



Centrum voor Wiskunde en Informatica

REPORTRAPPORT

Regular variation in a multi-source fluid queue

O.J. Boxma

Department of Operations Research, Statistics, and System Theory

BS-R9614 1996

Report BS-R9614
ISSN 0924-0659

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

CWI is the National Research Institute for Mathematics and Computer Science. CWI is part of the Stichting Mathematisch Centrum (SMC), the Dutch foundation for promotion of mathematics and computer science and their applications.

SMC is sponsored by the Netherlands Organization for Scientific Research (NWO). CWI is a member of ERCIM, the European Research Consortium for Informatics and Mathematics.

Copyright © Stichting Mathematisch Centrum
P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

Regular Variation in a Multi-Source Fluid Queue

O.J. Boxma

CWI

P.O. Box 94079, 1090 GB Amsterdam, The Netherlands;

Tilburg University, Faculty of Economics

P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Abstract

This paper considers a fluid queueing system, fed by N independent sources that alternate between silence and activity periods. We assume that the distribution of the activity periods of source 1 is a regularly varying function of index ζ , whereas all other sources have activity period distributions with an exponential tail. In addition, we assume that the inflow rate of each of the sources, when active, exceeds the outflow rate of the buffer. Under these assumptions, we show that the tail of the buffer content distribution is regularly varying of index $\zeta + 1$. In the special case that $\zeta \in (-2, -1)$, which implies long-range dependence of the input process, the buffer content does not even have a finite first moment.

Based on the obtained results and on a conjecture for the case that the outflow rate of the buffer is not necessarily exceeded by the inflow rates of the sources, we suggest a simple, effective-bandwidth-like, connection admission rule.

AMS Subject Classification (1991): 60K25, 68M20.

Keywords & Phrases: Fluid queue, regular variation, long-range dependence, buffer content.

1. INTRODUCTION

Recent measurements (see e.g. [15, 16]) have revealed that in high-speed telecommunication networks, like the ATM-based Broadband ISDN, traffic conditions can occur that exhibit long-range dependence and burstiness over an extremely wide range of time scales. The modeling and analysis of these phenomena is a difficult and important enterprise. Several interesting models have recently been proposed. Norros [14] considers a fluid queue with as input process Fractional Brownian Motion. This self-similar input process exhibits long-range dependence. Norros analyzes the buffer content process of this fluid queue. Erramilli et al. [10] try to capture long-range dependence and burstiness of traffic via chaos theory; a non-linear deterministic chaotic map

prescribes the successive states of an on/off source. The performance analysis of the resulting non-traditional queueing model still is a very challenging task. In the present paper, following [3, 4, 5, 13], we consider a queueing model that is more traditional and that - as will be demonstrated in the sequel - is accessible to a rather detailed qualitative and quantitative analysis.

Our model is a fluid queueing system, that is fed by N independent on/off sources. Each source alternates between activity periods and silence periods; in the former periods a source feeds work into the buffer at constant rate. This fluid queueing system has been found to be particularly well matched to the ATM environment. Since the fundamental publications of Anick et al. [1], Cohen [7] and Kosten [12], it has become a key model for the performance analysis of high-speed telecommunication networks. However, in [1, 12] and subsequent publications the activity and silence periods are exponentially distributed or determined by some Markov process, leading to exponential behaviour of the tail of the buffer content distribution. We are interested in the case that the activity period distribution of one of the sources has a non-exponential, fat, tail (this may, for example, arise in file transfers). More precisely, we consider the case that the tail of the activity period distribution of the first source is *regularly varying*. A measurable positive function f is called regularly varying of index ζ if, for all $x > 0$,

$$f(xt)/f(t) \rightarrow x^\zeta, \quad t \rightarrow \infty, \quad (1.1)$$

(cf. [2], p. 18). One writes $f \in R_\zeta$. When $\zeta = 0$, one speaks of a *slowly varying function* (to be denoted by $l(\cdot)$ in this paper); this could for instance be a constant, or a logarithmic function. We shall say that a distribution $F(\cdot)$ is regularly varying (or “has a regularly varying tail”) when $F^c(t) := 1 - F(t)$ is a regularly varying function. The Pareto distribution belongs to this class. Of particular interest to us is the case that an activity period distribution has a regularly varying tail of index $\zeta \in (-2, -1)$. In that case the first moment of the distribution exists, but the variance is infinite. This case is known to give rise to long-range dependence (see e.g. [3]).

For the traditional GI/G/1 queue, Cohen [6] has studied the effect of regularly varying interarrival or service time distributions on waiting-time and workload distributions. His main result states: the waiting-time distribution $W_{GI/G/1}(t)$ in the GI/G/1 queue is regularly varying of index $1 - \nu$ (with $\nu > 1$) iff the service time distribution $B_{GI/G/1}(t)$ is regularly varying of index $-\nu$. More explicitly, for $t \rightarrow \infty$ and $\nu > 1$, with β denoting mean service time, and ρ traffic load, and with $l(\cdot)$ a slowly varying function [6]:

$$B_{GI/G/1}^c(t) = (\nu - 1)(\beta/t)^\nu l(t) \Leftrightarrow W_{GI/G/1}^c(t) = \frac{\rho}{1 - \rho} (\beta/t)^{\nu-1} l(t). \quad (1.2)$$

Note Actually in the “if” (“only if”) part of (1.2) the second (first) equality sign should be replaced by “ \sim ” (or one should write $l_1(t)$ and $l_2(t)$, respectively); for briefness sake we use the above notation in (1.2), and in subsequent relations of the same type.

In [3] we have exploited this result in the following way. We have shown that the buffer content, at the end of periods in which all sources are silent, has the same distribution as the waiting time in an M/G/1 queue with mean interarrival time the mean time that all sources are silent, and with service time distribution the distribution of the net increment of the buffer content during a period in which at least one source is active. Using (1.2) it now suffices to show that the latter increment is regularly varying, in order to conclude that the buffer content distribution is regularly varying. This has been done in [3] for three cases in which the tail of the activity period distribution of source 1 is regularly varying: (i) $N = 1$; (ii) $N = 2$, source 2 having an exponential activity period distribution; (iii) $N = \infty$, all sources being identical.

