tau_n^{on} - c\tau_n^{off})^+, \quad n \ge 0. \tag{4.5}$$

Before we present our result let us state the following well known lemma on long-tailed distributions.

Lemma 2 *Let X and Y be two independent random variables distributed as F and G, respectively. If $X \in \mathcal{L}$, $\mathbb{E}X, \mathbb{E}Y < \infty$, $\mathbb{P}[Y \ge 0] = 1$, then*

(i) $\bar{F}(t) = o(\bar{F}_1(t))$ *as* $t \to \infty$.

(ii) $\mathbb{P}[X - Y > t] \sim \mathbb{P}[X > t] = \bar{F}(t)$ *as* $t \to \infty$.

Proof: Follows easily from the definition of $\mathcal{L}$. ◇

Recall that F and F_1 denote the distribution and the integrated tail distribution of τ^{on}.

Theorem 8 *If $r > c$, $(r-c)\mathbb{E}\tau_{on} < c\mathbb{E}\tau_{off}$, $F \in \mathcal{L}$, and $F_1 \in \mathcal{S}$ (or $F \in \mathcal{S}_d$), then*

$$\mathbb{P}[Q_n^P > x] \sim \frac{r-c}{c\mathbb{E}\tau_{off} - (r-c)\mathbb{E}\tau_{on}} \int_{x/(r-c)}^{\infty} \mathbb{P}[\tau^{on} > u]du \quad as \quad x \to \infty. \tag{4.6}$$

Proof: If we define $X_n = (r-c)\tau_n^{on} - c\tau_n^{off}$, then the theorem can be reduced to Theorem 7. To finish that, by Lemma 2 (ii), $\mathbb{P}[X_n > x] \sim \mathbb{P}[\tau_n^{on}(r-c) > x]$ as $x \to \infty$; also, since long-tailed and subexponential distributions are closed under the tail equivalence (Lemma 5 (ii) (A)), it follows that X_n is long-tailed, with its integrated tail distribution being subexponential. Therefore, Theorem 7 applies. Finally, by using $\mathbb{P}[X_n > x] \sim \mathbb{P}[\tau_n^{on}(r-c) > x]$ in Theorem 7 we obtain (4.6). This finishes the proof. ◇

4.2.1 Time Averages

Here, we will compute queue time averages based on the queue Palm probabilities computed in Theorem 8. For this we need a stationary version a_t^s of the On-Off arrival process a_t. Let $T_n^{on}, -\infty < n < \infty$, be a stationary point process that represents the beginnings of the On-Off periods, with a convention that $T_0^{on} < 0 \le T_1^{on}$. Then, according to [RES96], the random variable T_0^{on} can be represented as $-T_0^{on} = B(\tau_{(0)}^{off} + \tau_0^{on}) + (1-B)\tau_{(0)}^{on}$, where the random variables $B, \tau_{(0)}^{on}, \tau_{(0)}^{off}$ are independent of $\{\tau_n^{on}, \tau_n^{off}, n \le -1\}, \tau_0^{off}$, B is a Bernoulli random variable with $\mathbb{P}[B=0]=1-\mathbb{P}[B=1]=\mathbb{E}\tau^{on}/(\mathbb{E}\tau^{on}+\mathbb{E}\tau^{off})$, and $\tau^{on}_{(0)}, \tau^{off}_{(0)}$, are distributed as integrated tail distributions of τ^{on}, τ^{off}, respectively. Furthermore, the net increment of the load that comes to the queue in the interval $[T_0, 0]$ is given by the following equation

$$\int_{T_0^{on}}^{0}(a_t^s - c)dt = B[(r-c)\tau_0^{on} - c\tau_{(0)}^{off}] + (1-B)(r-c)\tau_{(0)}^{on} \tag{4.7}$$

Theorem 9 *If $r > c$, $(r-c)\mathbb{E}\tau_{on} < c\mathbb{E}\tau_{off}$, $F \in \mathcal{L}$ and $F_1 \in \mathcal{S}$ (or $F \in \mathcal{S}_d$), then*

$$\mathbb{P}[Q_t > x] \sim \mathbb{P}[Q_n^P > x] + \frac{1}{\mathbb{E}\tau^{off} + \mathbb{E}\tau^{on}}\int_{x/(r-c)}^{\infty}\mathbb{P}[\tau^{on} > u]du \tag{4.8}$$

$$\sim K\int_{x/(r-c)}^{\infty}\mathbb{P}[\tau^{on} > u]du \quad \text{as} \quad x \to \infty, \tag{4.9}$$

where

$$K = \frac{r-c}{c\mathbb{E}\tau_{off} - (r-c)\mathbb{E}\tau_{on}} + \frac{1}{\mathbb{E}\tau^{off} + \mathbb{E}\tau^{on}}. \tag{4.10}$$

Remarks: (i) This theorem improves on known results in [RES96, CHW95] that were obtained under the assumption of τ^{on} being regularly varying; (ii) The following proof can be carried out to establish the relationship between the Palm and time averages in much more general settings like semi-Markov fluid queues.

Proof: Let $\{Q_t, -\infty < t < \infty\}$ be a unique stationary solution to (4.5). Then, by using equation (4.7), and the independence of B of $Q_{T_0}, \tau_{(0)}^{off}, \tau_{(0)}^{on}, \tau_0^{off}$, we obtain

$$\begin{aligned}
\mathbb{P}[Q_0 > x] &= \mathbb{P}[Q_0 > x, B=1] + \mathbb{P}[Q_0 > x, B=0] \\
&= \mathbb{P}[Q_{T_0} + \tau_0^{on}(r-c) - c\tau_{(0)}^{off} > x, B=1] + \mathbb{P}[Q_{T_0} + (r-c)\tau_{(0)}^{on} > x, B=0] \\
&= \frac{\mathbb{E}\tau^{off}}{\mathbb{E}\tau^{on} + \mathbb{E}\tau^{off}}\mathbb{P}[Q_{T_0} + \tau_0^{on}(r-c) - c\tau_{(0)}^{off} > x] \qquad (4.11) \\
&\qquad + \frac{\mathbb{E}\tau^{on}}{\mathbb{E}\tau^{on} + \mathbb{E}\tau^{off}}\mathbb{P}[Q_{T_0} + (r-c)\tau_{(0)}^{on} > x].
\end{aligned}$$

(Note that $Q_0^P \equiv Q_{T_0}$.) Since Q_{T_0} and $\tau_{(0)}^{on}$ are independent, subexponential, and have asymptotically proportional (equivalent) tails, by applying Lemma 5 (ii) (A) it follows that

$$\mathbb{P}[Q_{T_0} + (r-c)\tau_{(0)}^{on} > x] \sim \mathbb{P}[Q_{T_0} > x] + \mathbb{P}[(r-c)\tau_{(0)}^{on} > x] \quad \text{as} \quad x \to \infty. \tag{4.12}$$

The independence of τ_0^{on} and $\tau_{(0)}^{off}$, and $\tau_0^{on} \in \mathcal{L}$, by Lemma 2 (ii) gives $\mathbb{P}[\tau_0^{on}(r-c) - c\tau_{(0)}^{off} > x] \sim \mathbb{P}[\tau_0^{on}(r-c) > x] = o(\mathbb{P}[Q_{T_0} > x])$ as $x \to \infty$. Subsequently, by applying Lemma 5 (i) (A)

$$\mathbb{P}[Q_{T_0} + \tau_0^{on}(r-c) - c\tau_{(0)}^{off} > x] \sim \mathbb{P}[Q_{T_0} > x] \quad \text{as} \quad x \to \infty. \tag{4.13}$$

Finally, by replacing asymptotic relations (4.12), (4.13), in (4.11), we obtain (4.8); combination of (4.6) and (4.8) gives (4.9). This completes the proof. $\diamond$

4.3 Multiplexing a Long-Tailed Process with Exponential Processes

In this section we consider multiplexing one long-tailed On-Off process with exponential processes (see Definition 4 below) in a fluid queue. The important conclusion to be drawn is that this queueing system is asymptotically interchangeable with a queueing system in which the On-Off process is arriving alone and the exponential processes are replaced by their mean values. To reach this conclusion we need the following definition.

Definition 3 *A distribution function F is* intermediate regular varying $F \in \mathcal{IR}$ *if*

$$\lim_{\delta \downarrow 1} \liminf_{t \to \infty} \frac{\bar{F}(\delta t)}{\bar{F}(t)} = 1.$$

Remark: For recent results on distributions of intermediate regular variation we refer the reader to [CLI94]. Some basic properties of $\mathcal{IR}$ are: $\mathcal{IR} \subset \mathcal{S}$; $\mathcal{R} \subset \mathcal{IR}$. Also, it is not very difficult to see that $\mathcal{IR} \subset \mathcal{S}_d$. Therefore, all of the results that we have obtained up to now apply for $\mathcal{IR}$. In addition, directly from the definition it can be shown that $F \in \mathcal{IR}, \int_0^\infty \bar{F}(t)dt < \infty, \Rightarrow F_1 \in \mathcal{IR}$.

Under the general large deviation Gärtner-Ellis conditions (see [WES95]) on the arrival process, it can be proved that the queue length distribution is exponentially bounded. To avoid stating Gärtner-Ellis conditions, we will define an arrival process e_t to be *exponential*, if whenever this process is fed into a constant server fluid queue, the queue length distribution is exponentially bounded.

Definition 4 *We say that a stationary and ergodic arrival process e_t is* exponential *if for any server capacity $c > \mathbb{E}e_t$ there exists $K \equiv K(c)$ and $\delta \equiv \delta(c) > 0$ such that*

$$\mathbb{P}[\sup_{t \geq 0} \int_{-t}^{0} (e_u - c)du > x] \leq K e^{-\delta x}.$$

Remark: The main examples when the conditions of this definition are satisfied (i.e. Gartnër-Ellis conditions hold) are finite state space Markov chains or processes. Also, in terms of the On-Off processes the conditions will hold whenever the distribution of On periods is exponentially bounded and Off periods have a finite mean.

