2005

Queuing Systems with Multiple FBM-Based Traffic Models

Mihaela Teodora Matache
University of Nebraska at Omaha, dvelcsov@unomaha.edu

Valentin Matache
University of Nebraska at Omaha, vmatache@unomaha.edu

QUEUING SYSTEMS FOR MULTIPLE FBM-BASED TRAFFIC MODELS

MIHAELA T. MATACHE[1] and VALENTIN MATACHE[2]

Abstract

A multiple Fractional Brownian Motion (FBM) based traffic model of the following form is considered

$$A(t) = mt + \sum_{j=1}^{M} \sigma_j B^{H_j}(t) + \tau W(t).$$

Here $B^{H_j}(t)$ are independent FBMs with Hurst parameters $1/2 < H_j < 1$ and $W(t)$ is a Brownian Motion independent of the FBMs. Various lower bounds for the overflow probability of the associated queuing system are obtained. Based on a probabilistic bound for the busy period of an ATM queuing system associated to a multiple FBM-based input traffic, a minimal dynamic bufer allocation function (DBAF) is obtained and a DBAF-allocation algorithm is designed. The purpose is to create an upper bound for the queuing system associated with the traffic. This upper bound, called DBAF, is a function of time, dynamically bouncing with the traffic. An envelope process associated to the multiple FBM-based traffic model is introduced and used to estimate the queue size of the queuing system associated to that traffic model.

Key words. queuing systems, overflow probability, Fractional Brownian Motion.

AMS Subject Classification. 60K25

———————

[1]Department of Mathematics, The University of Nebraska at Omaha, Omaha, NE 68182, USA, dmatache@mail.unomaha.edu

[2]Department of Mathematics, The University of Nebraska at Omaha, Omaha, NE 68182, USA, vmatache@mail.unomaha.edu

1. Introduction

Over the last years many studies ([1], [3], [7], [10], [11], [13], [16], [19], [20]) have shown that packet/cell traffic through telecommunication networks (like Ethernet, LAN, WAN, ISDN, ATM) exhibits long-range dependence and self-similarity, i.e. the autocorrelation function decays asymptotically as a power function with negative exponent, and the traffic looks the same when measured over various time scales.

These studies also show that traditional models used in traffic modeling, like Poisson models, cannot capture observed features of the telecommunication traffic. New parsimonious models are proposed in [5], [11], [14]. They involve the fractional Brownian motion (FBM) process whose properties make it a natural choice in modeling packet/cell traffic.

At the same time experimental and analytical studies ([5], [6], [12], [14]) show that the long-range dependence property can create big packet/cell losses in queuing systems, and have an important impact on engineering problems like buffer allocation or admission control. Most evidence is obtained exclusively through simulation experiments using trace data, since no queuing solution for fractional Brownian (FB) traffic models is known. However, in [5], [12], [14], approximations and bounds for the overflow probabilities in a queuing system driven by a FB traffic are presented. We will refer to such a traffic model using the abbreviation FB traffic, as opposed to the term multiple FBM-based traffic model which will designate a model where a standard Brownian motion is added to a superposition of independent FBMs.

On the other hand, when strongly variable but short-range dependent traffic is aggregated with long-range dependent traffic, the mixture could be described in a FB traffic only by reducing the available bandwidth, and the model would not be satisfactory at small time scales. In order to avoid this inflexibility, a possible solution would be to add a Brownian component to the FB traffic model ([17]).

For a classical FB traffic model, the following authors: Norros, Duffield, O'Connell, Mayor and Silvester ([5], [12], [14]), have performed analysis of the associated queuing system and obtained results related to the estimation of the busy period and the overflow probabilities.

In the sequel we extend these results to a multiple FBM-based traffic model, (see equality (2) in Section 2 for its exact definition). In Section 2, besides introducing the basic notions and setting up the notation, we obtain a lower bound for the overflow probability of the associated queuing system, (subsection 2.1). Also asymptotic lower bounds for the same probability are obtained (Theorem 2, subsection 2.2). In Section 3, based on a probabilistic upper bound for the busy period of an ATM queuing system with a multiple FBM-based input traffic model we introduce the notion of dynamic buffer

allocation function (DBAF) and show that a least DBAF exists (Proposition 2, subsection 3.1). In subsection 3.2 we use the least DBAF to design a dynamic buffer allocation algorithm. The algorithm is illustrated by graphs exhibiting MATLAB-simulated traffic (Figures 1 through 3). In Section 4 we introduce an envelope process associated to the multiple FBM-based traffic model and use it to obtain an upper bound for the busy period of an ATM queuing system. We obtain an upper bound for the queue size using the envelope process.

2. A Multiple FBM - Based Traffic Model

It has been observed that sometimes the FB traffic is not sufficient to model the traffic at small time scales when strongly variable, but short-range dependent traffic is mixed with long-range dependent traffic aggregated from a large number of sources. To overcome this difficulty, a natural model can be given by adding to the FB traffic model a short-range dependent component as in the following stochastic process

$$(1) \qquad A(t) = mt + \sqrt{ma}B^H(t) + \sqrt{mb}W(t).$$

Here m, a, b are positive constants, $\{B^H(t), t \geq 0\}$ is a standard FBM with Hurst parameter $H \in (\frac{1}{2}, 1)$, and $\{W(t), t \geq 0\}$ is a standard BM, independent of the FBM (see [17]). In this paper we will study a more general type of model with several FBM's given by the following equality

$$(2) \qquad A(t) = mt + \sum_{j=1}^{M} \sigma_j B^{H_j}(t) + \tau W(t).$$

As above m, σ_j, and τ are positive constants, $\{B^{H_j}(t), t \geq 0\}$, $j = 1, 2, \dots, M$ are independent standard FBMs with Hurst parameters H_j, $1/2 < H_j < 1$, $\forall j = 1, \dots, M$ respectively, and $\{W(t), t \geq 0\}$ is a standard BM independent of the FBMs. We will refer to this traffic model as **a multiple FBM-based traffic model** as opposed to the simple FBM-based traffic model (1). Observe that the class of the stochastic processes of type (2) is closed under superposition.