In the present paper we shall considerably extend the second result. We consider $N \geq 2$ sources, and we assume that the activity period distributions of sources $2, \dots, N$ have an exponential tail (e.g., the activity period distributions are Erlangian or hyperexponentially distributed). We also assume, as in [3], that the inflow rate of each of the sources, when active, exceeds the outflow rate of the buffer. We show that the buffer content distribution is regularly varying iff the activity period distribution of source 1 is regularly varying. Mathematically the case of *at least two* regularly varying activity period distributions is more difficult (we'll discuss this in a sequel to the present paper), but it is intuitively clear that a second regularly varying source can only make things "worse"; so for obtaining qualitative insight into the effect of a heavy-tailed activity period, and the additional effect of other more smoothly operating sources, it is already most interesting to study the case of one such badly behaved source.

The paper is organized as follows. The fluid queueing system under consideration is described in Section 2. Section 3 displays for this model some key results of Cohen [9] that form the starting-point of our approach. Section 4 considers the case that sources $2, \dots, N$ all have a negative exponential activity period distribution, while Section 5 considers the more general case that sources $2, \dots, N$ have activity period distributions with an exponential tail behaviour. In both cases we show that the buffer content distribution is regularly varying with parameter $1 - \nu$ iff the activity period distribution of source 1 is regularly varying with parameter $-\nu$; the smoothly behaving sources $2, \dots, N$ do not change the index of regular variation of the buffer content distribution. In Section 6 we propose a simple rule for deciding whether an additional source may be admitted to the system. This rule is based upon the results of Sections 4 and 5, and on a conjecture for the case that the outflow rate of the buffer is *not* necessarily exceeded by the inflow rates of the sources.

2. THE MODEL

Consider a fluid queueing system with an infinite storage capacity and constant, unit, output rate. This system receives input from N independent on/off sources. Source j has mutually independent alternating silence periods S_{ij} and activity periods A_{ij} , $j = 1, \dots, N$, $i = 1, 2, \dots$. Source j constantly transmits at rate r_j when active, so source j feeds $r_j A_{ij}$ traffic into the buffer during its i th activity period. The silence periods

S_{ij} have a negative exponential distribution with mean $1/\lambda_j$, and the activity periods A_{ij} have distribution $A_j(\cdot)$ with $A_j(0+) = 0$ and with mean α_j and Laplace-Stieltjes Transform (LST) $\alpha_j(\cdot)$. We assume that $r_j > 1$, $j = 1, \dots, N$. This assumption is restrictive and can be relaxed (see Section 6, and also Section 5 of [9]); we hope to discuss this issue in more detail in a future study.

The total traffic load offered to the buffer per unit time is assumed to be less than one:

$$L := \sum_{j=1}^N \frac{r_j \alpha_j \lambda_j}{1 + \alpha_j \lambda_j} < 1. \quad (2.1)$$

This is the ergodicity condition, cf. [9].

3. THE FLUID QUEUE

Let V_t denote the content of the buffer at time t . Assume that the buffer is empty at time zero. We call C_n , $n = 1, 2, \dots$, the length of the n th cumulative activity period after zero; this is a period in which at least one source is active. Such a period is followed by a period, with length I_{n+1} , in which all sources are silent. Denote by B_n the net increment of the buffer content during C_n , and by W_n the buffer content at the beginning of the n th cumulative activity period. It is easily seen (cf. Figure 1 and [9]) that

$$\begin{aligned} W_{n+1} &= \max[0, W_n + B_n - I_{n+1}], \quad n = 1, 2, \dots, \\ W_1 &= 0. \end{aligned} \quad (3.1)$$

The assumptions in the previous section about the silence periods imply that I_1, I_2, \dots

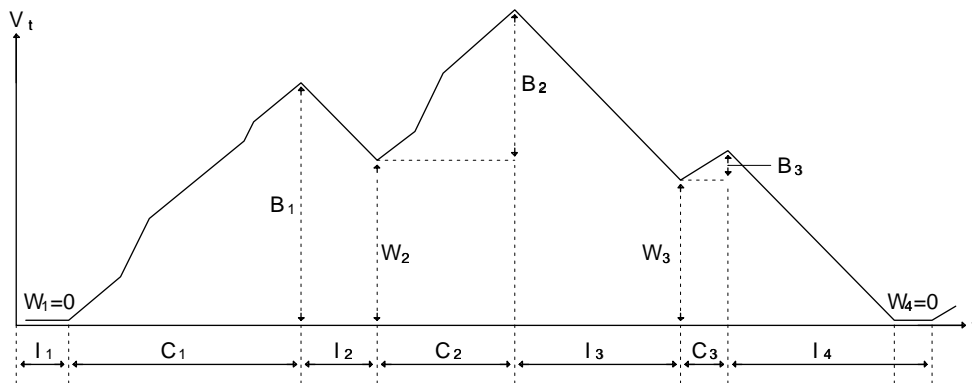


Figure 0.1: The buffer content

are independent and negative exponentially distributed with mean $1/\Lambda := 1/[\lambda_1 + \dots + \lambda_N]$. Furthermore, B_1, B_2, \dots are independent random variables; and W_n, B_n and I_{n+1} are mutually independent. Hence the sequence W_n , $n = 1, 2, \dots$, has the

same stochastic behaviour as the sequence of actual waiting times in an M/G/1 queue with arrival rate Λ and with service time distribution $B(\cdot)$, the distribution of B_n . So once we have determined $B(\cdot)$, M/G/1 theory tells us the steady-state distribution, $W(\cdot)$, of the buffer content at the start of cumulative activity periods. In this paper we shall concentrate on $W(\cdot)$. Knowledge about its behaviour will also enable us to draw conclusions about the buffer content distribution at arbitrary epochs; cf. Remark 4.2. The above relation to the M/G/1 queue allows us to obtain both qualitative and quantitative results. Firstly it gives access to Cohen's result (1.2): it now suffices to show that $B^c(t) \in R_{-\nu}$ in order to show that $W^c(t) \in R_{1-\nu}$. But it will even be possible to derive an explicit expression for the LST of $W(t)$ (see Sections 4 and 5).