Recall that F and F_1 represent the distribution and the integrated tail distribution of an On period, respectively.

Theorem 10 *Consider a single server queue with a capacity c and two independent arrival streams e_t and a_t. Assume that e_t is an exponential process (as in Definition 4) and a_t is an On-Off process with rate r, $F \in \mathcal{IR}$, and generally distributed Off periods with a finite mean. If $\mathbb{E}(e_t + a_t) < c$, $r > c' \stackrel{\text{def}}{=} c - \mathbb{E}e_t$, then the queue asymptotics of this queueing system is equal to the queue asymptotics in which only the On-Off process arrives and the server capacity is replaced by c', i.e., it is given by equation (4.9) in which c is replaced by c'.*

Remarks: (i) In Boxma (1996) [BOX96, BOX97], a precise asymptotics of the embeded queue distribution was obtained for multiplexing On-Off sources one of which had regularly varying On periods while the others had exponentially distributed On periods. A similar setting with intermediately varying On periods was investigated in [RSS97]. However, in both papers the equivalence relation of the original system to the system in which the exponential process is replaced by its mean has not been observed. (ii) The assumption of e_t being exponential can be weakened to $\mathbb{P}[\sup_{t\geq 0}\int_{-t}^{0}(e_u - c)du > x] = o(F_1(x))$, for all $c > \mathbb{E}e_t$, without any changes in the proof.

Proof: *Upper bound.* For any $\epsilon > 0$, we can make the following decomposition

$$\begin{aligned} W_t = \int_{-t}^{0}(e_u + a_u - c)du &= \int_{-t}^{0}(e_u - \mathbb{E}e_0 - \epsilon)du + \int_{-t}^{0}(a_u - (c - \mathbb{E}e_0 - \epsilon))du \\ &\stackrel{\text{def}}{=} W_t^e + W_t^s. \end{aligned}$$

Call $c_\epsilon \stackrel{\text{def}}{=} c - \mathbb{E}e_0 - \epsilon$. Observe that for all sufficiently small $\epsilon > 0$, such that $\mathbb{E}(e_t + a_t) + \epsilon < c$ and $c - \mathbb{E}e_t - \epsilon > 0$, $\mathbb{P}[\sup_{t\geq 0} W_t^s \leq x]$ represents the queue length distribution in a stable On-Off queue with arrival process a_t, and service capacity c_ϵ. By using Theorem 9 this implies that $\mathbb{P}[\sup_{t\geq 0} W_t^s \leq x] \in \mathcal{IR}$. Further,

$$\begin{aligned} \mathbb{P}[Q_t > x] &= \mathbb{P}[\sup_{t\geq 0}(W_t^e + W_t^s) > x] \\ &\leq \mathbb{P}[\sup_{t\geq 0} W_t^e + \sup_{t\geq 0} W_t^s > x] \\ &\sim \mathbb{P}[\sup_{t\geq 0} W_t^s > x] \quad \text{as} \quad x \to \infty, \end{aligned} \tag{4.14}$$

where (4.14) follows from Lemma 5 (i) (A), since $\sup_{t\geq 0} W_t^e$ and $\sup_{t\geq 0} W_t^s$ are two independent random variables, with $\mathbb{P}[\sup_{t\geq 0} W_t^e > x] = o(\mathbb{P}[\sup_{t\geq 0} W_t^s > x])$ as $x \to \infty$. Now,

$$\lim_{x\to\infty} \frac{\mathbb{P}[\sup_{t\geq 0} W_t^s > x]}{\int_{x/(r-c_\epsilon)}^{\infty} \mathbb{P}[\tau^{on} > u]du} = K_\epsilon, \tag{4.15}$$

where K_ϵ is given by equation (4.10) in Theorem 9, with c_ϵ in place of c. Subsequently, this leads to

$$\limsup_{x\to\infty} \frac{\mathbb{P}[Q_t > x]}{\int_{x/(r-c')}^{\infty} \mathbb{P}[\tau^{on} > u]du} \leq K_\epsilon \limsup_{x\to\infty} \frac{\int_{x/(r-c_\epsilon)}^{\infty} \mathbb{P}[\tau^{on} > u]du}{\int_{x/(r-c')}^{\infty} \mathbb{P}[\tau^{on} > u]du}. \tag{4.16}$$

Finally, if we pass $\epsilon \to 0$ in (4.16), we obtain

$$\limsup_{x\to\infty} \frac{\mathbb{P}[Q_t > x]}{\int_{x/(r-c')}^{\infty} \mathbb{P}[\tau^{on} > u]du} \leq K', \tag{4.17}$$

where K', similarly as K_ϵ, is given by equation (4.10) in Theorem 9, with c' in place of c. This concludes the upper bound proof.

Lower bound. For the lower bound we consider a different decomposition of W_t, i.e., we redefine W_t^e and W_t^s as follows ($\epsilon > 0$)

$$\begin{aligned} W_t = \int_{-t}^{0} (e_u + a_u - c) du &= -\int_{-t}^{0} (\mathbb{E}e_0 - \epsilon - e_u) du + \int_{-t}^{0} (a_u - (c - \mathbb{E}e_0 + \epsilon)) du \\ &\stackrel{\text{def}}{=} -W_t^e + W_t^s. \end{aligned}$$

Also, redefine $c_\epsilon \stackrel{\text{def}}{=} c - \mathbb{E}e_0 + \epsilon$. Similarly, as in the upper bound case, for $\epsilon < \mathbb{E}e_0 + \mathbb{E}a_0 - c$, $\mathbb{P}[\sup_{t\geq 0} W_t^s \leq x]$ represents a queue length distribution in a stable On-Off queue with arrival process a_t, and service capacity c_ϵ. Hence, $\mathbb{P}[\sup_{t\geq 0} W_t^s \leq x] \in \mathcal{IR}$. Further,

$$\begin{aligned} \mathbb{P}[Q_t > x] &= \mathbb{P}[\sup_{t\geq 0}(-W_t^e + W_t^s) > x] \\ &\geq \mathbb{P}[\inf_{t\geq 0} -W_t^e + \sup_{t\geq 0} W_t^s > x] \qquad (4.18) \\ &= \mathbb{P}[-\sup_{t\geq 0} W_t^e + \sup_{t\geq 0} W_t^s > x] \\ &\sim \mathbb{P}[\sup_{t\geq 0} W_t^s > x] \quad \text{as} \quad x \to \infty, \qquad (4.19) \end{aligned}$$

where in (4.18) we use that for any two functions $f, g : \mathbb{R} \to \mathbb{R}$, $\sup(f+g) \geq \inf f + \sup g$; asymptotics in (4.19) follows from the independence of a_t and e_t, $\mathbb{P}[\sup_{t\geq 0} W_t^s \leq x] \in \mathcal{L}$, $\mathbb{P}[\sup_{t\geq 0} W_t^e < \infty] = 1$, and Lemma 2 (ii). Subsequently, (4.19) leads to

$$\liminf_{x\to\infty} \frac{\mathbb{P}[Q_t > x]}{\int_{x/(r-c')}^{\infty} \mathbb{P}[\tau^{on} > u] du} \geq K_\epsilon \liminf_{x\to\infty} \frac{\int_{x/(r-c_\epsilon)}^{\infty} \mathbb{P}[\tau^{on} > u] du}{\int_{x/(r-c')}^{\infty} \mathbb{P}[\tau^{on} > u] du}, \qquad (4.20)$$

where again K_ϵ is computed from Theorem 9 with c_ϵ in place of c. Consequently, using the same arguments as in (4.17) we arrive at

$$\liminf_{x\to\infty} \frac{\mathbb{P}[Q_t > x]}{\int_{x/(r-c')}^{\infty} \mathbb{P}[\tau^{on} > u] du} \geq K', \qquad (4.21)$$

which together with (4.17) concludes the proof. $\diamond$

4.4 Subexponential M/G/∞ Arrival Process

In the next theorem we obtain a tight lower bound for the fluid queue asymptotics with M/G/∞ arrivals. For this fluid queue we denote its queue content process as Q_t^∞.

4.4.1 Lower Bound

Theorem 11 *Let $\rho \stackrel{\text{def}}{=} \mathbb{E}A_t^{\infty,s} = \Lambda r \mathbb{E}\tau^{on} < c$. If $r + \rho > c$, and $\tau^{on} \in \mathcal{IR}$, then*

$$\liminf_{x\to\infty} \frac{\mathbb{P}[Q_t^\infty > x]}{\int_{x/(r+\rho-c)}^{\infty} \mathbb{P}[\tau^{on} > u] du} \geq \frac{\Lambda r}{c - \rho}.$$

Proof: Let $\underline{A}^{\infty,y} = \{\underline{A}_t^{\infty,y}, t \geq 0\}$, $\overline{A}^{\infty,y} = \{\overline{A}_t^{\infty,y}, t \geq 0\}$, $y > 0$, be two independent M/G/∞ type processes with Poisson arrival rate, and On distributions, respectively given as $\underline{\Lambda}^y = \mathbb{P}[\tau^{on} \leq y]\Lambda$, $\overline{\Lambda}^y = \mathbb{P}[\tau^{on} > y]\Lambda$, $F_{\underline{y}}(x) = \mathbb{P}[\tau^{on} \leq x]/\mathbb{P}[\tau^{on} \leq y], 0 \leq x \leq y$, $F_{\overline{y}}(x) = \mathbb{P}[y < \tau^{on} \leq x]/\mathbb{P}[\tau^{on} > y], x > y$; all three processes are assumed to have the same parameter r. Now, we claim that

$$A^{\infty} \stackrel{d}{=} \underline{A}^{\infty,y} + \overline{A}^{\infty,y}, \tag{4.22}$$

where $\stackrel{d}{=}$ stands for equality in distribution. Observe that, for a fixed parameter r, any M/G/∞ process is uniquely defined with Possion arrival times $\{T_n\}$, and the lengths of On periods $\{\tau_n^{on}\}$; likewise, the pair $\{T_n\}$, $\{\tau_n^{on}\}$ uniquely defines a compound Poisson process (see [CIN75], pp. 90), that is a piecewise constant process with Poisson jump times $\{T_n\}$, and jump sizes $\{\tau_n^{on}\}$. Hence, proving (4.22) is equivalent to

$$Z \stackrel{d}{=} \underline{Z}^y + \overline{Z}^y, \tag{4.23}$$

where $Z, \underline{Z}^y, \overline{Z}^y$, are the compound Poisson processes corresponding to $A^{\infty}, \underline{A}^{\infty,y}, \overline{A}^{\infty,y}$, respectively. Since $Z, \underline{Z}^y, \overline{Z}^y$, are processes with stationary independent increments, (4.23) is equivalent to

$$Z_t \stackrel{d}{=} \underline{Z}_t^y + \overline{Z}_t^y, \quad t \geq 0. \tag{4.24}$$