REMARK. *Let $A_i(t), i = 1, 2, \dots N$, be the i-th FBM-based input traffic process defined as*

$$A_i(t) = m_i t + \sum_{j=1}^{M} \sigma_{ij} B_i^{H_j}(t) + \tau_i W_i(t), \qquad t \geq 0,$$

where m_i, σ_{ij}, τ_i are positive numbers, the processes $\{B_i^{H_j}(t), t \geq 0\}$ are independent standard FBMs with Hurst parameters H_j, and $\{W_i(t), t \geq 0\}$ are independent standard BMs which are also independent of the FBMs. Then the superposition

$$A(t) := \sum_{i=1}^{N} A_i(t)$$

can be written as

$$A(t) = mt + \sum_{j=1}^{M} \sigma_j B^{H_j}(t) + \tau W(t)$$

where $m = \sum_{i=1}^{N} m_i$, $\sigma_j = \sqrt{\sum_{i=1}^{N} \sigma_{ij}^2}$, $\tau = \sqrt{\sum_{i=1}^{N} \tau_i^2}$, $\{B^{H_j}(t), t \geq 0\}, j = 1, 2, \ldots, M$ are standard FBMs with parameters H_j, and $\{W(t), t \geq 0\}$ is a standard BM independent of the FBMs.

The proof is a simple application of basic properties of independent FBMs and BMs, and is left to the reader. It is known that the corresponding stationary queuing model can be described by the process $\{V(t), t \geq 0\}$, where

(3) $$V(t) = \sup_{s \leq t}(A(t) - A(s) - C(t - s)), \qquad t \geq 0.$$

In (3), C represents the constant service rate and satisfies $C > m$, and $m > 0$ is the mean input rate given in (2). Formula (3) gives the workload or the virtual waiting time in a FIFO (first in - first out) queuing system with the previously described parameters.

2.1. Lower Bound for the Complementary Distribution Function of the Queue Level In what follows we determine a lower bound for the overflow probability of the queue. The overflow probability, or the cell loss ratio is an important Quality of Service (QoS) parameter in telecommunications. The overflow probability ϵ, is defined as follows

$$\epsilon := \mathbf{P}(V(t) > x), \qquad t \geq 0.$$

Here x denotes a given buffer size. In what follows we will use the notation $\bar{\Phi}(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$, designating the complementary cumulative distribution function of a standard normal random variable. We use $\bar{\Phi}$ to obtain a lower estimate of the overflow probability.

THEOREM 2.1. *Let* $\{A(t), t \geq 0\}, A(t) = mt + \sum_{j=1}^{M} \sigma_j B^{H_j}(t) + \tau W(t)$ *be a multiple FBM-based traffic model and* $\{V(t), t \geq 0\}$ *be the stationary queuing process defined in* (3). *Then*

4

$$\mathbf{P}(V(t) > x) \geq \bar{\Phi}(\psi(u_1)), \qquad t \geq 0$$

where

$$\psi(t) = \frac{x - (m - C)t}{\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t}}$$

and u_1 is the unique real root of the following equation

$$(C - m)\sum_{j=1}^{M} \sigma_j^2 (1 - H_j) u^{2H_j} + \frac{1}{2}(C - m)\tau^2 u - \frac{1}{2}\tau^2 x - x \sum_{j=1}^{M} \sigma_j^2 H_j u^{2H_j - 1} = 0.$$

Proof. Observe that

$$\{\sup_{s \leq t}(A(t) - A(s) - C(t - s)) > x\} =$$

$$= \bigcup_{s \leq t}\{A(t) - A(s) - C(t - s) > x\}.$$

Using this equality we deduce that

$$\mathbf{P}(\sup_{s \leq t}(A(t) - A(s) - C(t - s)) > x) \geq$$

$$\geq \sup_{s \leq t}\mathbf{P}(A(t) - A(s) - C(t - s) > x) =$$

$$= \sup_{s \leq t}\mathbf{P}(A(t) - A(s) - C(t - s) - (m - C)(t - s) > x - (m - C)(t - s)) =$$

$$= \sup_{s \leq t}\int_{x - (m - C)(t - s)}^{\infty} \frac{1}{\sqrt{2\pi}\sigma(s)}\exp\{-\frac{y^2}{2\sigma^2(s)}\}dy$$

where $\sigma(s) = \sqrt{\sum_{j=1}^{M}\sigma_j^2(t - s)^{2H_j} + \tau^2(t - s)}$. By a change of variable we obtain

$$\mathbf{P}(\sup_{s \leq t}(A(t) - A(s) - C(t - s)) > x) \geq$$

$$\geq \sup_{s \leq t}\bar{\Phi}\left(\frac{x - (m - C)(t - s)}{\sigma(s)}\right) =$$

$$= \bar{\Phi}\left(\inf_{s \leq t}\frac{x - (m - C)(t - s)}{\sigma(s)}\right).$$

Thus we need to find $\inf_{s \leq t}\xi(s)$, where

$$\xi(s) = \frac{x - (m - C)(t - s)}{\sqrt{\sum_{j=1}^{M}\sigma_j^2(t - s)^{2H_j} + \tau^2(t - s)}}.$$

5

If we denote $u = t - s$, then by straightforward calculus considerations

$$\zeta(u) = \frac{x - (m - C)u}{\sqrt{\sum_{j=1}^{M} \sigma_j^2 u^{2H_j} + \tau^2 u}}$$

has a global minimum at some point $u_1 \in (0, \infty)$, decreases on $(0, u_1)$ and increases on (u_1, ∞). Therefore, if $t < u_1$, $\inf_{0 \leq u \leq t} \psi(u) = \psi(t)$, and if $t \geq u_1$, $\inf_{0 \leq u \leq t} \psi(u) = \psi(u_1)$. In conclusion, if $t < u_1$

$$\mathbf{P}(\sup_{s \leq t}(A(t) - A(s) - C(t - s)) > x) \geq \bar{\Phi}(\psi(t)) \geq \bar{\Phi}(\psi(u_1))$$

and if $t \geq u_1$

$$\mathbf{P}(\sup_{s \leq t}(A(t) - A(s) - C(t - s)) > x) \geq \bar{\Phi}(\psi(u_1)).$$