Determination of $B(\cdot)$ is a far from trivial problem if $N > 1$, because the number of active sources may fluctuate during a cumulative activity period. Cohen solves this problem in [9], obtaining the LST of the joint distribution of B and C (B and C denote generic random variables with joint distribution the joint limiting distribution of the random variables B_n and C_n). With $\text{Re } \omega \geq 0$, $\text{Re } s > 0$, and some $y > 0$, according to [9],

$$\frac{1}{s + \Lambda(1 - \mathbb{E}[e^{-\omega(B+C)-sC}])} = \int_{t=0}^{\infty} e^{-st} \prod_{j=1}^N \left\{ \frac{1}{2\pi i} \int_{-i\infty+y}^{i\infty+y} \frac{e^{ut}}{u + \lambda_j(1 - \alpha_j\{\omega r_j + u\})} du \right\} dt. \quad (3.2)$$

One can obtain the LST $\mathbb{E}[e^{-\omega B}]$, if the validity of (3.2) can be extended to the domain $\text{Re } \omega \geq 0$, $\text{Re } s > -K - \text{Re } \omega$ for some $K > 0$ so that $s = -\omega$ can be put; we shall return to this below Formula (4.7).

As a by-result of (3.2), Cohen [9] shows that

$$\Lambda \mathbb{E}[C] = -1 + \prod_{j=1}^N (1 + \alpha_j \lambda_j), \quad (3.3)$$

$$\Lambda \mathbb{E}[B] = 1 - (1 - L) \prod_{j=1}^N (1 + \alpha_j \lambda_j). \quad (3.4)$$

Note that the ergodicity condition $L < 1$ (cf. (2.1)) implies $\Lambda \mathbb{E}[B] < 1$.

4. ONE RV SOURCE AND $N - 1$ EXPONENTIAL SOURCES

In this section we assume that the first source has activity period distribution $A_1(t)$, with $A_1^c(t) \in R_{-\nu}$, whereas sources $2, \dots, N$ have a negative exponential activity period distribution; the silence periods of all sources are negative exponentially distributed. From [3, 5] we know that $W^c(t) \in R_{1-\nu}$ in the absence of the second source. Below we derive the LST of $W(t)$ for the present model; the obtained expression will allow us

to show that $W^c(t) \in R_{1-\nu}$ even in the presence of sources $2, \dots, N$. Moreover, it will also lead to the reverse statement that $A_1^c(t) \in R_{-\nu}$ if $W^c(t) \in R_{1-\nu}$.

Theorem 4.1 For the case of N sources with exponential silence periods, and with exponential activity periods for sources $2, \dots, N$, and for $t \rightarrow \infty$, $l(t) \in R_0$, and $\nu > 1$,

$$A_1^c(t) = (\nu - 1)(\alpha_1/t)^\nu l(t) \Leftrightarrow W^c(t) = \frac{\lambda_1 M}{(1 - L)(1 + \alpha_1 \lambda_1)} (M/t)^{\nu-1} l(t), \quad (4.1)$$

with

$$M := \alpha_1(r_1 - 1 + \sum_{j=2}^N \frac{\alpha_j \lambda_j r_j}{1 + \alpha_j \lambda_j}). \quad (4.2)$$

Proof Starting-point is Formula (3.2). It will be used to obtain $E[e^{-\omega B}]$, and then $\int_{t=0-}^{\infty} e^{-\omega t} dW(t)$. We first evaluate the j th term, X_j , $j = 2, \dots, N$, of the product in the right-hand side (RHS) of (3.2): for $t \geq 0$, $\text{Re } \omega \geq 0$, and some $y > 0$,

$$\begin{aligned} X_j &:= \frac{1}{2\pi i} \int_{-i\infty+y}^{i\infty+y} \frac{e^{ut}}{u + \lambda_j(1 - \alpha_j\{\omega r_j + u\})} du = \\ &= \frac{1}{2\pi i} \int_{-i\infty+y}^{i\infty+y} \frac{e^{ut}(1 + \alpha_j \omega r_j + \alpha_j u)}{(u + \lambda_j)(1 + \alpha_j \omega r_j + \alpha_j u) - \lambda_j} du = \\ &= \frac{1}{2\pi i \alpha_j} \int_{-i\infty+y}^{i\infty+y} \frac{e^{ut}(1 + \alpha_j \omega r_j + \alpha_j u)}{(u - u_+^{(j)})(u - u_-^{(j)})} du, \end{aligned} \quad (4.3)$$

with poles, for $j = 2, \dots, N$:

$$u_{\pm}^{(j)} := \frac{-[1 + \alpha_j \omega r_j + \alpha_j \lambda_j] \pm [(1 + \alpha_j \omega r_j + \alpha_j \lambda_j)^2 - 4\alpha_j^2 \lambda_j \omega r_j]^{1/2}}{2\alpha_j}. \quad (4.4)$$

Note that both poles $u_{\pm}^{(j)}$ of the integrand in (4.3) are in the left half-plane. This is obvious for real ω . To see it for complex ω , apply Rouché's theorem. Consider the region G inside the closed contour \hat{G} that consists of the line from $-iR$ to iR and a semi-circle to its left with radius R , and then let $R \rightarrow \infty$. The functions $f_1(u) := (u + \lambda_j)(1 + \alpha_j \omega r_j + \alpha_j u)$ and $f_2(u) := \lambda_j$ (appearing in the second line of (4.3)) are analytic in G and continuous up to the boundary, while $|f_1(u)| > |f_2(u)|$ on \hat{G} ; the statement now follows because $f_1(u)$ has two zeros in G .