Evidently (4.24) is implied by

$$\begin{aligned} \mathbb{E}e^{-s(\underline{Z}_t^y + \overline{Z}_t^y)} &= \mathbb{E}e^{-s\underline{Z}_t^y}\mathbb{E}e^{-s\overline{Z}_t^y} \\ &= \exp\left(-t\underline{\Lambda}^y \int_0^{\infty}(1-e^{-su})dF_{\underline{y}}(u)\right)\exp\left(-t\overline{\Lambda}^y \int_0^{\infty}(1-e^{-su})dF_{\overline{y}}(u)\right) \qquad (4.25) \\ &= \exp\left(-t\Lambda \int_0^{y}(1-e^{-su})dF(u)\right)\exp\left(-t\Lambda \int_y^{\infty}(1-e^{-su})dF(u)\right) \\ &= \exp\left(-t\Lambda \int_0^{\infty}(1-e^{-su})dF(u)\right) = \mathbb{E}e^{-sZ_t}; \qquad (4.26) \end{aligned}$$

equalities (4.25), (4.26), are well known expressions for compound Poisson processes (see [CIN75], pp. 94). This proves (4.22).

Note that process $\underline{A}_t^{\infty,y}$ has bounded On periods, and in conclusion, it is an *exponential* process (i.e., it satisfies Definition 4). Also, $\overline{A}_t^{\infty,y} \geq r1(\overline{A}_t^{\infty,y} > 0)$. Therefore, $\mathbb{P}[Q_t^{\infty} > x]$ is stochastically larger than a queueing process $\underline{Q}_t^{\infty}$ obtained by feeding $\underline{A}_t^{\infty,y} + r1(\overline{A}_t^{\infty,y} > 0)$ into it. Let $Q_t^{a^y}$ be a queueing process with a subexponential On-Off arrival process $a_t^y \stackrel{\text{def}}{=} r1(\overline{A}_t^{\infty,y} > 0)$, and a server capacity $c_y = c - \mathbb{E}\underline{A}_t^{\infty,y} = c - r\mathbb{P}[\tau^{on} \leq y]\Lambda\mathbb{E}\tau^{on}$. Here, by Theorem 10, we obtain

$$\mathbb{P}[Q_t^{\infty} > x] \geq \mathbb{P}[\underline{Q}_t^{\infty} > x] \sim \mathbb{P}[Q_t^{a^y} > x] \text{ as } x \to \infty. \tag{4.27}$$

In addition, a_t^y has Off period $I^{off,y}$ exponentially distributed with parameter $\overline{\Lambda}^y$, On period $I^{on,y}$ with mean $\mathbb{E}I^{on,y} = \frac{1}{\overline{\Lambda}^y}(e^{\overline{\Lambda}^y \mathbb{E}\tau^{on,y}} - 1) = \frac{1}{\overline{\Lambda}^y}(e^{\Lambda \mathbb{E}\tau^{on} 1(\tau^{on} > y)} - 1)$, and asymptotics (by Theorem 5)

$$\begin{aligned} \mathbb{P}[I^{on,y} > t] &\sim e^{\overline{\Lambda}^y \mathbb{E}\tau^{on,y}} \mathbb{P}[\tau^{on,y} > t] \ \text{ as } \ t \to \infty \\ &= \frac{e^{\Lambda \mathbb{E}\tau^{on} 1(\tau^{on} > y)}}{\mathbb{P}[\tau^{on} > y]} \mathbb{P}[\tau^{on} > t]. \end{aligned} \tag{4.28}$$

Hence, Theorem 9 and (4.28) lead to

$$\mathbb{P}[Q_t^{a^y} > x] \sim K_y \int_{x/(r-c_y)}^{\infty} \mathbb{P}[\tau^{on} > u] du \ \text{ as } \ x \to \infty, \tag{4.29}$$

where

$$K_y \stackrel{\text{def}}{=} \left(\frac{r - c_y}{c_y \mathbb{E}I^{off,y} - (r - c_y)\mathbb{E}I^{on,y}} + \frac{1}{\mathbb{E}I^{off,y} + \mathbb{E}I^{on,y}} \right) \frac{e^{\Lambda \mathbb{E}\tau^{on} 1(\tau^{on} > y)}}{\mathbb{P}[\tau^{on} > y]}.$$

Combining (4.27) and (4.29) produces

$$\liminf_{x \to \infty} \frac{\mathbb{P}[Q_t^{\infty} > x]}{\int_{x/(r-c')}^{\infty} \mathbb{P}[\tau^{on} > u] du} \geq K_y \liminf_{x \to \infty} \frac{\int_{x/(r-c_y)}^{\infty} \mathbb{P}[\tau^{on} > u] du}{\int_{x/(r-c')}^{\infty} \mathbb{P}[\tau^{on} > u] du}, \tag{4.30}$$

where $c' \stackrel{\text{def}}{=} c - \rho$. Here, observe that $K_y \to \frac{\Lambda r}{c'}$ as $y \to \infty$, and, by assumption $F \in \mathcal{IR}$,

$$\lim_{y \to \infty} \liminf_{x \to \infty} \frac{\int_{x/(r-c_y)}^{\infty} \mathbb{P}[\tau^{on} > u] du}{\int_{x/(r-c')}^{\infty} \mathbb{P}[\tau^{on} > u] du} = 1.$$

Finally, if we take the limit with respect to y in (4.30), we obtain the statement of the theorem. $\diamond$

4.4.2 Precise Queue Asymptotics

Let $Q_n^{P,\infty}$ be the queue size observed at the beginning of the nth activity period of the M/G/∞ arrival process.

Theorem 12 *Let $\rho = \mathbb{E}A_t^{s,\infty} = \Lambda r \mathbb{E}\tau^{on} < c$. If $c \leq r$, and τ^{on} is regularly varying with noninteger exponent $\alpha > 1$, then*

$$\lim_{x \to \infty} \frac{\mathbb{P}[Q_t^{P,\infty} > x]}{\int_{x/(\rho+r-c)}^{\infty} \mathbb{P}[\tau^{on} > u] du} = \Lambda \left(\frac{r}{c - \rho} - 1 \right).$$

Proof: Proof follows directly from Theorem 7 and Theorem 6, by taking $X_n \stackrel{\text{def}}{=} D_{c,n} - cI_n^{off}$, observing that (by Lemma 2) $\mathbb{P}[X_n > x] \sim \mathbb{P}[D_{c,n} > x]$ as $x \to \infty$, and $\mathbb{E}X_n = e^{\Lambda \mathbb{E}\tau^{on}}(\Lambda r \mathbb{E}\tau^{on} - c)/\Lambda$. $\diamond$

Theorem 13 *Let* $\rho = \mathbb{E}A_t^{s,\infty} = \Lambda r \mathbb{E}\tau^{on} < c$. *If Conjecture 1 holds,* $c \leq r$, *and* τ^{on} *is regularly varying with noninteger exponent* $\alpha > 1$, *then*

$$\lim_{x\to\infty} \frac{\mathbb{P}[Q_t^\infty > x]}{\int_{x/(\rho+r-c)}^{\infty} \mathbb{P}[\tau^{on} > u]du} = \Lambda \frac{r}{c-\rho}.$$

Proof: This theorem follows from Theorem 13, Conjecture 1, and exactly the same arguments as in the proof of Theorem 9. We omit the details. $\diamond$

Remark: The asymptotic result in this theorem is the same as the lower bound obtained in Theorem 11.

4.5 Finite Number of Subexponential On-Off Processes

For the finite number of long-tailed On-Off processes we can easily obtain the upper bound given by Theorem 18, Appendix C. Similarly, utilization of $A_t^N \geq r1(A_t^N > 0)$, Theorem 3, and Theorem 9, can easily produce a lower bound. Unfortunately, these bounds are very weak. For that reason, we resort back to Theorems 11, 12, and 13, and suggest an approximation for the finite number of processes.

4.5.1 M/G/∞ Approximation: Simulation Results

Based on Theorems 11, 12, and 13, we suggest that the queueing probabilities obtained by multiplexing N long-tailed On-Off processes $a_t^i, 1 \leq i \leq N$, are approximated as

$$\mathbb{P}[Q_t^N > x] \approx \frac{\Lambda_N r}{c_N} \int_{x/(r-c_N)}^{\infty} \mathbb{P}[\tau^{on} > u]du, \tag{4.31}$$

where $c_N \stackrel{\text{def}}{=} c - N\mathbb{E}a_t^i$, and $\Lambda_N \stackrel{\text{def}}{=} N\mathbb{E}a_t^i/(r\mathbb{E}\tau^{on})$. We term this approximation an M/G/∞ approximation. This approximation is to be used when the queue is stable and $r + (N-1)\mathbb{E}a_t^i > c$ *is satisfied.*

For simulation purposes we consider a discrete time "fluid" queue. Correspondingly, we replace exponential Off periods, with geometrically distributed random variables $\mathbb{P}[\tau^{off} = t] = p(1-p)^{t-1}, t = 1, 2, 3, \ldots$. For On periods we consider the Pareto family $\mathbb{P}[\tau^{on} \geq t] = 1/t^\alpha, t = 1, 2, \ldots, \alpha > 0$. Here, for the discrete Pareto case we use

$$\mathbb{P}[Q_t^N = x] \approx \frac{\Lambda_N r}{c_N}(r - c_N)^{\alpha-1}x^{-\alpha}, \tag{4.32}$$

where c_N, and Λ_N are as defined earlier.