Clearly, u_1 is the root of $\zeta'(u)$ which, by a short computation, is the same as the unique real root of the following equation

$$(C - m)\sum_{j=1}^{M} \sigma_j^2(1 - H_j)u^{2H_j} + \frac{1}{2}(C - m)\tau^2 u - \frac{1}{2}\tau^2 x - x\sum_{j=1}^{M} \sigma_j^2 H_j u^{2H_j - 1} = 0.$$

$\square$

2.2. Asymptotic Lower Bounds for the Overflow Probability

THEOREM 2.2. *Let* $\{A(t), t \geq 0\}$ *be the multiple FBM-based input traffic process defined in* (2), *and* $\{V(t), t \geq 0\}$ *the workload process defined in* (3). *Then*

(a)

$$\liminf_{x \to \infty} \frac{\sum_{j=1}^{M} \sigma_j^2 x^{2H_j} + \tau^2 x}{x^2} \log \mathbf{P}(V(t) > x) \geq -\frac{1}{2}\frac{(C - m)^{2H}}{H^{2H}(1 - H)^{2-2H}}$$

where $H = \max\{H_j, j = 1, 2, \ldots, M\}$.
(b)

$$\liminf_{t \to \infty} \frac{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t}{t^2} \log \mathbf{P}(V(t) > x) \geq -\frac{(x + C - m)^2}{2}.$$

Proof. **(a)** It is known from [5], that

$$\liminf_{x \to \infty} h(x)^{-1} \log \mathbf{P}(V(t) > x) \geq -\inf_{u > 0} g(u)\lambda^\star(u^+).$$

6

if the following hypotheses are satisfied.

(i) There exist functions $a, v : [0, \infty) \to [0, \infty)$ that increase to infinity, such that for each $\theta \in \mathbf{R}$, the cumulant generating function defined as the limit

$$\lambda(\theta) := \lim_{t \to \infty} v(t)^{-1} \log \mathbf{E}[\exp\{\frac{\theta v(t) U(t)}{a(t)}\}]$$

exists in $[-\infty, \infty]$. Moreover, $\lambda(\theta)$ is essentially smooth and lower semi-continuous. Here $U(s) := A(t) - A(t - s) - Cs$ for some $t \geq 0$ and for $0 \leq s \leq t$.

Note that $V(t)$ can be written as $V(t) = \sup_{0 \leq s \leq t} U(s)$.

(ii) There exists $\theta > 0$ for which $\lambda(\theta) < 0$.

(iii) There exists an increasing function $h : [0, \infty) \to [0, \infty)$ such that the limit

$$g(u) := \lim_{t \to \infty} \frac{v(a^{-1}(\frac{t}{u}))}{h(t)}$$

exists for each $u > 0$, where $a^{-1}(t) := \sup_{s \geq 0}\{a(s) \leq t\}$. $\lambda^\star$ represents the Fenchel - Legendre transform of λ, i.e. the function defined as $\lambda^\star(x) := \sup_{\theta \geq 0}\{\theta x - \lambda(\theta)\}$.

The three hypotheses are satisfied using the following functions

$$a, v : [0, \infty) \to [0, \infty), \quad a(s) = s, \quad v(s) = h(s) = \frac{s^2}{\sum_{j=1}^{M} \sigma_j^2 s^{2H_j} + \tau^2 s}.$$

Under these conditions the conclusion in [5] holds, and

$$\liminf_{x \to \infty} \frac{1}{v(x)} \log \mathbf{P}(V(t) > x) \geq - \inf_{u > 0} g(u)\lambda^\star(u).$$

Observe that $\lambda^\star(x) := \sup_{\theta \geq 0}\{\theta x - \lambda(\theta)\} = \sup_{\theta \geq 0}\{-\frac{1}{2}\theta^2 + \theta(x + C - m)\} = \frac{(x+C-m)^2}{2}$. Thus

$$- \inf_{u > 0} g(u)\lambda^\star(u) = -\frac{1}{2} \inf_{u > 0} \frac{(u + C - m)^2}{u^{2-2H}} = -\frac{1}{2} \frac{(C - m)^{2H}}{H^{2H}(1 - H)^{2-2H}}.$$

In conclusion,

$$\liminf_{x \to \infty} \frac{\sum_{j=1}^{M} \sigma_j^2 x^{2H_j} + \tau^2 x}{x^2} \log \mathbf{P}(V(t) > x) \geq -\frac{1}{2} \frac{(C - m)^{2H}}{H^{2H}(1 - H)^{2-2H}}.$$

(b) For the second part of the theorem, let $a(t)$, $v(t)$, $\lambda(\theta)$, and $\lambda^\star(x)$ be as above. Since $\lambda^\star$ is continuous, one has that

$$-\lambda^\star(x) = \lim_{t \to \infty} \frac{\log \mathbf{P}(\frac{U(t)}{a(t)} > x)}{v(t)}$$

7

by the Gärtner-Ellis Theorem (see for example [4], Theorem 2.3.6, or [5]). Observe that $\mathbf{P}(V(t) > x) \geq \mathbf{P}(U(t) > x) \geq \mathbf{P}(U(t) > tx) = \mathbf{P}(\frac{U(t)}{t} > x)$ for all $t \geq 1$. Therefore

$$\frac{\log \mathbf{P}(V(t) > x)}{v(t)} \geq \frac{\log \mathbf{P}(\frac{U(t)}{a(t)} > x)}{v(t)}$$

for all $t \geq 1$. Letting $t \to \infty$ in the expression above, one gets

$$\liminf_{t \to \infty} \frac{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t}{t^2} \log \mathbf{P}(V(t) > x) \geq -\lambda^\star(x) = -\frac{(x + C - m)^2}{2}.$$

$\square$

3. Dynamic Buffer Allocation

3.1. A Probabilistic Bound   The maximum busy period of an ATM queuing system is very important since it provides a bound for the delay of the ATM cells in the queue ([18]).

If we define
$$\hat{d}_H := \inf\{t \geq 1 : \mathbf{P}(Q(t) > 0) \leq \epsilon\}$$

with $\epsilon \ll 1$, then the busy period will exceed $\hat{d}_H$ with probability $\epsilon \ll 1$. For more about this quantity we refer to [2] or [12]. Here $Q(t) = A(t) - Ct$, $\{A(t), t \geq 0\}$ is the input traffic process (2), and C the positive service rate.