Subsequently take the contour integral of the integrand in (4.3) along \hat{G} . The contribution along the semi-circle disappears for $R \rightarrow \infty$, and what remains is the sum of the residues at the poles $u_+^{(j)}$ and $u_-^{(j)}$:

$$\begin{aligned} X_j &= \frac{1 + \alpha_j \omega r_j + \alpha_j u_+^{(j)}}{(u_+^{(j)} - u_-^{(j)})\alpha_j} e^{u_+^{(j)} t} - \frac{1 + \alpha_j \omega r_j + \alpha_j u_-^{(j)}}{(u_+^{(j)} - u_-^{(j)})\alpha_j} e^{u_-^{(j)} t} = \\ &= Y_+^{(j)} e^{u_+^{(j)} t} + Y_-^{(j)} e^{u_-^{(j)} t}, \quad t \geq 0, \text{ Re } \omega \geq 0; \end{aligned} \quad (4.5)$$

here the definition of $Y_+^{(j)}$ and $Y_-^{(j)}$ should be self-evident from the above formula. Substitution in (3.2) gives: for $\text{Re } \omega \geq 0$, $\text{Re } s > 0$, and some $y > 0$,

$$\frac{1}{s + \Lambda(1 - \mathbb{E}[e^{-\omega(B+C)-sC}])} = \int_{t=0}^{\infty} e^{-st} \prod_{j=2}^N \{Y_+^{(j)} e^{u_+^{(j)}t} + Y_-^{(j)} e^{u_-^{(j)}t}\} \frac{1}{2\pi i} \int_{-i\infty+y}^{i\infty+y} \frac{e^{ut}}{u + \lambda_1(1 - \alpha_1\{\omega r_1 + u\})} du dt. \quad (4.6)$$

The product in the RHS of (4.6) leads to 2^{N-1} different terms, each one contributing an exponential term with in the exponent a sum of $N - 1$ factors $u_{\pm}^{(j)}$. Now we make the following observation. The last integral in the RHS of (4.6) is the inverse of a Laplace transform. The first integral in each of the 2^{N-1} terms is a Laplace transform, with $(s - (u_{\pm}^{(2)} + \dots + u_{\pm}^{(N)}))t$ as argument of the exponent. Successive application of the two integral operators, i.e., a Laplace transform of the inverse of a Laplace transform, yields the integrand with u replaced by $s - u_{\pm}^{(2)} - \dots - u_{\pm}^{(N)}$ (note that $\text{Re}(s - u_{\pm}^{(2)} - \dots - u_{\pm}^{(N)}) > 0$): for $\text{Re } \omega \geq 0$, $\text{Re } s > 0$,

$$\frac{1}{s + \Lambda(1 - \mathbb{E}[e^{-\omega(B+C)-sC}])} = \sum \frac{a_2 \dots a_N}{s - (u_{\pm}^{(2)} + \dots + u_{\pm}^{(N)}) + \lambda_1(1 - \alpha_1\{\omega r_1 + s - (u_{\pm}^{(2)} + \dots + u_{\pm}^{(N)})\})}. \quad (4.7)$$

Here the sum in the RHS is taken over all 2^{N-1} possibilities, with $a_j = Y_+^{(j)}$ respectively $Y_-^{(j)}$ and correspondingly $u_{\pm}^{(j)} = u_+^{(j)}$ respectively $u_-^{(j)}$.

Multiply both sides of (4.7) by $(\Lambda E B - 1)\omega$, and then take $s = -\omega$. Using the principle of permanence, one can show that it is allowed to continue (4.7) analytically to such values of s . The LST $\alpha_1\{\cdot\}$ in the RHS of (4.7) is analytic for these values of s and ω . The LHS of (4.7) (after multiplication by ω) is analytic at least upto $s = -\omega$. In fact, multiplying (4.7) by $(\Lambda E B - 1)\omega$ and then taking $s = -\omega$ gives exactly the LST of the waiting-time distribution $W(t)$ (cf. (3.1), and use the Pollaczek-Khintchine formula for the waiting-time LST in the M/G/1 queue). For $\text{Re } \omega \geq 0$:

$$\int_{t=0-}^{\infty} e^{-\omega t} dW(t) = \frac{(1 - \Lambda E[B])\omega}{\omega - \Lambda(1 - \mathbb{E}[e^{-\omega B}])} = \quad (4.8)$$

$$(1 - \Lambda E[B])\omega \sum \frac{a_2 \dots a_N}{\omega + (u_{\pm}^{(2)} + \dots + u_{\pm}^{(N)}) - \lambda_1(1 - \alpha_1\{\omega(r_1 - 1) - (u_{\pm}^{(2)} + \dots + u_{\pm}^{(N)})\})},$$

with $\Lambda E[B]$ specified in (3.4).

We now restrict ourself to the case $\nu \in (1, 2)$, the case of long-range dependence. Given (4.1) we can write, applying Lemma A.1 (which relates the regularly varying tail behaviour of a distribution to the behaviour of its LST at the origin):

$$\alpha_1\{\omega\} - 1 + \alpha_1\omega \sim \Gamma(2 - \nu)(\alpha_1\omega)^\nu, \quad \omega \rightarrow 0. \quad (4.9)$$

It easily follows from (4.4) that, for $\omega \rightarrow 0$,

$$u_+^{(j)} = -\omega \frac{\alpha_j \lambda_j r_j}{1 + \alpha_j \lambda_j} + O(\omega^2), \quad (4.10)$$

$$u_-^{(j)} = -\frac{1 + \alpha_j \lambda_j}{\alpha_j} - \omega \frac{r_j}{1 + \alpha_j \lambda_j} + O(\omega^2). \quad (4.11)$$