The efficacy of the approximation (4.32) is tested in the following simulation experiments. For all experiments we fixed $p = 0.05$. In each experiment, the number of simulated On-Off intervals was at least 10^8, or equivalently, the length of the simulated aggregated process was $\approx 2 \times 10^9$. This was necessary to ensure the desired precision of the simulation outcomes.

Experiment 1 Choose $\alpha = 2, r = 1, c = 1$. This gives $\mathbb{E}\tau^{on} = 1.645$, and $\mathbb{E}a_t^i = 0.0759$. Then, for $N = 8, 10, 12$, processes, the approximations are given by β/x^2, $\beta = 0.573$, 1.46, 5.73, respectively. Comparisons between the simulation results and these approximations are given in Figure 1.

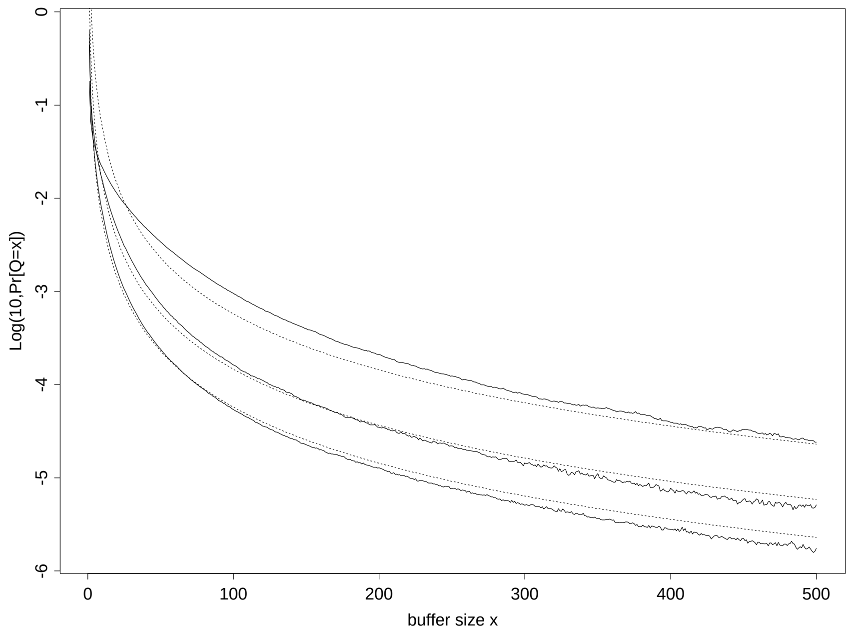

Figure 1: Illustration for Example 1.

Experiment 2 Choose $\alpha = 2.5, r = 1, c = 1$. This gives $\mathbb{E}\tau^{on} = 1.341$, and $\mathbb{E}a_t^i = 0.0628$. Then, for $N = 10, 12, 14$, processes, the approximations are given by $\beta/x^{2.5}$, $\beta = 0.63$, $1.50, 4.51$, respectively. Excellent agreement between the simulation results and these approximations is displayed in Figure 2.

Experiment 3 Choose $\alpha = 3, r = 2, c = 3$. This gives $\mathbb{E}\tau^{on} = 1.202$, and $\mathbb{E}a_t^i = 0.113$. Then, for $N = 20, 25$, processes, the approximations are given by $\beta/x^3, \beta = 4.14, 48.04$, respectively. Again, the desirable closeness between the simulation results and the approximations is represented in Figure 3. It is interesting to observe that in this case the peak rate of each individual process is smaller than the capacity of the server.

Notice that in the last experiment (3), for the case of $N = 20$ processes, the probabilities are very small ($\approx 10^{-8}$). Hence, in order to achieve reasonable simulation accuracy, we had to choose a very large number (10^9) of simulated On-Off intervals. This means that the aggregate process was approximately 2×10^{10} samples long. The simulation of this case took *77 hours*!!! on a modern (200 MIPS) IBM workstation. *On the other hand, it is needless to say that the evaluation of (4.32), or (4.31), only takes a negligible amount of time!*

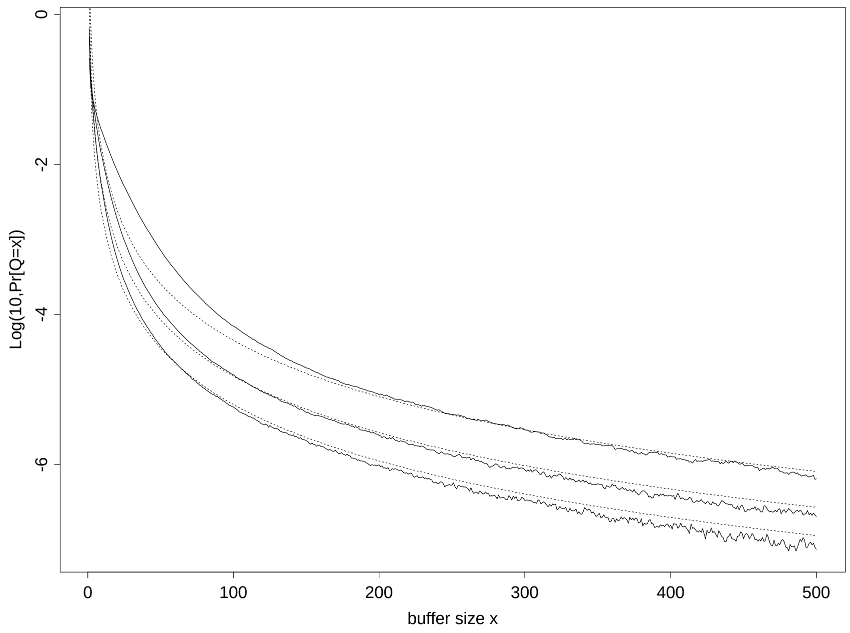

Figure 2: Illustration for Example 2.

5. Conclusion

In this paper we established a precise asymptotic characterization of the activity period of an arrival process obtained by multiplexing On-Off processes with exponential Off periods and subexponential On periods. This characterization was done both for a finite number of processes as well as for the limiting $M/G/\infty$ case.

For a simple subexponential On-Off fluid flow queue we obtained a precise asymptotic relation between the Palm queue distribution and the time average queue distribution. Furthermore, exponential processes, when multiplexed with a subexponential On-Off process, were shown to contribute to the large buffer asymptotics only through their mean value.

In the limiting $M/G/\infty$ case (e.g., large number of subexponential On-Off processes) with regularly varying On periods (with noninteger exponents) we obtained a precise queue asymptotics observed at the beginning of the arrival process activity periods. The asymptotic (time average) queue lower bound was derived under more general assumptions of intermediately varying On periods.

Based on these asymptotic results, a computationally efficient approximation was suggested for the large buffer probabilities of finitely many subexponential On-Off processes. The accuracy of this approximation was verified using extensive simulation experiments.

The results in this paper bring us closer to understanding the subexponential queueing asymptotics of multiplexed long-tailed processes. The precision and low computational complexity of the $M/G/\infty$ approximation has a practical impact on improving the efficacy of ATM admission controllers.

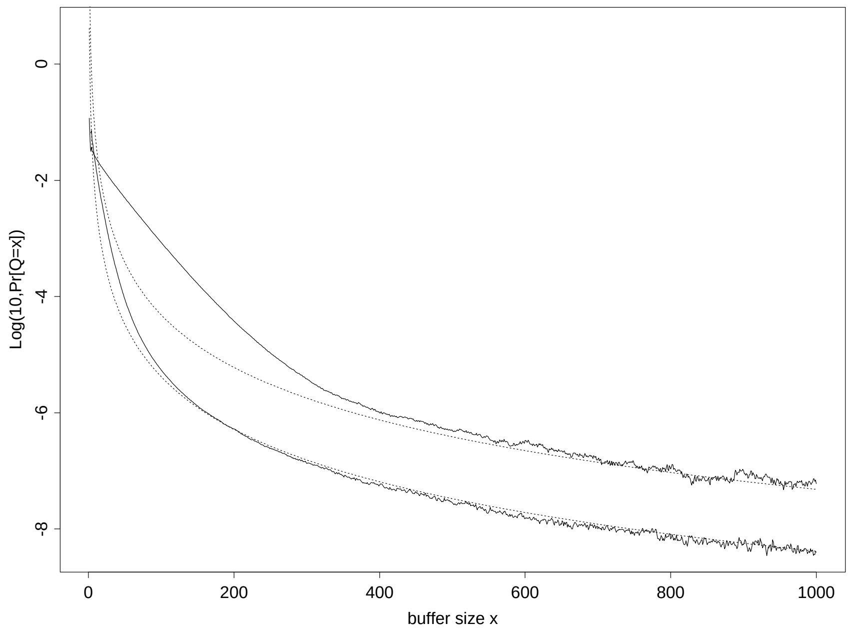

Figure 3: Illustration for Example 3.

Acknowledgements

The authors are very grateful to Onno Boxma for the preprint of his paper, and to Claudia Klüppelberg for her helpful comments.

6. Appendix

A: Basic Results on Subexponential and Long-Tailed Distributions

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

Lemma 3 *(Athreya and Ney, [ATN72])* $\mathcal{S} \subset \mathcal{L}$.

Lemma 4 *If* $F \in \mathcal{L}$ *then* $(1 - F(x))e^{\alpha x} \to \infty$ *as* $x \to \infty$, *for all* $\alpha > 0$.

Note: Lemma 4 clearly shows that for long-tailed distributions Cramér type conditions are not satisfied.

The proof of the following result can be found in Embrechts, Goldie and Veraverbeke [EGV79].