PROPOSITION 3.1. $\hat{d}_H$ can be calculated by the following formula

$$\hat{d}_H = \eta^{-1}\left(\frac{\bar{\Phi}^{-1}(\epsilon)}{C - m}\right)$$

where

(4) $$\eta(t) = \frac{t}{\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t}}.$$

Proof. We have
$$\mathbf{P}(Q(t) > 0) =$$

$$= \mathbf{P}\left(\frac{\sum_{j=1}^{M} \sigma_j B^{H_j}(t) + \tau W(t)}{\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t}} > \frac{(C - m)t}{\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t}}\right) =$$

$$= \bar{\Phi}\left(\frac{(C - m)t}{\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t}}\right).$$

So

$$\hat{d}_H = \inf\{t \geq 1 : \bar{\Phi}\left(\frac{(C-m)t}{\sqrt{\sum_{j=1}^{M}\sigma_j^2 t^{2H_j} + \tau^2 t}}\right) \leq \epsilon\} =$$

$$= \inf\{t \geq 1 : \frac{t}{\sqrt{\sum_{j=1}^{M}\sigma_j^2 t^{2H_j} + \tau^2 t}} \geq \frac{\bar{\Phi}^{-1}(\epsilon)}{C-m}\}.$$

If we define

$$(4) \qquad\qquad \eta(t) = \frac{t}{\sqrt{\sum_{j=1}^{M}\sigma_j^2 t^{2H_j} + \tau^2 t}},$$

it can be easily seen that $\eta(t)$ is invertible so that we can write

$$\hat{d}_H = \inf\{t \geq 1 : t \geq \eta^{-1}\left(\frac{\bar{\Phi}^{-1}(\epsilon)}{C-m}\right)\} = \eta^{-1}\left(\frac{\bar{\Phi}^{-1}(\epsilon)}{C-m}\right).$$

Note that since $\epsilon \ll 1$, we can assume that $\epsilon < \frac{1}{2}$, so that $\bar{\Phi}^{-1}(\epsilon) > 0$. This is required since $\eta : [0, \infty) \to [0, \infty)$. $\qquad\square$

For fixed $\delta > 0$ we want to determine a positive function $M(t)$, which will be called a dynamic buffer allocation function, such that

$$\mathbf{P}(Q(t) > M(t)) \leq \delta, \qquad \delta > 0.$$

Of course, we are interested in $\delta \ll 1$, which means that there is a very small overflow probability. On the other hand, this analysis is significant for a time interval where the queuing process $\{V(t), t \geq 0\}$ can be approximated by the related process $\{Q(t), t \geq 0\}$. By our introductory comments in this subsection, this time interval is given by $\hat{d}_H$.

PROPOSITION 3.2. *There is a least dynamic buffer allocation function and it is given by*

$$\xi(t) = \bar{\Phi}^{-1}(\delta)\sqrt{\sum_{j=1}^{M}\sigma_j^2 t^{2H_j} + \tau^2 t} - (C-m)t.$$

Proof. Observe that

$$\mathbf{P}(Q(t) > M(t)) = \mathbf{P}(A(t) > M(t) + Ct) =$$

$$= \bar{\Phi}\left(\frac{M(t) + (C-m)t}{\sqrt{\sum_{j=1}^{M}\sigma_j^2 t^{2H_j} + \tau^2 t}}\right)$$

9

so $\mathbf{P}(Q(t) > M(t)) \le \delta$ is equivalent to

$$\bar{\Phi}\left(\frac{M(t) + (C-m)t}{\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t}}\right) \le \delta \Leftrightarrow$$

$$\Leftrightarrow M(t) \ge \bar{\Phi}^{-1}(\delta)\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t} - (C-m)t.$$

Then if we define

$$\xi(t) := \bar{\Phi}^{-1}(\delta)\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t} - (C-m)t$$

this function is the least dynamic buffer allocation function, since $M(t) \ge \xi(t)$.

$\square$

Elementary calculus considerations can be used to see that $\xi(t)$ has a unique positive root t_0. We would like to have $\hat{d}_H \le t_0$ in order to maintain a nonnegative buffer allocation function.

REMARK. $\hat{d}_H \le t_0$ *if and only if* $\delta \le \epsilon$.

Proof.

$$\hat{d}_H \le t_0 \Leftrightarrow \eta^{-1}\left(\frac{\bar{\Phi}^{-1}(\epsilon)}{C-m}\right) \le t_0$$

where $\eta(t)$ is the function given in (4). Since

$$\sqrt{\sum_{j=1}^{M} \sigma_j^2 t_0^{2H_j} + \tau^2 t_0} = \frac{(C-m)t_0}{\bar{\Phi}^{-1}(\delta)}$$

we obtain

$$\bar{\Phi}^{-1}(\epsilon) \le \bar{\Phi}^{-1}(\delta) \Rightarrow \delta \le \epsilon.$$

$\square$

In Figure 1 a multiple FBM-based traffic queue is simulated using MAT-LAB. The corresponding Dynamic Buffer Allocation Function is graphed to illustrate how it bounds the queue from above. The traffic utilized was of the form

$$A(t) = mt + \sigma_1 B^{H_1}(t) + \sigma_2 B^{H_2}(t) + \tau W(t).$$

10

The parameters a, b, and c which are specified in the figure have the following significance