Substitute these formulas in (4.8), splitting the sum in the RHS into the term T_+ in which $a_j = Y_+^{(j)}$ for all j (hence all $u_+^{(j)}$ are selected) and the remaining $2^{N-1} - 1$ terms. Observe that, for $\omega \rightarrow 0$, all those $2^{N-1} - 1$ terms contribute $O(1)$, and hence after multiplication by $(1 - \Lambda E[B])\omega$ they yield $O(\omega)$. But the term T_+ behaves like $1/(C_1\omega + C_2\omega^\nu)$ for $\omega \rightarrow 0$; the constants C_1 and C_2 are easily calculated. Taking (3.4) into account one eventually obtains, for $\omega \rightarrow 0$:

$$1 - \int_{t=0-}^{\infty} e^{-\omega t} dW(t) \sim \Gamma(2 - \nu) \frac{\lambda_1}{(1 - L)(1 + \alpha_1 \lambda_1)} \alpha_1^\nu (r_1 - 1 + \sum_{j=2}^N \frac{\alpha_j \lambda_j r_j}{1 + \alpha_j \lambda_j})^\nu \omega^{\nu-1}. \quad (4.12)$$

The \Rightarrow part of the theorem now follows by application of Lemma 4.1. The \Leftarrow part is obtained similarly, again exploiting the relation (4.8) between the LST's of $A_1(t)$ and $W(t)$. For $\nu \geq 2$ a similar result can be obtained from (4.8), at the expense of more lengthy calculations. For $\nu \in (n, n + 1)$ the only change this will bring about in the RHS of (4.12) is the replacement of the first term ("1") in the LHS of (4.12) by

$$\sum_{j=0}^{n-1} E[W^j] \frac{(-\omega)^j}{j!}. \quad (4.13)$$

Note that a less direct way to prove the theorem would have been to apply Lemma A.1 to $E[e^{-\omega B}]$ (as specified by (4.7) with $s = -\omega$) to prove that $B^c(t) \in R_{-\nu}$ if $A_1^c(t) \in R_{-\nu}$. Subsequently $W^c(t) \in R_{1-\nu}$ follows by exploiting the M/G/1-like relation (3.1) between B_n and W_{n+1} , and the GI/G/1 result (1.2).

Remark 4.1 Taking $\alpha_2 = \dots = \alpha_N = 0$ yields the case of a single source. Formula (4.12) reduces to a result of [3]:

$$1 - \int_{t=0-}^{\infty} e^{-\omega t} dW(t) \sim \Gamma(2 - \nu) \frac{\lambda_1}{1 - \lambda_1 \alpha_1 (r_1 - 1)} \alpha_1^\nu (r_1 - 1)^\nu \omega^{\nu-1}, \quad \omega \rightarrow 0. \quad (4.14)$$

The presence of the exponential sources apparently does not change the fact that $W^c(t) \in R_{1-\nu}$, but the coefficient is increased by a multiplicative factor

$$\frac{1 - \lambda_1 \alpha_1 (r_1 - 1)}{(1 - L)(1 + \alpha_1 \lambda_1)} \left(1 + \frac{1}{r_1 - 1} \sum_{j=2}^N \frac{\alpha_j \lambda_j r_j}{1 + \alpha_j \lambda_j}\right)^\nu \geq 1. \quad (4.15)$$

Remark 4.2 So far for W_n ; what about the buffer content V_t at some arbitrary time t ? In an ordinary M/G/1 queue these two quantities have the same limiting distribution. That is not true in the present model, but they are clearly quite closely related. Observe that the distribution of the sum of the independent random variables W_n and B_n has a regularly varying tail of index $1 - \nu$ iff the same holds for W_n (cf. the last statement of the appendix). Relating V_t to the immediately preceding end or beginning of an activity period (whichever has come last) makes it clear that the tail of its distribution is regularly varying of index $1 - \nu$ iff the same holds for W_n .

Let us also consider V_{max} , the maximum of the buffer content during a busy period (a period in which the buffer is never empty). Note that it is the maximum of $W_n + B_n$ over a busy period; since W_n and B_n have the same distributions as the corresponding quantities in the ordinary M/G/1 queue, the same holds for V_{max} . Cohen ([8], Formula (III.7.67)) shows that, in the M/G/1 queue,

$$\mathbf{P}\{V_{max} \geq t\} = \frac{1}{\Lambda} \frac{dW(t)}{dt}, \quad t > 0. \quad (4.16)$$

The monotone density theorem ([2], p. 39) tells us that, for $t \rightarrow \infty$, $W^c(t) \sim t^{1-\nu} l(t) \Rightarrow dW^c(t)/dt \sim (1 - \nu)t^{-\nu} l(t)$. Hence $\mathbf{P}\{V_{max} \geq t\} \in R_{-\nu}$ when $W^c(t) \in R_{1-\nu}$; the tail of the distribution of V_{max} is regularly varying *of the same index* as the tail of the activity period distribution of source 1.

5. ONE RV SOURCE AND $N - 1$ SOURCES WITH AN EXPONENTIAL ACTIVITY TAIL

In this section we relax the assumptions of the previous section in one respect: instead of assuming that the activity periods of sources $2, \dots, N$ are exponentially distributed, we assume that these periods have distributions with an exponential tail. More precisely, we assume that their LST's are rational functions:

$$\alpha_j\{\omega\} = \frac{\alpha_{jN}\{\omega\}}{\alpha_{jD}\{\omega\}}, \quad j = 2, \dots, N. \quad (5.1)$$

Here $\alpha_{jD}\{\omega\}$ (D for denominator) is a polynomial in ω of degree n_j , and $\alpha_{jN}\{\omega\}$ (N for numerator) is a polynomial in ω of degree at most $n_j - 1$ (for simplicity we assume that $A_j(0+) = 0$). Since $\alpha_j\{\omega\}$ is analytic in $\text{Re } \omega \geq 0$, the n_j zeros of $\alpha_{jD}\{\omega\}$ all lie in the left half-plane. Partial fraction expansion shows that $A_j(t)$ is a mixture of exponential distributions, and hence has an exponential tail. It is our goal in this section to prove:

Theorem 5.1 Theorem 4.1 remains valid when the condition that sources $2, \dots, N$ have exponential activity periods is replaced by the condition that the LST's of the

activity period distributions of sources $2, \dots, N$ are rational functions.

