

QUEUEING SYSTEMS FOR MULTIPLE FBM-BASED TRAFFIC MODELS

MIHAELA T. MATACHE¹ and VALENTIN MATACHE¹

(Received 24 July, 2003; revised 19 July, 2004)

Abstract

A multiple fractional Brownian motion (FBM)-based traffic model is considered. Various lower bounds for the overflow probability of the associated queueing system are obtained. Based on a probabilistic bound for the busy period of an ATM queueing system associated with a multiple FBM-based input traffic, a minimal dynamic buffer allocation function (DBAF) is obtained and a DBAF-allocation algorithm is designed. The purpose is to create an upper bound for the queueing system associated with the traffic. This upper bound, called a DBAF, is a function of time, dynamically bouncing with the traffic. An envelope process associated with the multiple FBM-based traffic model is introduced and used to estimate the queue size of the queueing system associated with that traffic model.

1. Introduction

Over the last years many studies [1, 3, 7–9, 11, 13, 16, 17] have shown that packet/cell traffic through telecommunication networks (like Ethernet, LAN, WAN, ISDN, ATM) exhibits long-range dependence and self-similarity, that is, 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 modelling, like Poisson models, cannot capture observed features of the telecommunication traffic. New parsimonious models are proposed in [5, 9, 12]. They involve the fractional Brownian motion (FBM) process whose properties make it a natural choice in modelling packet/cell traffic.

At the same time experimental and analytical studies [5, 6, 10, 12] show that the long-range dependence property can create big packet/cell losses in queueing systems,

¹Department of Mathematics, The University of Nebraska at Omaha, Omaha, NE 68182, USA; e-mail: dmatache@mail.unomaha.edu and vmatache@mail.unomaha.edu.

© Australian Mathematical Society 2005, Serial-fee code 1446-1811/05

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 queueing solution for fractional Brownian (FB) traffic models is known. However, in [5, 10, 12], approximations and bounds for the overflow probabilities in a queueing system driven by 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 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 [14].

For a classical FB traffic model, the following authors: Norros, Duffield, O'Connell, Mayor and Silvester [5, 10, 12], have analysed the associated queueing system and obtained results related to the estimation of the busy period and the overflow probabilities.

In what follows we extend these results to a multiple FBM-based traffic model (see (2.2) in Section 2 for its exact definition). In Section 2, besides introducing the basic ideas and setting up the notation, we obtain a lower bound for the overflow probability of the associated queueing system (Subsection 2.1). Also, asymptotic lower bounds for the same probability are obtained (Theorem 2.2). In Section 3, based on a probabilistic upper bound for the busy period of an ATM queueing system with a multiple FBM-based input traffic model we introduce the notion of a dynamic buffer allocation function (DBAF) and show that a least DBAF exists (Proposition 3.2). 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 1–3). In Section 4 we introduce an envelope process associated with the multiple FBM-based traffic model and use it to obtain an upper bound for the busy period of an ATM queueing 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 FB traffic is not sufficient to model 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:

$$A(t) = mt + \sqrt{ma}B^H(t) + \sqrt{mb}W(t). \quad (2.1)$$

Here m, a, b are positive constants, $\{B^H(t), t \geq 0\}$ is a standard FBM with Hurst parameter $H \in (1/2, 1)$, and $\{W(t), t \geq 0\}$ is a standard BM, independent of FBM (see [14]). In this paper we will study a more general type of model with several FBMs given by the following equality:

$$A(t) = mt + \sum_{j=1}^M \sigma_j B^{H_j}(t) + \tau W(t). \tag{2.2}$$

As above, m, σ_j , and τ are positive constants, $\{B^{H_j}(t), t \geq 0\}, j = 1, \dots, M$ are independent standard FBMs with Hurst parameters $H_j, 1/2 < H_j < 1$, for all $j = 1, \dots, M$ 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 (2.1). Observe that the class of the stochastic processes of type (2.2) is closed under superposition.

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

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

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

$$A(t) = mt + \sum_{j=1}^M \sigma_j B^{H_j}(t) + \tau W(t), \tag{2.3}$$

where

$$m = \sum_{i=1}^N m_i, \quad \sigma_j = \sqrt{\sum_{i=1}^N \sigma_{ij}^2}, \quad \tau = \sqrt{\sum_{i=1}^N \tau_i^2},$$

$$\{B^{H_j}(t), t \geq 0\}, \quad j = 1, \dots, M$$

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 queueing model can be described by the process $\{V(t), t \geq 0\}$, where

$$V(t) = \sup_{s \leq t} (A(t) - A(s) - C(t - s)), \quad t \geq 0. \tag{2.4}$$

In (2.4), C represents constant service rate and satisfies $C > m$, and $m > 0$ is the mean input rate given in (2.2). Formula (2.4) gives the workload or the virtual waiting time in a FIFO (first in–first out) queueing 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: $\epsilon := \mathbf{P}(V(t) > x), t \geq 0$. Here x denotes a given buffer size. In what follows we will use the notation

$$\bar{\Phi}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy,$$

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 $\{A(t), t \geq 0\}$ in (2.3) be a multiple FBM-based traffic model and $\{V(t), t \geq 0\}$ be the stationary queueing process defined in (2.4). Then*