$$\sigma_1 = \sqrt{ma}, \quad \sigma_2 = \sqrt{mc}, \quad \tau = \sqrt{mb}.$$

The probabilistic precisions ϵ and δ are specified as well. Overflows like the one in Figure 1 are rather hard to obtain even when coarser precisions like the one we utilized there are used. In most cases the queue is much smaller than the DBAF and it took hours of simulations to produce graphs where overflows occur (which would be natural, given the probabilistic methods used). In Figure 1 the DBAF is graphed over the time interval between its two roots. We wish to observe that changing parameters result in dramatic changes in the DBAF. For instance increasing the probabilistic degree of precision ϵ and δ results in very large DBAFs with large values of $\hat{d}_H$, the comment being that in practice the size of the buffer one can use is limited, so one might want to trade between quality of service and sparing buffer-space. We use this function to create an upper bound, that is a function of time, larger than the queue associated to the traffic, usable for the whole duration of the process, and dynamically bouncing with the traffic so that buffer-space could be spared and, say allotted to a different queue. We do this in the next subsection, and illustrate our construction in Figures 2 and 3.
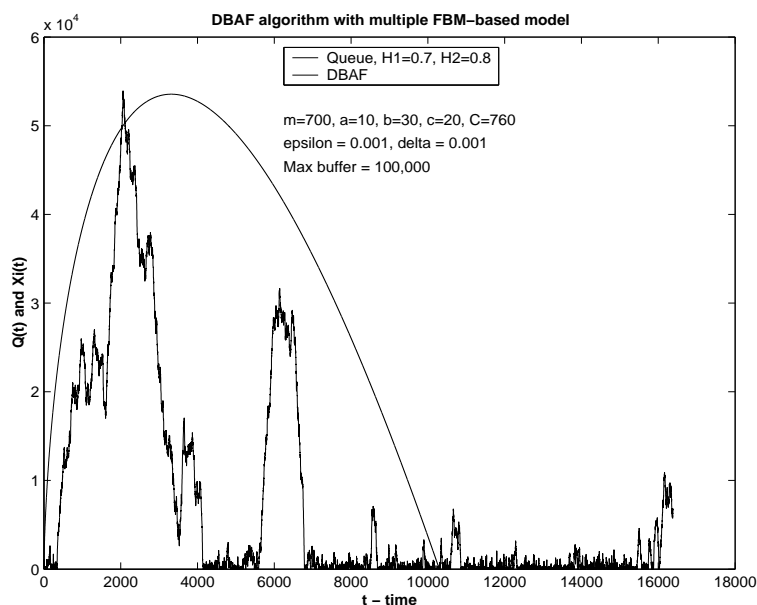

FIGURE 1. Single DBAF Function

3.2. The DBAF-Algorithm Given that the DBAF is a curve which can be used as an upper bound for the queue on a limited time-interval we

11

propose the following algorithm where we partition the time-interval and concatenate copies of the DBAF shifted by the size of the queue. Here is the description in detail of our algorithm.

• Establish the time interval over which the queue is observed. Say this interval is $[0, T]$.

• Choose ϵ and δ such that $0 < \delta \leq \epsilon << 1$.

• Compute

$$\hat{d}_H = \eta^{-1}\left(\frac{\bar{\Phi}^{-1}(\epsilon)}{C - m}\right)$$

where

$$\eta(t) = \frac{t}{\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t}}.$$

• Partition the time interval $[0, T]$ in adjacent intervals of standard length $\hat{d}_H$, namely $[n\hat{d}_H, (n+1)\hat{d}_H)$ for $n = 0, 1, 2, \ldots, N - 1$, where $N = \left[\frac{T}{\hat{d}_H}\right]$, i.e. the integer part of $\frac{T}{\hat{d}_H}$.

• Define the DBAF as follows

$$M(t) := Q(n\hat{d}_H) + \xi(t - n\hat{d}_H), \quad t \in [n\hat{d}_H, (n+1)\hat{d}_H), \, n = 0, 1, 2, \ldots, N-1,$$

where $\xi(t) = \bar{\Phi}^{-1}(\delta)\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t} - (C - m)t$.

Thus, at the beginning of each time interval we reset the clock to 0 and we shift the initial DBAF upwards by the size of the queue at the left end-point of each time subinterval. On each of these subintervals we have

$$\mathbf{P}(Q(t) > M(t)) \leq \delta.$$

We illustrate the algorithm in Figures 2 and 3. In Figure 2 we produce a sample path of traffic generated with the parameters a, b, c, C, m, H_1, and H_2 specified. The value of $\hat{d}_H$ is about 1000 time-units. We apply the algorithm to a trace of about 4000 time-units. Observe that the peak of the DBAF is about 9000. We also provide simulation of more intense traffic in Figure 3 where the mean input rate m is 740 (in Figure 2 it was 700). The resulting DBAF has a peak of about $30,000$. Had we maintained the high precision in probability as in Figure 2 (where $\epsilon = \delta = 0.00001$), this would have made the DBAF much larger. In Figure 3 we chose to relax the precision by taking a $\delta = 0.001$ and $\epsilon = 0.01$. This generates a lower DBAF and some cell-loss as it can be seen. We wish to note that the graphs in this paper are few of many similar ones obtained by the authors for various parameter combinations.
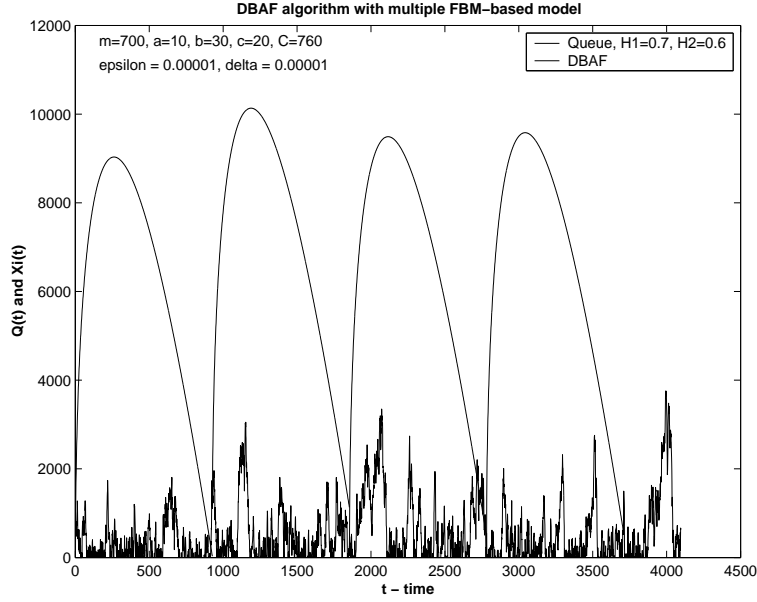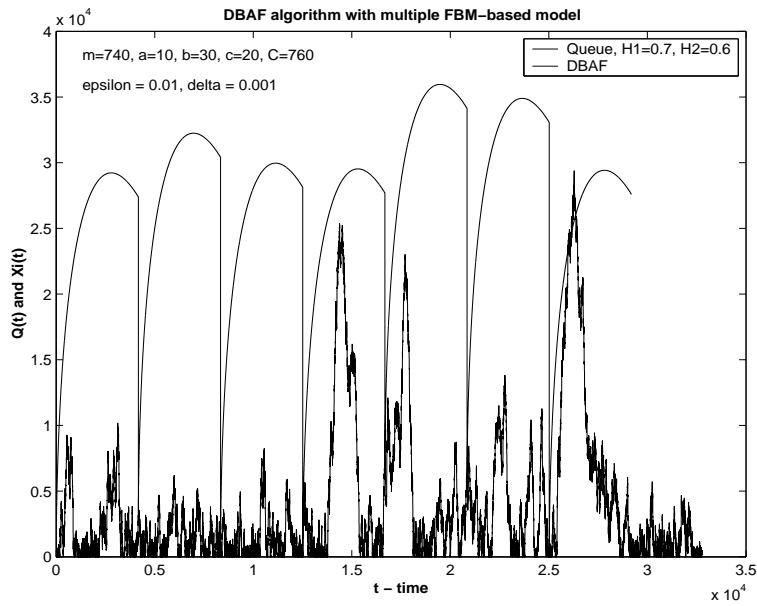
FIGURE 2. Concatenated DBAF Functions