Lemma 5 *Let* $F \in \mathcal{S}$. *Then,*

(i) *If* G *is a probability distribution such that* $\bar{G}(x) = o(\bar{F}(x))$ *as* $x \to \infty$, *then* $\overline{F * G}(x) \sim \bar{F}(x)$ *as* $x \to \infty$.

(ii) *If* $\lim_{x\to\infty} \bar{G}(x)/\bar{F}(x) = c \in (0, \infty)$, *where* G *is a distribution function on* $[0, \infty)$, *then* $G \in \mathcal{S}$, *and* $\overline{F * G}(x) \sim \bar{F}(x) + \bar{G}(x)$ *as* $x \to \infty$.

The next result is due to Athreya and Ney (see [ATN72], pp. 147-150).

Lemma 6 *If* $F \in \mathcal{S}$, *then*

(i) $\overline{F^{*n}}(x)/\bar{F}(x) \to n$ *as* $x \to \infty$, *for all* $n \in \mathbb{N}$.

(ii) For each $\epsilon > 0$ *there exists a constant* $C_\epsilon (< \infty)$ *such that* $\overline{F^{*n}}(x) \leq C_\epsilon (1+\epsilon)^n \bar{F}(x)$ *for all* x *and* n.

This lemma directly gives the asymptotics of a renewal measure with the following Pollazcek-Khintchine representation

$$G(x) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} \gamma^n F^{*n}(x). \tag{6.1}$$

Using dominated convergence and the previous lemma it is easy to prove the following very useful theorem.

Theorem 14 *If* $F \in \mathcal{S}$, *and* $-1 < \gamma < 1$, *then*

$$\lim_{x\to\infty} \frac{\bar{G}(x)}{\bar{F}(x)} = \frac{\gamma}{(1-\gamma)^2}.$$

Proof: Follows easily from Lemma 6 and dominated convergence (see [AHK94]).

Sometimes when F is absolutely continuous, i.e., F has a density function f, it is of interest to calculate the density g of G. If we take a derivative in (6.1) with respect to x we obtain

$$g(x) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} \gamma^n f^{\otimes n}(x), \tag{6.2}$$

where $f^{\otimes n}(x)$ is the density of F^{n*}; $f^{\otimes 2}(x) = \int_0^x f(x-u)f(u)du$, and $f^{\otimes(n+1)}(x) = \int_0^x f^{\otimes n}(x-u)f(u)du$. The investigation of the asymptotics of g requires the investigation of the asymptotics of $f^{\otimes n}(x)$. This has motivated the introduction of subexponential density functions in [KLU89].

Definition 5 *A function $f : \mathbb{R}_+ \to \mathbb{R}_+$ such that $f(x) > 0$ on $[A, \infty)$ for some $A \in \mathbb{R}_+$ belongs to the class $\mathcal{S}_d$ if f is long-tailed, and*

$$\lim_{x\to\infty} \frac{f^{\otimes 2}(x)}{f(x)} = 2d.$$

Then, equivalent results to Lemma 6, and Theorem 14 (in its most general form) were obtained in Theorem 3.2, [KLU89]. For convenience we state here simplified versions of these results.

Lemma 7 *Let $f \in \mathcal{S}_d$, and $\int_0^\infty f(x)dx = 1$. Then, both (i) and (ii) of Lemma 6 are true with $\bar{F}$ replaced by f and $\overline{F^{*n}}(x)$ replaced by $f^{\otimes n}$.*

Theorem 15 *If $f \in \mathcal{S}_d$, with $\int_0^\infty f(x)dx = 1$, and $-1 < \gamma < 1$, then*

$$g(x) \sim \frac{\gamma}{(1-\gamma)^2} f(x) \quad as \ \ x \to \infty.$$

An extensive treatment of subexponential distributions (and further references) can be found in [CLI86, KLU88]. For a recent survey of the application of subexponential distributions in queueing theory the reader is referred to [ASM96].

B: Proof of Theorem 6

As we have already mentioned the proof of this result is based on Karamata's Tauberian/Abelian theorem for distribution functions of regular variation. This theorem relates the tail behavior of a distribution function to the asymptotic behavior of its Laplace transform at the origin. For convenience we state the following result due to Bingham and Doney [BID74] ([BGT87], pp. 333).

Let F be a distribution function on $[0, \infty)$, and let $\tilde{F}(s)$ be its Laplace-Stieltjes transform. Denote by $m_n = \mathbb{E}X^n = \int_{[0,\infty)} x^n dF(x), n = 0, 1, \ldots$. When $m_n < \infty$, $\tilde{F}(s)$ may be expanded in a Taylor series as far as the s^n term:

$$\tilde{F}(s) = \sum_{k=0}^{n} m_k(-s)^k/k! + o(s^n) \ \text{ as } \ s \downarrow 0.$$

To compare the tail behavior of F with the behavior of $\tilde{F}$ at the origin, one needs to eliminate the Taylor polynomial $\sum_{k=0}^{n} m_k(-s)^k/k!$. This may be done by subtraction or repeated differentiation, i.e., let

$$\begin{aligned} f_n(s) &\stackrel{\text{def}}{=} (-1)^{n+1}\left\{\tilde{F}(s) - \sum_{k=0}^{n} m_k(-s)^k/k!\right\} \\ g_n(s) &\stackrel{\text{def}}{=} d^n f_n(s)/ds^n = m_n - (-1)^n \tilde{F}^{(n)}(s), \quad n \geq 0, \end{aligned}$$

where $\tilde{F}^{(n)}(s)$ denotes the nth derivative of $\tilde{F}(s)$; in general for any function G we will use $G^{(n)}$ to denote its nth derivative.

Theorem 16 *Let l be a slowly varying function, $n \in \mathbb{N}_0$, and $\alpha = n + \beta, 0 < \beta < 1$ (α-noninteger). Then, the following are equivalent*

$$f_n(s) \sim s^\alpha l(1/s) \ \text{ as } \ s \downarrow 0, \tag{6.3}$$

$$g_n(s) \sim \frac{\Gamma(\alpha+1)}{\Gamma(\beta+1)} s^\beta l(1/s) \ \text{ as } \ s \downarrow 0, \tag{6.4}$$

$$(-1)^{n+1}\tilde{F}^{(n+1)}(s) \sim \frac{\Gamma(\alpha+1)}{\Gamma(\beta)} s^{\beta-1} l(1/s) \ \text{ as } \ s \downarrow 0, \tag{6.5}$$

$$1 - F(x) \sim \frac{(-1)^n}{\Gamma(1-\alpha)} x^{-\alpha} l(x) \ \text{ as } \ x \to \infty, \tag{6.6}$$

where Γ stands for the gamma function.

Now, in order to determine the tail behavior of $D_{c,n}$, we will investigate asymptotics of its LS transform at the origin. Note that without loss of generality we can set $r = 1$, since $D_{c,r} = rD_{c/r,1}$; $D_{c,r}$ stands for D_c where the arrival process has parameter r. To simplify the notation we use $D \equiv D_c$, $\tau \equiv \tau^{on}$. Slight modification of Theorem 2.10, [COH74], reads as follows.

Theorem 17 *Let $0 \leq c \leq 1$ ($r = 1$), $\mathbb{E}\tau < \infty$. Then, for $s > 0$,*

$$\frac{c - \Lambda \mathbb{E}\tau \tilde{F}_1(s)}{c - \Lambda \mathbb{E} D \tilde{D}_1(s)} = 1 - \Lambda \int_0^\infty \mathbb{E}\left\{e^{-s(\tau - ct)} 1(\tau \geq t)\right\} e^{-f(s,t)} dt, \tag{6.7}$$

where

$$f(s,t) = \Lambda\{t(1 - \tilde{F}(s)) + \mathbb{E}[\tau \exp(-s\tau)] - \mathbb{E}[(\tau - t)\exp(-s\tau) 1(\tau \geq t)]\},$$

$\tilde{F}(s) = \mathbb{E}e^{-s\tau}$, *and* $\tilde{F}_1(s) = (1-\mathbb{E}e^{-s\tau})/(s\mathbb{E}\tau)$, $\tilde{D}_1(s) = (1-\mathbb{E}e^{-sD})/(s\mathbb{E}D)$, *are the LS transforms of the integrated tail distributions of* F *and* D, *respectively;* $\mathbb{E}D = (\mathbb{E}\tau - c/\Lambda)e^{\Lambda\mathbb{E}\tau} + c/\Lambda$.

Proof of Theorem 6: Let $r = 1$. In order to simplify the usage of Theorem 16, without loss of generality we assume that $1-F(x) \sim (-1)^n x^{-\alpha} l(x)/\Gamma(1-\alpha)$, $\alpha = n+\beta, n \in \mathbb{N}, 0 < \beta < 1$. Let us denote the integral in (6.7) as

$$\phi(s) = \int_0^\infty g(s,t)e^{-f(s,t)}dt,$$

where $g(s,t) \stackrel{\text{def}}{=} \mathbb{E}\left\{e^{-s(\tau-ct)}1(\tau \geq t)\right\}$. Next, we want to find the nth derivative of $\phi(s)$ and compare it with the nth derivative of the left-hand side of (6.7) as $s \downarrow 0$. The *main technical difficulty* in doing this is to prove the following two lemmas.

Lemma 8 *For any* $\rho > \epsilon > 0$, *there exist* $s_0 > 0, t_0 > 0$, *such that for all* $0 < s < s_0$, $t > t_0$, $0 \leq k \leq n$,

$$(1-\epsilon)\rho^k t^k e^{-s(\rho+\epsilon)t-\rho} \leq (-1)^k \frac{\partial^k}{\partial s^k} e^{-f(s,t)} \leq (1+\epsilon)\rho^k t^k e^{-s(\rho-\epsilon)t-\rho}. \qquad (6.8)$$

Proof: Given at the end of this section. ◇

Lemma 8 is crucial in proving the following result which directly leads to the proof of Theorem 6. Unfortunately, the proof of this lemma is very technical and, due to the space limitations, we do not present it here.