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

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:

$$(C - m) \sum_{j=1}^M \sigma_j^2 (1 - H_j) u^{2H_j} + \frac{1}{2} (C - m) \tau^2 u - \frac{1}{2} \tau^2 x - x \sum_{j=1}^M \sigma_j^2 H_j u^{2H_j - 1} = 0. \tag{2.5}$$

PROOF. Observe that

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

Using this equality we deduce that

$$\begin{aligned} & \mathbf{P}\left(\sup_{s \leq t} (A(t) - A(s) - C(t-s)) > x\right) \\ & \geq \sup_{s \leq t} \mathbf{P}(A(t) - A(s) - C(t-s) > x) \\ & = \sup_{s \leq t} \mathbf{P}(A(t) - A(s) - C(t-s) - (m-C)(t-s) > x - (m-C)(t-s)) \\ & = \sup_{s \leq t} \int_{x-(m-C)(t-s)}^{\infty} \frac{1}{\sqrt{2\pi}\sigma(s)} e^{-y^2/2\sigma^2(s)} dy, \end{aligned}$$

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

$$\begin{aligned} \mathbf{P}\left(\sup_{s \leq t} (A(t) - A(s) - C(t-s)) > x\right) & \geq \sup_{s \leq t} \bar{\Phi}\left(\frac{x - (m-C)(t-s)}{\sigma(s)}\right) \\ & = \bar{\Phi}\left(\inf_{s \leq t} \frac{x - (m-C)(t-s)}{\sigma(s)}\right). \end{aligned}$$

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

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

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

$$\inf_{0 \leq u \leq t} \psi(u) = \begin{cases} \psi(t), & \text{if } t < u_1, \\ \psi(u_1), & \text{if } t \geq u_1. \end{cases}$$

In conclusion,

$$\mathbf{P}\left(\sup_{s \leq t} (A(t) - A(s) - C(t-s)) > x\right) \geq \begin{cases} \bar{\Phi}(\psi(t)) \geq \bar{\Phi}(\psi(u_1)), & \text{if } t < u_1, \\ \bar{\Phi}(\psi(u_1)), & \text{if } t \geq u_1. \end{cases}$$

Clearly, u_1 is the root of $\zeta'(u)$ which, by a short computation, is the same as the unique real root of (2.5).

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 (2.2), and $\{V(t), t \geq 0\}$ the workload process defined in (2.4). Then*

(a)

$$\liminf_{x \rightarrow \infty} \frac{\sum_{j=1}^M \sigma_j^2 x^{2H_j} + \tau^2 x}{x^2} \log \mathbf{P}(V(t) > x) \geq -\frac{1}{2} \frac{(C - m)^{2H}}{H^{2H} (1 - H)^{2-2H}}, \tag{2.6}$$

where $H = \max\{H_j, j = 1, \dots, M\}$.

(b)

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

PROOF. (a) It is known from [5] that

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

if the following hypotheses are satisfied.

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

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

exists in $[-\infty, \infty]$. Moreover, $\lambda(\theta)$ is essentially smooth and lower semicontinuous. 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 $h : [0, \infty) \rightarrow [0, \infty)$ such that the limit

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

exists for each $u > 0$, where $a^{-1}(t) := \sup_{s \geq 0} \{a(s) \leq t\}$. Here λ^* represents the Fenchel-Legendre transform of λ , that is, the function defined as

$$\lambda^*(x) := \sup_{\theta \geq 0} \{\theta x - \lambda(\theta)\}.$$

The three hypotheses are satisfied using the following functions:

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

Under these conditions the conclusion in [5] holds and

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

Observe that

$$\lambda^*(x) := \sup_{\theta \geq 0} \{\theta x - \lambda(\theta)\} = \sup_{\theta \geq 0} \left\{ -\frac{1}{2} \theta^2 + \theta(x + C - m) \right\} = \frac{(x + C - m)^2}{2}.$$

Thus

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

In conclusion, we have that (2.6) holds.

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

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

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

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

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

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

3. Dynamic buffer allocation

3.1. A probabilistic bound The maximum busy period of an ATM queueing system is very important since it provides a bound for the delay of the ATM cells in the queue [15].

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 [10]. Here $Q(t) = A(t) - Ct$, $\{A(t), t \geq 0\}$ is the input traffic process (2.2) and C the positive service rate.