FIGURE 3. Concatenated DBAF Functions

4. An Upper Bound for the Busy Period Using an Envelope Process

4.1. The Envelope Process In [12], the authors introduce a traffic model based on a FBM probabilistic envelope process

$$\hat{A}(t) = mt + k\sqrt{ma}t^H$$

13

and use this process to determine approximations for the overflow probabilities of an ATM queuing system. The input traffic in their case is a FB traffic. It is stated that the same framework can be applied for other arrival processes, as long as a "suitable envelope process" $\{\hat{A}(t), t \geq 0\}$ could be defined. Suitable means that the following condition must hold

$$\mathbf{P}(A(t) > \hat{A}(t)) = \bar{\Phi}(k).$$

In the sequel we solve the problem suggested by the authors of [12]. To this aim we introduce the envelope process associated with a multiple FBM-based traffic as follows.

$$\hat{A}(t) = mt + k\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t}$$

The parameter k determines the probability that $A(t)$ be larger than $\hat{A}(t)$ at time t. More precisely

$$\mathbf{P}(A(t) > \hat{A}(t)) = \mathbf{P}\left(\frac{\sum_{j=1}^{M} \sigma_j B^{H_j}(t) + \tau W(t)}{\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t}} > k\right) = \bar{\Phi}(k).$$

Thus, if we require that $P(A(t) > \hat{A}(t)) \leq \epsilon$, for some $\epsilon > 0$, meaning that we are looking for a big probability that the envelope be an upper bound for the input process, then we get $k \geq \bar{\Phi}^{-1}(\epsilon)$.

In Figure 4 a simple FBM-Based traffic-model $A(t)$ of type (1) is considered. The Hurst parameter is $H = 0.8$. For $\epsilon = 0.0001$ the envelope process is seen to be an upper bound for the traffic. It is observed that if one relaxes the precision in probability the traffic doesn't stay below the envelope at all times.

It is important to have an increasing subadditive envelope process $\hat{A}(t)$ in order to use the following well-known property of such functions

$$\inf_{t \geq 1} \frac{\hat{A}(t)}{t} = \lim_{t \to \infty} \frac{\hat{A}(t)}{t}$$

(see [2]). Therefore, we wish to prove that the envelope we introduced has these properties.

PROPOSITION 4.1. *The envelope function $\hat{A}(t)$ is increasing and subadditive.*

Proof. Monotonicity is obvious. To show that $\hat{A}(t + s) \leq \hat{A}(t) + \hat{A}(s)$, we need to show that

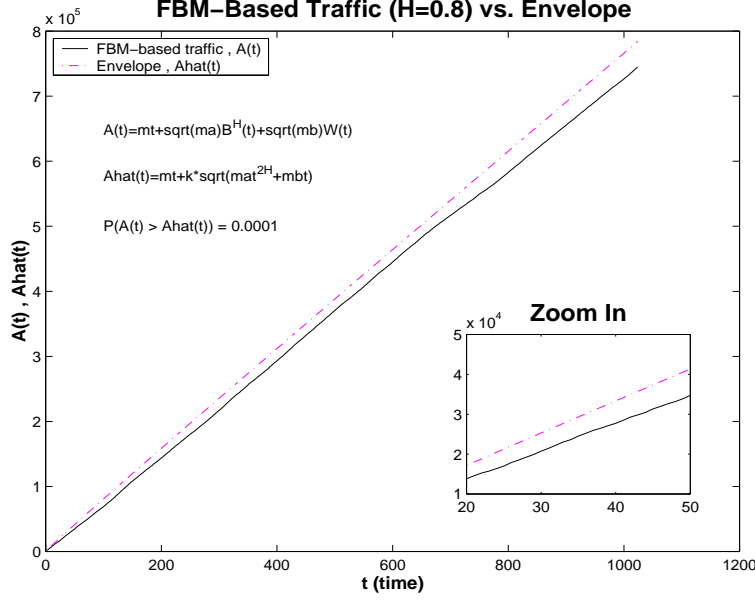$$\sum_{j=1}^{M} \sigma_j^2 (t + s)^{2H_j} \leq$$

14

FIGURE 4. Traffic versus envelope process

$$\leq \sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \sum_{j=1}^{M} \sigma_j^2 s^{2H_j} + 2\sqrt{(\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t)(\sum_{j=1}^{M} \sigma_j^2 s^{2H_j} + \tau^2 s)}.$$

We will show that actually the following inequality holds.

$$\sqrt{\sum_{j=1}^{M} \sigma_j^2 (t+s)^{2H_j}} \leq \sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j}} + \sqrt{\sum_{j=1}^{M} \sigma_j^2 s^{2H_j}}$$

Obviously, this inequality implies the previous one. Consider the space $X = \{1, 2, \ldots, M\}$, the σ-algebra of all parts on the space X, and the weighted counting measure determined by

$$\mu(\{j\}) := \sigma_j^2, \qquad j = 1, 2, \ldots, M.$$

Set $f \in \mathbf{L}^2(X)$, $f(j) := t^{2H_j}$ for $j = 1, 2, \ldots, M$, and similarly set $g \in \mathbf{L}^2(X)$, $g(j) := s^{2H_j}$, for $j = 1, 2, \ldots, M$. The Minkowski inequality

$$||f + g||_2 \leq ||f||_2 + ||g||_2$$

produces

$$\sqrt{\sum_{j=1}^{M} \sigma_j^2 (t+s)^{2H_j}} \leq \sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j}} + \sqrt{\sum_{j=1}^{M} \sigma_j^2 s^{2H_j}}$$

$\square$

As in [2], we set

$$d := \inf\{t \geq 1 : \hat{A}(t) - Ct \leq 0\}.$$

The following proposition exhibits the relation between d and the least dynamic buffer allocation function $\xi(t)$.