Lemma 9 *As* $s \downarrow 0$

$$\phi^{(n)}(s) \sim \frac{d^n}{ds^n}\int_0^\infty \mathbb{E}\left\{e^{-s(\tau-ct)}1(\tau \geq t)\right\} e^{-\rho st-\rho}dt, \qquad (6.9)$$

where $\rho = \Lambda\mathbb{E}\tau$.

Proof: Given at the end of this section. ◇

At this point we are ready to finish the proof of Theorem 6. By using Lemma 9 we derive

$$\begin{aligned}
\lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1}\phi^{(n)}(s) &= \lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1}e^{-\rho}\frac{d^n}{ds^n}\mathbb{E}e^{-s\tau}\int_0^\tau e^{-(\rho-c)st}dt \\
&= \lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1}e^{-\rho}\frac{d^n}{ds^n}\mathbb{E}e^{-s\tau}\frac{e^{-(\rho-c)s\tau}-1}{-s(\rho-c)} \\
&= \lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1}e^{-\rho}\frac{d^n}{ds^n}\mathbb{E}\frac{e^{-s\tau}-e^{-(\rho+1-c)s\tau}}{s(\rho-c)} \\
&= \frac{e^{-\rho}\mathbb{E}\tau}{\rho-c}\lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1}\frac{d^n}{ds^n}\left[\mathbb{E}\frac{1-e^{-(\rho+1-c)s\tau}}{s\mathbb{E}\tau} - \mathbb{E}\frac{1-e^{-s\tau}}{s\mathbb{E}\tau}\right] \\
&= \frac{e^{-\rho}\mathbb{E}\tau}{\rho-c}\lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1} \\
&\qquad \times\frac{d^n}{ds^n}\left[(\rho+1-c)\tilde{F}_1(s(\rho+1-c)) - \tilde{F}_1(s)\right]; \quad (6.10)
\end{aligned}$$

recall that $\tilde{F}_1(s)$ is the LS transform of the integrated tail distribution of τ. From the assumption on F it follows that

$$1-F_1(x) \sim (-1)^n x^{-(\alpha-1)} l(x)/(\Gamma(1-\alpha)(\alpha-1)\mathbb{E}\tau) = (-1)^{n-1} x^{-(\alpha-1)} l(x)/(\Gamma(2-\alpha)\mathbb{E}\tau) \quad x \to \infty.$$

Consequently, by Theorem 16, equation (6.5),

$$\tilde{F}_1^{(n)}(s) \sim (-1)^n \frac{\Gamma(\alpha)}{\Gamma(\beta)\mathbb{E}\tau} s^{\beta-1} l(1/s), \quad \text{as} \quad s \downarrow \infty. \tag{6.11}$$

The latter result, when replaced in (6.10), produces

$$-\Lambda \lim_{s \downarrow 0} s^{1-\beta}(l(1/s))^{-1}\phi^{(n)}(s) = \frac{\Lambda e^{-\rho}}{\rho - c} \times \frac{(-1)^n \Gamma(\alpha)}{\Gamma(\beta)} \left[1 - (\rho + 1 - c)^\alpha\right], \tag{6.12}$$

which represents the result of taking the operator $\lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1} d^n/ds^n$ on the right hand side of (6.7). To finish the proof we have to compute the result of applying the same operator on the left-hand side of (6.7). Let us start with the derivative

$$\frac{d^n}{ds^n}\frac{c - \Lambda\mathbb{E}\tau\tilde{F}_1(s)}{c - \Lambda\mathbb{E}D\tilde{D}_1(s)} = \frac{-\Lambda\mathbb{E}\tau\tilde{F}_1^{(n)}(s)}{c - \Lambda\mathbb{E}D\tilde{D}_1(s)} + \frac{(c - \Lambda\mathbb{E}\tau\tilde{F}_1(s))\Lambda\mathbb{E}D\tilde{D}_1^{(n)}(s)}{(c - \Lambda\mathbb{E}D\tilde{D}_1(s))^2} + R_{n-1}(s), \tag{6.13}$$

where $R_{n-1}(s)$ is a rational function that contains only the first $n-1$ derivatives of $\tilde{D}_1(s)$, and $\tilde{F}_1(s)$. Note that, $|\tilde{F}_1^{(k)}(s)| \leq |\tilde{F}_1^{(k)}(0)| < \infty$, for $0 \leq k \leq n-1$, $s \geq 0$. Furthermore, by induction and by taking successive derivatives in (6.7), it is easy to show that $|\tilde{D}_1^{(k)}(s)| \leq |\tilde{D}_1^{(k)}(0)| < \infty$, for $0 \leq k \leq n-1$, $s \geq 0$. Therefore,

$$\lim_{s\downarrow 0} |R_{n-1}(s)| < \infty. \tag{6.14}$$

Combining (6.11) and $\mathbb{E}D = (\mathbb{E}\tau - c/\Lambda)e^\rho + c/\Lambda$ we obtain

$$\frac{-\Lambda\mathbb{E}\tau\tilde{F}_1^{(n)}(s)}{c - \Lambda\mathbb{E}D\tilde{D}_1(s)} \sim \frac{\Lambda e^{-\rho}}{\rho - c} \times \frac{(-1)^n\Gamma(\alpha)}{\Gamma(\beta)} s^{\beta-1} l(1/s) \quad s \to 0. \tag{6.15}$$

Finally, $\mathbb{E}D = (\mathbb{E}\tau - c/\Lambda)e^\rho + c/\Lambda$, (6.12), (6.13), (6.14), and (6.15), yield

$$\begin{aligned}\lim_{s\downarrow 0} s^{1-\beta} l(1/s)\tilde{D}_1^{(n)}(s) &= \frac{(c - \Lambda\mathbb{E}D)^2}{(c-\rho)\Lambda\mathbb{E}D} \times \frac{-\Lambda e^{-\rho}}{\rho - c} \times \frac{(-1)^n\Gamma(\alpha)}{\Gamma(\beta)}(\rho + 1 - c)^\alpha \\ &= e^\rho(\rho + 1 - c)^\alpha \frac{(-1)^n\Gamma(\alpha)}{\Gamma(\beta)\mathbb{E}D},\end{aligned} \tag{6.16}$$

which is, by Theorem 16, equivalent to

$$\int_x^\infty \mathbb{P}[D > u]du \sim e^\rho(\rho + 1 - c)^\alpha (-1)^{n-1} x^{-(\alpha-1)} l(x)/\Gamma(2-\alpha) \quad \text{as} \quad x \to \infty;$$

finally, by Monotone Density Theorem ([BGT87]), pp. 39)

$$\mathbb{P}[D>x] \sim e^{\rho}(\rho+1-c)^{\alpha}(-1)^{n}x^{-\alpha}l(x)/\Gamma(1-\alpha) \sim e^{\rho}(\rho+1-c)^{\alpha}\mathbb{P}[\tau>x] \sim e^{\rho}\mathbb{P}[\tau(\rho+1-c)>x]$$

as $x\to\infty$; this finishes the proof of Theorem 6 for the case $r=1$. For $r\neq 1$, $D_{c,r}=rD_{c/r,1}$; thus

$$\mathbb{P}[D_{c,r}>x]=\mathbb{P}[rD_{c/r,1}>x]\sim e^{\rho}(r+\rho-c)^{\alpha}\mathbb{P}[\tau>x] \quad \text{as} \quad x\to\infty,$$

where $\rho=r\Lambda\mathbb{E}\tau$. This completes the proof of Theorem 6. $\diamond$

Proof of Lemma 8 : Let us start with the case $k=0$. Since $\mathbb{E}[(\tau-t)\exp(-s\tau)1(\tau\geq t)]$ $\leq \mathbb{E}[(\tau-t)1(\tau\geq t)]\to 0$, as $t\to\infty$, and $(1-\tilde{F}(s))/s\to\mathbb{E}\tau$, $\mathbb{E}[\tau\exp(-s\tau)]\to\mathbb{E}\tau$ as $s\to 0$, it follows that for any $\epsilon>0$, $\epsilon<\rho$, there exist $s_0>0, t_0>0$, such that for all $0<s<s_0$, $t>t_0$

$$(1-\epsilon)e^{-s(\rho+\epsilon)t-\rho}\leq e^{-f(s,t)}\leq(1+\epsilon)e^{-s(\rho-\epsilon)t-\rho}. \tag{6.17}$$

For $1\leq k\leq n$ let us first show that for any $\rho>\epsilon>0$,

$$(1-\epsilon)\rho^{k}t^{k}\leq(-1)^{k}e^{f(s,t)}\frac{\partial^{k}}{\partial s^{k}}e^{-f(s,t)}\leq(1+\epsilon)\rho^{k}t^{k}, \tag{6.18}$$

for all sufficiently small s and all sufficiently large t. To show this, notice that for all $1\leq k\leq(n-1)$, and all $s>0$,

$$\left|\frac{\partial^{k}}{\partial s^{k}}f(s,t)+\Lambda t\tilde{F}^{(k)}(s)\right|\leq\Lambda\mathbb{E}[\tau^{k+1}]+\Lambda t\mathbb{E}[\tau^{k}1(\tau\geq t)]. \tag{6.19}$$

For $k=n$

$$\left|\frac{\partial^{n}}{\partial s^{n}}f(s,t)+\Lambda t\tilde{F}^{(n)}(s)\right|\leq\Lambda\mathbb{E}[\tau^{n+1}1(\tau<t)]+\Lambda t\mathbb{E}[\tau^{n}1(\tau\geq t)]=o(t), \tag{6.20}$$

as $t\to\infty$. Hence, since $|\tilde{F}^{(k)}(s)|\leq\mathbb{E}\tau^{k}<\infty, s\geq 0, 0\leq k\leq n$, (6.19) and (6.20) imply that

$$\left|\frac{\partial^{k}}{\partial s^{k}}f(s,t)\right|=O(t), \tag{6.21}$$

uniformly for all $1\leq k\leq n$, and all $s>0$. Furthermore, for $k=1$, for any $\epsilon>0, \epsilon<\rho$,

$$(1-\epsilon)\rho t\leq\frac{\partial}{\partial s}f(s,t)\leq(1+\epsilon)\rho t, \tag{6.22}$$

for all sufficiently small s and all sufficiently large t. Using (6.21), after some straightforward algebra, which we skip here, we arrive at

$$\left|(-1)^{k}e^{f(s,t)}\frac{\partial^{k}}{\partial s^{k}}e^{-f(s,t)}-\left(\frac{\partial}{\partial s}f(s,t)\right)^{k}\right|=O(t^{k-1}), \tag{6.23}$$

uniformly in $s > 0$. Subsequently, combination of (6.22) and (6.23) yields (6.18) (with possibly two different ϵ in (6.22) and (6.18)). Finally, (6.17) and (6.18), give (6.8), for $1 \leq k \leq n$. This finishes the proof of Lemma 8. $\diamond$