Proof The proof follows the same steps as those of Theorem 4.1. However, the steps are somewhat more complex. For expository reasons, we have therefore decided to present the case of exponential activity period distributions first, in Section 4.

Starting-point again is Formula (3.2). The j th term of the product in its RHS is again denoted by X_j . Consider the denominator $d_j(u) := u + \lambda_j(1 - \alpha_j\{\omega r_j + u\})$ of its integrand. It is easily seen that $d_j(u)$ has no zeros for $\text{Re } u > 0$, with $\text{Re } \omega \geq 0$. But substitution of (5.1) readily shows that the integrand of X_j , viz., $e^{ut}/d_j(u)$, has a denominator which is a polynomial of degree $n_j + 1$; so the integrand has exactly $n_j + 1$ poles. Hence these must all be located in the left half-plane or on the imaginary axis. Denote them by $u_1^{(j)}, \dots, u_{n_j+1}^{(j)}$ (in the notation we suppress their dependence on ω). Similarly to (4.5) we can write X_j as the sum of the $n_j + 1$ residues at the poles $u_k^{(j)}$:

$$X_j = \sum_{k=1}^{n_j+1} \frac{e^{u_k^{(j)}t}}{1 - \lambda_j \alpha_j' \{\omega r_j + u_k^{(j)}\}}. \quad (5.2)$$

Substitution in (3.2) yields, as in (4.6), a product of $N - 1$ terms; but now those terms have $n_j + 1$ instead of 2 components. Accordingly, the product in the RHS of (4.6) is replaced by $(n_2 + 1) \dots (n_N + 1)$ terms, each one contributing an exponential term with in the exponent a sum (over j) of $N - 1$ factors $u_k^{(j)}$. We recognize here $(n_2 + 1) \dots (n_N + 1)$ Laplace transforms of inverses of Laplace transforms, leading to the following generalization of (4.7): for $\text{Re } \omega \geq 0$, $\text{Re } s > 0$,

$$\frac{1}{s + \Lambda(1 - \mathbb{E}[e^{-\omega(B+C)-sC}])} = \sum \frac{a_2 \dots a_N}{s - (u_{k_2}^{(2)} + \dots + u_{k_N}^{(N)}) + \lambda_1(1 - \alpha_1\{\omega r_1 + s - (u_{k_2}^{(2)} + \dots + u_{k_N}^{(N)})\})}. \quad (5.3)$$

Here the sum in the RHS is taken over all $(n_2 + 1) \dots (n_N + 1)$ possibilities, with a_j successively denoting the residue in X_j of the pole $u_k^{(j)}$, $k = 1, \dots, n_j + 1$. With this proviso Formula (4.8) is also extended in a straightforward way: take $s = -\omega$ in (5.3), and multiply by $(\Lambda \mathbb{E}[B] - 1)\omega$.

The next step is to let $\omega \rightarrow 0$. As in the proof of Theorem 4.1, we first restrict ourself to the case $\nu \in (1, 2)$, so that (4.9) holds. Observe that $d_j(0) = 0$, and hence $u_1^{(j)} = 0$ for $\omega = 0$. This is a simple zero; $u_k^{(j)} \neq 0$ for $\omega = 0$, for $k = 2, \dots, N$. For $\omega \rightarrow 0$ we can write, analogously to (4.10) and (4.11):

$$u_1^{(j)} = -\omega \frac{\alpha_j \lambda_j r_j}{1 + \alpha_j \lambda_j} + O(\omega^2), \quad (5.4)$$

$$u_k^{(j)} = G_{kj} + O(\omega), \quad k = 2, \dots, n_j + 1, \quad (5.5)$$

with $G_{kj} \neq 0$.

Split the sum in the RHS of (5.3) into the term T_+ in which all $u_1^{(j)}$ are taken, and the remaining $(n_2 + 1) \dots (n_N + 1) - 1$ terms. For $\omega \rightarrow 0$, all the latter terms contribute $O(1)$ whereas T_+ behaves like $1/(D_1\omega + D_2\omega^\nu)$ for $\omega \rightarrow 0$. A rather tedious calculation finally shows that (4.12) again holds. The proof of the theorem can now be completed in the same way as the proof of Theorem 4.1.

Remark 5.1 Theorem 5.1 shows that the result of Theorem 4.1 is not changed by allowing more general activity period distributions (as long as they have an exponential tail). For $\nu \in (n, n + 1)$, $n \geq 2$, the higher moments of the activity period distributions will affect the terms of the development (4.13), but in the term with $\omega^{\nu-1}$ only the factors $\alpha_j \lambda_j$ play a role.

6. A CONNECTION ADMISSION CONTROL RULE

The goal of this section is to formulate a simple rule for deciding whether an additional source may be admitted to the system. We base this rule on the insight that has been developed in [3] and the present paper, concerning conditions under which relatively large buffers are required in broadband communication networks: the buffer content distribution has a regularly varying tail if each source, when active, transmits at a rate that exceeds the output rate of the buffer ($r_i > 1$), while at least one source has a regularly varying activity period distribution. We first pose the question, whether regular variation can occur even if not all these conditions are satisfied.

(i) For $N = 1$ the conditions cannot be relaxed. The buffer is permanently empty when $r_1 \leq 1$, and if $r_1 > 1$ then the buffer content distribution has a regularly varying tail iff the activity period distribution has a regularly varying tail (cf. [3] and Remark 4.1).