15

PROPOSITION 4.2. *If $\hat{A}(1) \leq C$ then $d = 1$. Otherwise, d is the unique positive root of the equation $\zeta(t) = 0$, where*

$$\zeta(t) = k\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t} - (C - m)t.$$

For $k = \bar{\Phi}^{-1}(\epsilon)$, i. e. $\zeta = \xi$ the minimal dynamic buffer allocation function, and if $m + k\sqrt{\sum_{j=1}^{M} \sigma_j^2 + \tau^2} > C$, we have that $d = \hat{d}_H$.

Proof. Clearly, $\{t \geq 1 : \hat{A}(t) - Ct \leq 0\} \neq \emptyset$ since

$$\lim_{t \to \infty} \frac{\hat{A}(t)}{t} = m < C.$$

If $\hat{A}(1) - C \leq 0 \Leftrightarrow m + k\sqrt{\sum_{j=1}^{M} \sigma_j^2 + \tau^2} \leq C$, it follows that $d = 1$. Otherwise $d > 1$.

Obviously, $\hat{A}(d) - Cd \leq 0$. If we assume that $\hat{A}(d) - Cd < 0$, by the continuity of $\hat{A}(t)$ there is a $d' \in (1, d)$ such that $\hat{A}(d') - Cd' \leq 0$, which is a contradiction with the definition of d. Thus, $\hat{A}(d) = Cd$ which means that

$$k\sqrt{\sum_{j=1}^{M} \sigma_j^2 d^{2H_j} + \tau^2 d} - (C - m)d = 0.$$

If we set $\zeta(t) = k\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t} - (C - m)t$, then by the considerations about the function $\xi(t)$ defined in Section 3, it follows that there is a unique $d > 0$ satisfying the previous equality. Observe also that we are in the case where $m + k\sqrt{\sum_{j=1}^{M} \sigma_j^2 + \tau^2} > C$, which implies $d > 1$.

According to [2], under these conditions, any busy period is bounded above by d.

Now, recall that

$$\hat{d}_H = \eta^{-1}\left(\frac{\bar{\Phi}^{-1}(\epsilon)}{C - m}\right)$$

where

$$\eta(t) = \frac{t}{\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t}}.$$

Then

$$\eta(\hat{d}_H) = \frac{\bar{\Phi}^{-1}(\epsilon)}{C - m} \Leftrightarrow \frac{\hat{d}_H}{\sqrt{\sum_{j=1}^{M} \sigma_j^2 \hat{d}_H^{2H_j} + \tau^2 \hat{d}_H}} = \frac{\bar{\Phi}^{-1}(\epsilon)}{C - m}$$

which means that $\hat{d}_H$ is a root of

$$\bar{\Phi}^{-1}(\epsilon)\sqrt{\sum_{j=1}^{M}\sigma_j^2 t^{2H_j} + \tau^2 t} - (C-m)t = 0.$$

If $k = \bar{\Phi}^{-1}(\epsilon)$, by the first part of the proof, this root is exactly d. So $d = \hat{d}_H$, since the root is unique. $\square$

4.2. An Upper Bound for the Queue Size Using the Envelope Process We have seen that the stationary queuing process $\{V(t), t \geq 0\}$ can be approximated by the easier to handle process $\{Q(t), t \geq 0\}$ during the busy period of an ATM queuing system. Recall that

$$V(t) = \sup_{s \leq t}(A(t) - A(s) - C(t-s))$$

and

$$Q(t) = A(t) - Ct.$$

At the same time, if we consider the envelope process defined in the previous subsection by

$$\hat{A}(t) = mt + k\sqrt{\sum_{j=1}^{M}\sigma_j^2 t^{2H_j} + \tau^2 t},$$

then we could require that $\mathbf{P}(A(t) > \hat{A}(t)) = \epsilon$, where $\epsilon = \bar{\Phi}(k)$. But

$$\mathbf{P}(A(t) > \hat{A}(t)) = \mathbf{P}(A(t) - Ct > \hat{A}(t) - Ct) = \mathbf{P}(Q(t) > \hat{Q}(t))$$

where $\hat{Q}(t) := \hat{A}(t) - Ct$. Thus, $\mathbf{P}(Q(t) > \hat{Q}(t)) = \epsilon$, and with probability $1 - \epsilon$ the maximum value of $Q(t)$ is bounded by the maximum value of $\hat{Q}(t)$. We want to find the maximum of

$$\hat{Q}(t) = (m - C)t + k\sqrt{\sum_{j=1}^{M}\sigma_j^2 t^{2H_j} + \tau^2 t}.$$

This function has a global maximum at $t_0 > 0$, so that

$$\hat{Q}_{\max} = \hat{Q}(t_0)$$

and with probability $1 - \epsilon$ this is an upper bound for the queue size $Q(t)$.