Proof of Lemma 9: Now, the nth derivative of $\phi(s)$ is given by

$$\begin{aligned}
\frac{d^n}{ds^n}\phi(s) &= \sum_{k=0}^{n} \binom{n}{k} \int_0^\infty \frac{\partial^{n-k}}{\partial s^{n-k}} g(s,t) \frac{\partial^k}{\partial s^k} e^{-f(s,t)} dt && (6.24)\\
&\stackrel{\text{def}}{=} \sum_{k=0}^{n} \binom{n}{k} (-1)^n I_k(s); && (6.25)
\end{aligned}$$

since the interchange of the differentiation and integration in (6.24) is granted by dominated convergence.

Hence, the asymptotic behavior of $\phi^{(n)}(s)$ at the origin will follow if we derive the asymptotic behavior of $I_k(s), 0 \leq k \leq n$, at the origin. Note that $I_k(s)$ can be written as

$$\begin{aligned}
I_k(s) &= \int_0^\infty \mathbb{E}\left\{(\tau - ct)^{n-k} e^{-s(\tau - ct)} 1(\tau \geq t)\right\} (-1)^k \frac{\partial^k}{\partial s^k} e^{-f(s,t)} dt && (6.26)\\
&= \int_0^{t_0} + \int_{t_0}^\infty,
\end{aligned}$$

for some $t_0 > 0$. Now, for any $\epsilon > 0$, we can choose t_0, s_0, according to Lemma 8, such that for all $0 < s < s_0$, and a sufficiently large constant $C_{t_0} < \infty$,

$$\begin{aligned}
I_k(s) &\leq C_{t_0} + (1+\epsilon)\rho^k \int_0^\infty \mathbb{E}\left\{(\tau - ct)^{n-k} e^{-s(\tau - ct)} 1(\tau \geq t)\right\} t^k e^{-s(\rho-\epsilon)t - \rho} dt \\
&= C_{t_0} + (1+\epsilon)\rho^k e^{-\rho} \sum_{j=0}^{n-k} \binom{n-k}{j} (-c)^j \mathbb{E} \int_0^\tau \tau^{n-k-j} t^{k+j} e^{-s(\tau + (\rho - \epsilon - c)t)} dt \\
&\stackrel{\text{def}}{=} C_{t_0} + (1+\epsilon)\rho^k e^{-\rho} \sum_{j=0}^{n-k} \binom{n-k}{j} (-c)^j I_{kj}^{\epsilon}(s). && (6.27)
\end{aligned}$$

Here, let $\delta_\epsilon = \rho - \epsilon - c$ and $p = k + j$, then

$$\begin{aligned}
I^\epsilon_{kj}(s) &= \mathbb{E}\tau^{n-p}e^{-s\tau}\int_0^\tau t^p e^{-s\delta_\epsilon t}dt \\
&= \left(\frac{-1}{\delta_\epsilon}\right)^p \mathbb{E}\tau^{n-p}e^{-s\tau}\frac{\partial^p}{\partial s^p}\int_0^\tau e^{-s\delta_\epsilon t}dt \\
&= \left(\frac{-1}{\delta_\epsilon}\right)^p \mathbb{E}\tau^{n-p}e^{-s\tau}\frac{\partial^p}{\partial s^p}\frac{e^{-s\delta_\epsilon\tau}-1}{-s\delta_\epsilon} \\
&= \frac{-1}{\delta_\epsilon^{p+1}}\mathbb{E}\left\{\tau^{n-p}e^{-s\tau}p!s^{-p-1}\left[\sum_{q=0}^p\frac{(\tau\delta_\epsilon s)^q}{q!}e^{-\tau\delta_\epsilon s}-1\right]\right\} \\
&= \frac{-1}{\delta_\epsilon^{p+1}}p!s^{-p-1}\left\{\sum_{q=0}^p\frac{(\delta_\epsilon s)^q}{q!}\mathbb{E}\tau^{n-p+q}e^{-\tau(1+\delta_\epsilon)s} - \mathbb{E}\tau^{n-p}e^{-\tau s}\right\} \\
&\stackrel{\text{def}}{=} \frac{-1}{\delta_\epsilon^{p+1}}p!s^{-p-1}h(s).
\end{aligned} \tag{6.28}$$

Careful examination of h shows that

$$\lim_{s\downarrow 0} h^{(k)}(s) = 0, \quad \text{for} \quad 0 \le k \le p. \tag{6.29}$$

Also, it is easy to see that the $(p+1)$st derivative of h has the following form

$$h^{(p+1)}(s) = \sum_{q=0}^p c_q(\epsilon)s^q\mathbb{E}\tau^{n+1+q}e^{-\tau(1+\delta_\epsilon)s} - (-1)^{p+1}\mathbb{E}\tau^{n+1}e^{-s\tau}, \tag{6.30}$$

for some constants $c_q(\epsilon), 0 \le q \le p$. Note that, by Theorem 16, all $s^q\mathbb{E}\tau^{n+1+q}e^{-\tau(1+\delta_\epsilon)s}, 0 \le q \le p$, and $\mathbb{E}\tau^{n+1}e^{-s\tau}$, are asymptotically proportional to $l(1/s)s^{\beta-1}$ as $s \downarrow 0$. Therefore, in conjuction with (6.29), we obtain that $h(s) \sim C_h(\epsilon)s^{p+\beta}l(1/s)$, as $s \downarrow 0$, for some constant $C_h(\epsilon)$ continuous in ϵ. Hence,

$$\lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1}I^\epsilon_{kj}(s) = C_{kj}(\delta_\epsilon), \tag{6.31}$$

for some constant $C_{kj}(\delta_\epsilon)$. By observing that $C_{kj}(\delta_\epsilon)$ is continuous in δ_ϵ, we obtain

$$\begin{aligned}
\lim_{\epsilon\to 0}\lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1}I^\epsilon_{kj}(s) &= \lim_{\epsilon\to 0} C_{kj}(\delta_\epsilon) \\
&= \lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1}I^0_{kj}(s).
\end{aligned} \tag{6.32}$$

Now, by taking the limits with respect to ϵ, and s, in (6.27), and utilizing (6.32) to interchange the limits, we arrive at

$$\limsup_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1} I_k(s) \;\le\; \rho^k e^{-\rho} \sum_{j=0}^{n-k} \binom{n-k}{j} (-c)^j \lim_{\epsilon\to 0}\lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1} I^{\epsilon}_{kj}(s)$$

$$= \lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1} \int_0^\infty \mathbb{E}\left\{(\tau - ct)^{n-k} e^{-s(\tau-ct)} 1(\tau \ge t)\right\} t^k \rho^k e^{-s\rho t-\rho} dt.$$

By repeating the same arguments one can prove the lower bound, which results in

$$\lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1} I_k(s) \tag{6.33}$$

$$= \lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1} \int_0^\infty \mathbb{E}\left\{(\tau - ct)^{n-k} e^{-s(\tau-ct)} 1(\tau \ge t)\right\} t^k \rho^k e^{-s\rho t-\rho} dt.$$

Finally, by applying (6.33) in (6.24) we obtain

$$\lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1} \phi^{(n)}(s) \;=\; \sum_{k=0}^{n} \binom{n}{k} (-1)^n \lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1} I_k(s)$$

$$= \lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1} \sum_{k=0}^{n} \binom{n}{k} (-1)^n$$

$$\times \int_0^\infty \mathbb{E}\left\{(\tau - ct)^{n-k} e^{-s(\tau-ct)} 1(\tau \ge t)\right\} t^k \rho^k e^{-s\rho t-\rho} dt$$

$$= \lim_{s\downarrow 0} s^{1-\beta}(l(1/s))^{-1} e^{-\rho} \frac{d^n}{ds^n} \int_0^\infty \mathbb{E}\left\{e^{-s(\tau-ct)} 1(\tau \ge t)\right\} e^{-\rho s t} dt$$

which finishes the proof of Lemma 9. $\diamond$

C: Finite Number of Subexponential On-Off Processes: Upper Bound

For a finite number of long-tailed On-Off processes we can obtain the following general upper bound.

Theorem 18 *Let $F \in \mathcal{L}$, and $F_1 \in \mathcal{S}$. If $r > c/N$, then*

$$\lim_{x\to\infty} \frac{\mathbb{P}[Q^N > x]}{\int_{x/(r-c_N)}^\infty \mathbb{P}[\tau^{on} > u]du} \le N K_N, \tag{6.34}$$

where $c_N = c/N$, and K_N is given by Theorem 9, equation (4.10), with c_N in place of c.