(ii) If $N \geq 2$ and $r_1 > 1$ while $A_1^c(t) \in R_{-\nu}$ (the other activity period distributions having exponential tails), then $W^c(t) \in R_{1-\nu}$ even if not all r_2, \dots, r_N are larger than one. This can be seen as follows. Using the bounding approach in Section 2 of Choudhury and Whitt [5], one obtains a *lower* bound for the complementary buffer content distribution by choosing $r_2 = \dots = r_N = 0$. But that choice corresponds to the model with only source 1 (see also Remark 4.1), for which it has been established that $W^c(t) \in R_{1-\nu}$.

Note Assuming monotonicity of $W^c(t)$ in r_2, \dots, r_N (which we have not proven), one can increase r_2, \dots, r_N so that they are all larger than one, and thus obtain an *upper* bound for $W^c(t)$. According to Theorem 5.1, this upper bound is also regularly varying of index $1 - \nu$; hence it is of the *same* index as the lower bound.

Remark 6.1 So far we have not yet discussed the situation in which more than

one source has a regularly varying activity period distribution. For $N = 2$, with $A_i^c(t) \in R_{-\nu_i}$ and $r_i > 1$, $i = 1, 2$, it will be proven in a forthcoming paper that $W^c(t) \in R_{1-\min(\nu_1, \nu_2)}$. In relation to this it is interesting to consider the following limiting case. Take $r_i \rightarrow \infty$ and $\alpha_i \rightarrow 0$ such that $0 < \beta_i := r_i \alpha_i < \infty$ for $i = 1, \dots, N$. This corresponds to having instantaneous input. In the limit the fluid queue becomes an ordinary M/G/1 queue with service time distribution $\sum_{i=1}^N \frac{\lambda_i}{\Lambda} B_i(t)$, with $\Lambda := \sum_{i=1}^N \lambda_i$ and $B_i(t) := \lim A_i(t/r_i)$, the limit being taken in the way indicated above. For this M/G/1 queue one can directly apply Cohen's result (1.2) to conclude the following. If $B_i^c(t) \in R_{-\nu_i}$ for $i \in U$, a subset of $(1, \dots, N)$, then $W^c(t) \in R_{1-\min_{i \in U} \nu_i}$.

(iii) The most interesting case is $N \geq 2$ and $r_1 < 1$. Several things may happen.

(iii-a) At least one of the sources $2, \dots, N$ has a regularly varying activity period distribution. If source 2, say, has $A_2^c(t) \in R_{-\nu_2}$ but $r_2 > 1$, then the same bounding technique of [5] shows that the complementary buffer content distribution is bounded from below by a regularly varying function of index $1 - \nu_2$ (take $r_j = 0$ for $j \neq 2$). We refrain for the moment from discussing the case in which also $r_2 \leq 1$; we return to it in Conjecture 6.1 below.

(iii-b) All sources $2, \dots, N$ have activity period distributions with an exponential tail (this is the case of Section 5, except that we now assume that $r_1 \leq 1$, and make no assumptions about r_2, \dots, r_N). Without source 1 (so with $r_1 = 0$), the buffer content distribution has an exponential tail. Now take $0 < r_1 \leq 1$, and consider the probability $p_k(X) = \mathbf{P}(W_k > X)$ of the buffer content exceeding some large value X at the beginning of the k -th cumulative activity period. Assume that this epoch occurs during a busy period of the fluid queue that has included one or more activity periods of source 1; otherwise it is as if source 1 is absent, and then $p_k(X)$ will certainly be exponentially small in X . Since for activity periods A_{i1} of source 1 one has $\mathbf{P}\{A_{11} + \dots + A_{n1} > x\} \approx n\mathbf{P}\{A_{11} > x\}$ (because of the regular variation), it is most likely that *only one* long activity period of source 1 is responsible for W_k exceeding the large value X . But even if such a long activity period of source 1 occurs, it does not necessarily lead to an increment of the buffer content (measured between the beginning and the end of that activity period), because $r_1 \leq 1$. Since we are considering a long activity period, say A_{11} , the law of large numbers tells us that the buffer increment during A_{11} equals approximately $A_{11}[r_1 - 1 + \sum_{j=2}^N \alpha_j \lambda_j r_j / (1 + \alpha_j \lambda_j)]$ (cf. the factor in (4.12)!). This leads us to the conjecture that if $r_1 - 1 + \sum_{j=2}^N \alpha_j \lambda_j r_j / (1 + \alpha_j \lambda_j) > (\leq) 0$ then $A_1^C(t) \in R_{-\nu}$ does (does not) imply $W^c(t) \in R_{1-\nu}$. So while the presence of exponential sources in the case $r_1 > 1$ only influences the coefficient of the regularly varying function $W^c(t)$ (see the factor in (4.12)), in the case $r_1 \leq 1$ these exponential sources may make the difference between $W^c(t)$ being regularly varying or not.

Combining this conjecture with the proven results, we go one step further and state

the following

Conjecture 6.1 Let U be the subset of $(1, \dots, N)$ for which the activity period distributions are regularly varying (of index $-\nu_i$ for source i). Consider the subset $U_1 \subset U$ of indices i for which holds:

$$r_i + \sum_{j \neq i} \frac{\alpha_j \lambda_j r_j}{1 + \alpha_j \lambda_j} > 1. \quad (6.1)$$

Then $W^c(t) \in R_{1 - \min_{i \in U_1} \nu_i}$.

This conjecture indicates in what direction we believe that the condition $r_i > 1$ in the present paper can be relaxed.

Connection admission control

The conjecture gives rise to a simple connection admission control rule. To decide whether an additional source $N + 1$ may be admitted to the system, check whether its addition gives rise to a regularly varying buffer content distribution; if it does, the source is not admitted.

More specifically, one distinguishes between the cases in which the activity period distribution of source $N + 1$ is regularly varying or not. If it is, one checks whether (6.1) holds for $i = N + 1$. If it is not, one checks for all $i \in U$ whether (6.1) holds, with $j = N + 1$ included in the summation. This will indicate whether addition of source $N + 1$ gives rise to a regularly varying buffer content distribution.