Now we would like to determine the service rate $C > 0$ that yields $\hat{Q}_{\max} = x$, where x is a given buffer size. In this case, the buffer will overflow with probability ϵ. This is obtained from

$$\hat{Q}_{\max} = x \Leftrightarrow (m - C)t_0 + k\sqrt{\sum_{j=1}^{M}\sigma_j^2 t_0^{2H_j} + \tau^2 t_0} = x,$$

17

and t_0 is the solution of $\hat{Q}'(t) = 0$, that is the solution of the following equation

$$(5) \qquad (m - C)\sqrt{\sum_{j=1}^{M} \sigma_j^2 t^{2H_j} + \tau^2 t} + \frac{1}{2}k(2\sum_{j=1}^{M} \sigma_j^2 H_j t^{2H_j - 1} + \tau^2) = 0.$$

Thus t_0 and C will satisfy the system formed by the equations (5) and (6).

$$(6) \qquad\qquad (m - C)t_0 + k\sqrt{\sum_{j=1}^{M} \sigma_j^2 t_0^{2H_j} + \tau^2 t_0} = x$$

For a traffic-model with parameters identical to those in Figure 1 we solve the system (5), (6) geometrically first (Figure 5). The buffer-size chosen for this example is 15000. Using a numerical solver one gets the solution $t_0 = 406.21654274$ and $C = 828.73281605$. We used MAPLE to solve the problem geometrically as shown in Figure 5 and MATLAB to solve it numerically. The results obtained are consistent. The traffic-simulations and in general, all figures except 5 were produced using MATLAB.
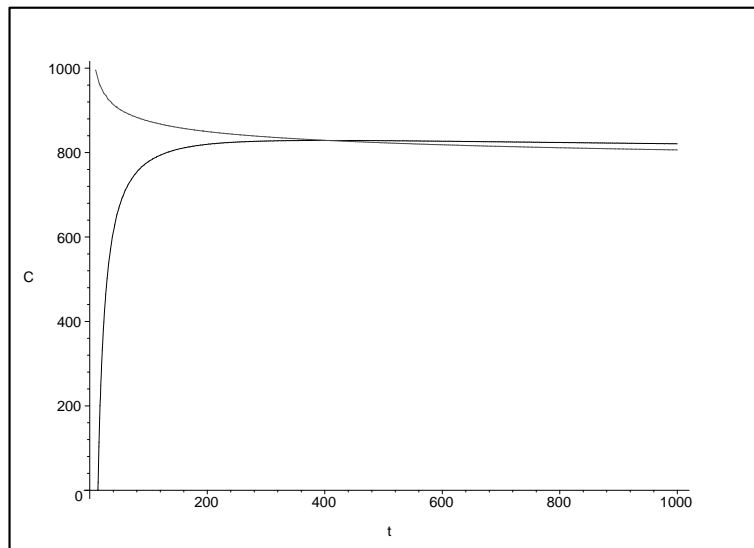

Figure 5. Geometric Solution of (5) and (6)

References

[1] J. Beran, R. Sherman, M. S. Taqqu, W. Willinger, *Long-Range Dependence in Variable-Bit-Rate Video Traffic*, IEEE Transactions on Communications, Vol. 43, No. 2/3/4, February/March/April 1995, 1566-1579.

[2] C. Chang, *Stability, Queue Length, and Delay of Deterministic and Stochastic Queuing Networks*, IEEE Transactions on Automatic Control, May 1994.

[3] M. E. Crovella, A. Bestavros, *Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes*, IEEE/ACM Transactions on Networking, vol. 5, no. 6, pages 835-846, 1997.

[4] A. Dembo, O. Zeitouni, *Theory of Large Deviation Techniques and Applications*, Jones and Bartlett, Boston-London 1993.

[5] N. G. Duffield, N. O'Connell, *Large Deviations and Overflow Probabilities for the General Single-Server Queue, with Applications*, Math. Proc. Cam. Phil. Soc., No. 118, pages 363-374, 1995.

[6] A. Erramilli, O. Narayan, W. Willinger, *Experimental Queuing Analysis with Long-Range Dependent Packet Traffic*, IEEE/ACM Transactions on Networking, vol.4, no. 2, pages 209-223, 1996.

[7] M. W. Garrett, W. Willinger, *Analysis, Modeling and Generation of Self-Similar VBR Video Traffic*, Proceedings of ACM SIGComm '94, London, pages 269-280, 1994.

[8] H. E. Hurst, *Long-Term Storage Capacity of Reservoirs*, Trans. Amer. Soc. Civil Eng., no. 116, pages 770-808, 1951.

[9] H. E. Hurst, *Methods of Using Long-Term Storage in Reservoirs*, Proc. Inst. Civil Eng., no. 1, pages 519-543, 1956.

[10] W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, *On the Self-Similar Nature of Ethernet Traffic*, Proc. ACM Sigcom, San Francisco, pages 183-193, 1993.

[11] W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, *On the Self-Similar Nature of Ethernet Traffic (Extended Version)*, IEEE/ACM Transactions on Networking, vol. 2, no. 1, pages 1-15, 1994.

[12] G. Mayor, J. Silvester, *Time Scale Analysis of an ATM Queuing System With Long-Range Dependent Traffic*, IEEE INFOCOM, pages 205-212, 1997.

[13] K. Meier-Hellstern, P. E. Wirth, Y.-L. Yan, D. A. Hoeflin, *Traffic Models for ISDN Data Users: Office Automation Application*, Proc. 13th ITC, Coopenhagen, Denmark, pages 167-172, 1991.

[14] I. Norros, *A storage model with self-similar input*, Queuing Systems , vol. 16, pages 387-396, 1994.

[15] I. Norros, *The Management of Large Flows of Connectionless Traffic on the Basis of Self-Similar Modeling*, Proc. of IEEE ICC, 1995.

[16] V. Paxson, S. Floyd, *Wide-Area Traffic: The Failure of Poisson Modeling*, Proc. ACM Sigcom, London, UK, pages 257-268, 1994.

[17] J. Roberts, U. Mocci, J. Virtamo (Eds.), *Broadband Network Traffic*, Springer-Verlag, 1996.

[18] M. Schwartz, *Broadband Integrated Networks*, Prentice Hall, 1996.

[19] W. Willinger, M.S. Taqqu, W.E. Leland, D.V. Wilson, *Self-Similarity in High-Speed Packet Traffic: Analysis and Modelling of Ethernet Traffic Measurements*, Statistical Science, vol. 10, no. 1, pages 67-85, 1995.

[20] W. Willinger, M. S. Taqqu, R. Sherman, V. Wilson, *Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level*, IEEE/ACM Transactions on Networking, vol. 5, no. 6, pages 71-86, 1997.