Proof: Similarly, as in the proof of Theorem 10, we have the decomposition

$$\begin{aligned} W_t = \int_{-t}^{0} (A_u^N - c)du \;&=\; \sum_{i=1}^{N} \int_{-t}^{0} (a_u^i - c/N)du \\ &\stackrel{\text{def}}{=}\; \sum_{i=1}^{N} w_t^i. \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{P}[Q_t^N > x] \;&=\; \mathbb{P}[\sup_{t\geq 0} \sum_{i=1}^{N} w_t^i > x] \\ &\leq\; \mathbb{P}[\sum_{i=1}^{N} \sup_{t\geq 0} w_t^i > x] \\ &\sim\; N\mathbb{P}[\sup_{t\geq 0} w_t^1 > x] \;\text{ as }\; x \to \infty \qquad (6.35) \\ &=\; NK_N, \qquad (6.36) \end{aligned}$$

where (6.35) follows from Lemma 6 (i), since $\sup_{t\geq 0} w_t^i, 1 \leq i \leq N$, are i.i.d. subexponential (by Theorem 9) random variables; (6.36) follows also from Theorem 9. This proves (6.34). $\diamond$

References

[ACW94] J. Abate, G.L. Choudhury, and W. Whitt, "Waiting-time tail probabilities in queues with long-tail service-time distributions.," *Queueing systems.*, vol. 16, no. 3/4, pp. 311–338, 1994.

[AHK94] S. Asmussen, L. F. Henriksen, and C. Klüppelberg, "Large claims approximations for risk processes in a Markovian environment," *Stochastic Processes and their Applications*, vol. 54, pp. 29–43, 1994.

[AMS82] D. Anick, D. Mitra, and M. M. Sondhi, "Stochastic theory of a data handling system with multiple sources," *Bell Syst. Techn. J.*, vol. 61, pp. 1871–1894, 1982.

[ANA95] V. Anantharam, "On the sojourn time of sessions at an ATM buffer with long-range dependent input traffic," in *Proceedings of the 34th IEEE Conference on Decision and Control*, December 1995.

[ASC97] S. Asmussen, H. Schmidli, and V. Schmidt, "Tail probabilities for non-standard risk and queueing processes with subexponential jumps." preprint, 1997.

[ASM87] S. Asmussen, *Applied Probability and Queues.* John Wiley and Sons, 1987.

[ASM96] S. Asmussen, "Rare events in the presence of heavy tails," in *Stochastic Networks: Stability and Rare Events* (P. Glasserman, K. Sigman, and D. D. Yao, editors), Lecture Notes, Springer-Verlag, 1996.

[ATN72] K. B. Athreya and P. E. Ney, *Branching Processes.* Springer-Verlag, 1972.

[BAB94] F. Baccelli and P. Bremaud, *Elements of Queueing Theory: Palm-Martingale Calculus and Stochastic Recurence.* Springer Verlag, 1994.

[BGT87] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular Variation.* Cambridge University Press, Cambridge, 1987.

[BID74] N. H. Bingham and R. A. Doney, "Asymptotic properties of super-critical branching processes I: the Galton-Watson process," *Advances in Applied Probability*, vol. 6, pp. 711–731, 1974.

[BOX96] O. J. Boxma, "Fluid queues and regular variation," *Performance Evaluation*, vol. 27&28, pp. 699–712, 1996.

[BOX97] O. J. Boxma, "Regular variation in a multi-source fluid queue," in *ITC 15*, pp. 391–402, Washington, D.C., USA, June 1997.

[CHI64] V. P. Chistyakov, "A theorem on sums of independent positive random variables and its application to branching random processes," *Theor. Probab. Appl.*, vol. 9, pp. 640–648, 1964.

[CHW95] G. L. Choudhury and W. Whitt, "Long-tail buffer-content distributions in broadband networks," *Performance Evaluation*, vol. 30, pp. 177–190, 1997.

[CIN72] E. Cinlar, "Superposition of point processes," in *Stochastic Point Processes: Statistical Analysis, Theory and Application* (P. A. W. Lewis, editor), pp. 594–606, New York: Wiley, 1972.

[CIN75] E. Cinlar, *Introduction to Stochastic Processes.* Prentice-Hall, 1975.

[CLI86] D. B. H. Cline, "Convolution tails, product tails and domains of attraction," *Probab. Th. Rel. Fields*, vol. 72, no. 1, pp. 529–557, 1986.

[CLI94] Daren B. H. Cline, "Intermediate regular and π variation," *Proc. London Math. Soc.*, 1994.

[COH73] J. W. Cohen, "Some results on regular variation for distributions in queueing and fluctuation theory," *J. Appl. Probability*, vol. 10, pp. 343–353, 1973.

[COH74] J. W. Cohen, "Superimposed renewal processes and storage with gradual input," *Stochastic Processes and their Applications*, vol. 2, pp. 31–58, 1974.

[COH94] J. W. Cohen, "On the effective bandwidth in buffer design for the multi-server channels," Technical report, CWI Report BS-R9406, 1994.

[COS54] D. R. Cox and W. L. Smith, "On the superposition of renewal processes," *Biometrika*, vol. 41, pp. 91–99, 1954.

[DUC95] N. G. Duffield and N. O'Connell, "Large deviations and overflow probabilities for the general single-server queue with applications," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 118, pp. 363–374, 1995.

[EGV79] P. Embrechts, C. M. Goldie, and N. Veraverbeke, "Subexponentiality and infinite divisibility," *Z. Wahr. verw. Gebiete*, vol. 49, pp. 335–347, 1979.

[EHL95] A. Elwalid, D. Heyman, T. V. Lakshman, D. Mitra, and A. Weiss, "Fundamental bounds and approximations for ATM multiplexers with applications to video teleconferencing," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 1004–1016, August 1995.

[GLW94] P. V. Glynn and W. Whitt, "Logarithmic asymptotics for steady-state tail probabilities in a single-server queue," in *Studies in Applied Probability* (J. Galambos and J. Gani, editors), vol. 31A (special issue of *J. of Appl. Prob.*), pp. 131–156, Applied Probability Trust, Sheffield, England, 1994.

[HEL96] D. P. Heyman and T. V. Lakshman, "Source models for VBR broadcast-video traffic," *IEEE J. Select. Areas in Communications*, vol. 4, pp. 40–48, February 1996.

[HRS96] D. Heath, S. Resnick, and G. Samorodnitsky, "Heavy tails and long range dependence in on/off processes and associated fluid models," preprint.

[JLA95g] P. R. Jelenković and A. A. Lazar, "Subexponential asymptotics of a Markov-modulated random walk with queueing applications," *Journal of Applied Probability*, vol. 35, June 1998.

[JLA96b] P. R. Jelenković and A. A. Lazar, "Multiplexing on-off sources with subexponential on periods: Part I," in *Proceedings of INFOCOM'97*, Kobe, Japan, April 1997.

[JLA96b2] P. R. Jelenković and A. A. Lazar, "Multiplexing on-off sources with subexponential on periods: Part II," in *ITC 15*, pp. 965–974, Washington, D.C., USA, June 1997.

[JLA96h] P. R. Jelenković and A. A. Lazar, "Multiple time scale and subexponential asymptotic behavior of a network multiplexer," in *Stochastic Networks: Stability and Rare Events* (P. Glasserman, K. Sigman, and D. D. Yao, editors), Lecture Notes, pp. 215–235, Springer-Verlag, 1996.

[JLS95b] P. R. Jelenković, A. A. Lazar, and N. Semret, "Multiple time scales and subexponentiality in MPEG video streams," in *International IFIP-IEEE Conference on Broadband Communications*, April 1996.

[JLS96] P. R. Jelenković, A. A. Lazar, and N. Semret, "The effect of multiple time scales and subexponentiality of MPEG video streams on queueing behavior," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 1052–1071, August 1997.

[KAR30] J. Karamata, "Sur un mode de croissance régulière des fonctions," *Mathematica (Cluj)*, vol. 4, pp. 38–53, 1930.

[KEW92] O. Kella and W. Whitt, "A storage model with a two-state random environment," *Oper. Res.*, vol. 40, pp. 257–262, 1992.

[KLU88] C. Klüppelberg, "Subexponential distributions and integrated tails," *J. Appl. Prob.*, vol. 25, pp. 132–141, 1988.

[KLU89] C. Klüppelberg, "Subexponential distributions and characterizations of related classes," *Probability Theory and Related Fields*, vol. 82, pp. 259–269, 1989.

[LOY62] R. M. Loynes, "The stability of a queue with non-independent inter-arrival and service times," *Proc. Cambridge Philos Soc.*, vol. 58, pp. 497–520, 1968.

[LTG95] N. Likhanov, B. Tsybakov, and N. D. Georganas, "Analysis of an ATM buffer with self-similar ("fractal") input traffic," in *INFOCOM'95*, pp. 985–991, Boston, Masssachusetts, April 1995.

[LTW93] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic," in *SIGCOMM'93*, pp. 183–193, 1993.

[NOR94] I. Norros, "A storage model with self-similar input," *Queueing Systems*, vol. 16, pp. 387–396, 1994.

[PAK75] A.G. Pakes, "On the tails of waiting-time distribution," *J. Appl. Probab.*, vol. 12, pp. 555–564, 1975.

[PAM96] M. Parulekar and A. M. Makowski, "Tail probabilitites for a multiplexer with self-similar traffic," in *INFOCOM'96*, San Francisco, California, March 1996.

[PAM97] M. Parulekar and A. M. Makowski, "Tail probabilitites for M/G/∞ input processes (I): preliminary asymptotices," *Queueing Systems*, vol. 27, pp. 271–296, 1997.

[RES96] S. Resnick and G. Samorodnitsky, "Performance decay in a single server queueing model with long range dependence," preprint, 1996.

[RSS97] T. Rolski, S. Schlegel, and V. Schmidt, "Asymptotics of palm-stationary buffer content distribution in fluid flow queues," in *9th INFORMS Applied Probability Conference*, Cambridge, Massachusetts, June 1997.

[RUB73] M. Rubinovitch, "The output of a buffered data communication system," *Stochastic Processes and their Applications*, vol. 1, pp. 375–380, 1973.

[RYL96] B. K. Ryu and S. B. Lowen, "Point process approaches to the modeling and analysis of self-similar traffic - part I: Model construction," in *INFOCOM'96*, San Francisco, California, March 1996.

[WES95] A. Weiss and A. Shwartz, *Large Deviations for Performance Analysis: Queues, Communications, and Computing.* New York: Chapman & Hall, 1995.

[WIT92] 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*, vol. 10, pp. 295–312, 1992.