Of course the above connection admission control rule is a bit simplistic. For instance, a source i with a regularly varying activity period distribution and $r_i > 1$ would never be admitted. In such a case, some form of traffic shaping should be applied first.

Remark 6.2 The concept of effective bandwidth has proved to be very valuable for connection admission control. Gibbens and Hunt [11] and Cohen [9] have derived expressions for the effective bandwidth in fluid queues; but those expressions are not valid for the case of regularly varying activity period distributions. For the latter case, (6.1) might be a useful alternative. As remarked by Gibbens and Hunt [11], p. 22, the factor $\alpha_j \lambda_j r_j / (1 + \alpha_j \lambda_j)$ is the *mean bandwidth* of source j , a lower bound on the effective bandwidth.

ACKNOWLEDGEMENT

The author is indebted to J.W. Cohen for interesting discussions.

REFERENCES

1. D. Anick, D. Mitra, M.M. Sondhi (1982). *Stochastic theory of a data-handling system with multiple sources*, Bell Syst. Techn. J. 61, 1871-1894.
2. N.H. Bingham, C.M. Goldie, J.L. Teugels (1987). *Regular Variation*. Cambridge University Press, Cambridge.
3. O.J. Boxma (1996). *Fluid queues and regular variation*, CWI Report BS-R9608; to appear in *Performance Evaluation*, 1996.
4. F. Brichet, J. Roberts, A. Simonian, D. Veitch (1995). *Heavy traffic analysis of a storage model with long range dependent on/off sources*, Report FT-CNET.
5. G.L. Choudhury, W. Whitt (1995). *Long-tail buffer-content distributions in broadband networks*, Report AT&T Bell Laboratories.
6. J.W. Cohen (1973). *Some results on regular variation for distributions in queueing and fluctuation theory*, J. Appl. Probab. 10, 343-353.
7. J.W. Cohen (1974). *Superimposed renewal processes and storage with gradual input*, Stoch. Proc. and their Appl. 2, 31-58.
8. J.W. Cohen (1982). *The Single Server Queue*. North-Holland Publ. Cy., Amsterdam (2nd ed.).
9. J.W. Cohen (1994). *On the effective bandwidth in buffer design for the multi-server channels*, CWI Report BS-R9406.
10. A. Erramilli, R.P. Singh, P. Pruthi (1994). *Chaotic maps as models of packet traffic*, In: J. Labetoulle and J.W. Roberts (eds.), *The Fundamental Role of Teletraffic in the Evolution of Telecommunications Networks*, Proceedings of the 14th International Teletraffic Congress (North-Holland Publ. Cy., Amsterdam) pp. 329-338.
11. R.J. Gibbens, P.J. Hunt (1991). *Effective bandwidths for the multi-type UAS channel*, Queueing Systems 11, 17-27.
12. L. Kosten (1984). *Stochastic theory of data handling systems with groups of multiple sources*, In: *Performance of Computer-Communication Systems*, eds. W. Bux and H. Rudin (North-Holland Publ. Cy. Amsterdam) pp. 321-331.
13. P.R. Jelenković, A.A. Lazar (1995). *Subexponential asymptotics of a Markov-modulated G/G/1 queue*, Report Columbia University.
14. I. Norros (1994). *A storage model with self-similar input*, Queueing Systems 16, 387-396.
15. W.E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson (1994). *On the self-similar nature of Ethernet traffic (extended version)*, IEEE/ACM Transactions on Networking 2, 1-15.
16. W. Willinger, M.S. Taqqu, W.E. Leland, D.V. Wilson (1995). *Self-similarity in high-speed packet traffic: analysis and modeling of Ethernet traffic measurements*, Statistical Science 10, 67-85.

Appendix - Regular variation

Regular variation is an important concept in probability theory and various other fields. The main reference text is the book [2], to which we refer for proofs and detailed discussions of the results gathered below. See (1.1) for the definition of regular variation. A slowly varying function (see below (1.1)) is denoted by $l(\cdot)$.

The next lemma plays a key role in the paper. It links the behaviour of the tail of a probability distribution $F(t)$ on $[0, \infty)$ to the behaviour of its LST $\phi(s)$ for $s \rightarrow 0$. It is part of Theorem 8.1.6 on p. 333/334 of [2], originally due to Bingham and Doney. Assume that the first n moments μ_1, \dots, μ_n of $F(\cdot)$ are finite. Define

$$\phi_n(s) := (-1)^{n+1} \left[\phi(s) - \sum_{j=0}^n \mu_j \frac{(-s)^j}{j!} \right]. \quad (6.2)$$

Lemma A.1 Let $\nu = n + \psi$ with $0 < \psi < 1$. The following are equivalent:

$$\phi_n(s) \sim s^\nu l(1/s), \quad s \rightarrow 0, \quad (6.3)$$

$$F^c(t) \sim \frac{(-1)^n}{\Gamma(1-\nu)} t^{-\nu} l(t), \quad t \rightarrow \infty. \quad (6.4)$$

In Theorem 8.1.6 of [2] the somewhat more complicated cases $\psi = 0, 1$ are also discussed.

In the paper we also use the following two results, that follow quickly from Lemma A.1 (but can also be obtained by more direct probabilistic and analytic arguments).

$$(i) \quad F^c(t) \in R_{-\nu} \Rightarrow \int_t^\infty F^c(x) dx \in R_{1-\nu}, \quad (6.5)$$

and in fact (cf. [2], p. 28, as part of the Karamata theorem):

$$\frac{tF^c(t)}{\int_t^\infty F^c(x) dx} \rightarrow \nu - 1, \quad t \rightarrow \infty. \quad (6.6)$$

This follows from Lemma A.1 by observing that $\int_0^t F^c(x) dx$ has LST $[1 - \phi(s)]/s$.

(ii) The sum of two independent nonnegative random variables with regularly varying distributions of index ν_1 respectively ν_2 is again regularly varying, of index $\max(\nu_1, \nu_2)$.