PROPOSITION 3.1. *The bound \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

$$\eta(t) = \frac{t}{\sqrt{\sum_{j=1}^M \sigma_j^2 t^{2H_j} + \tau^2 t}}. \tag{3.1}$$

PROOF. We have

$$\begin{aligned} \mathbf{P}(Q(t) > 0) &= \mathbf{P} \left(\frac{\sum_{j=1}^M \sigma_j B^{H_j}(t) + \tau W(t)}{\sqrt{\sum_{j=1}^M \sigma_j^2 t^{2H_j} + \tau^2 t}} > \frac{(C - m)t}{\sqrt{\sum_{j=1}^M \sigma_j^2 t^{2H_j} + \tau^2 t}} \right) \\ &= \bar{\Phi} \left(\frac{(C - m)t}{\sqrt{\sum_{j=1}^M \sigma_j^2 t^{2H_j} + \tau^2 t}} \right). \end{aligned}$$

So

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

If we define $\eta(t)$ as in (3.1), it can be easily seen that $\eta(t)$ is invertible so that we can write

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

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

For fixed $\delta > 0$, we want to determine a positive function $M(t)$, which will be called a DBAF, such that $\mathbf{P}(Q(t) > M(t)) \leq \delta$. 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 queueing 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 DBAF and it is given by*

$$\xi(t) = \bar{\Phi}^{-1}(\delta) \sqrt{\sum_{j=1}^M \sigma_j^2 t^{2H_j} + \tau^2 t - (C - m)t}. \tag{3.2}$$

PROOF. Observe that

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

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

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

which is equivalent to

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

Then if we define $\xi(t)$ as in (3.2), this function is the least DBAF, since $M(t) \geq \xi(t)$.

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

REMARK 2. The inequality $\hat{d}_H \leq t_0$ holds if and only if $\delta \leq \epsilon$.

PROOF. The inequality $\hat{d}_H \leq t_0$ is equivalent to

$$\eta^{-1} \left(\frac{\bar{\Phi}^{-1}(\epsilon)}{C - m} \right) \leq t_0,$$

where $\eta(t)$ is given in (3.1). 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 $\bar{\Phi}^{-1}(\epsilon) \leq \bar{\Phi}^{-1}(\delta)$ and therefore $\delta \leq \epsilon$.

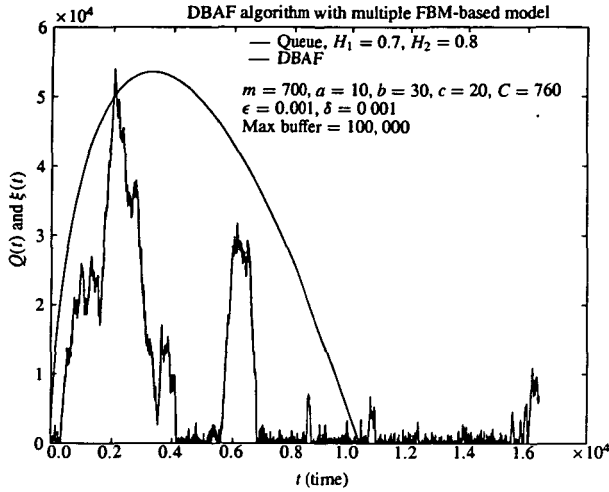


FIGURE 1. Single DBAF

In Figure 1 a multiple FBM-based traffic queue is simulated using MATLAB. The corresponding DBAF is graphed to illustrate how it bounds the queue from above. The traffic utilised was of the form

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

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 utilised 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 make the observation that changing parameters results 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 with 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.

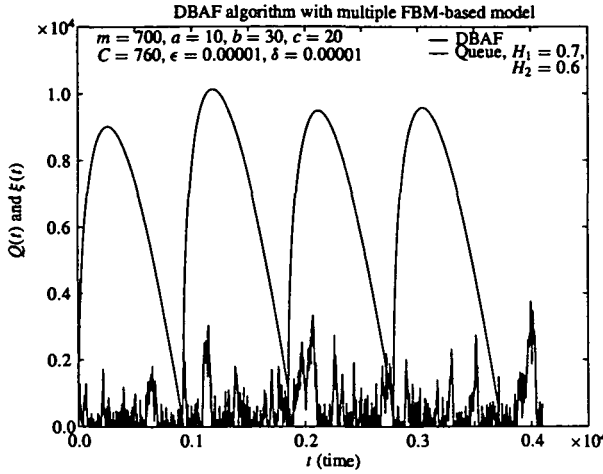


FIGURE 2. Concatenated DBAFs

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 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 $0 < \delta \leq \epsilon \ll 1$.
- Compute $\hat{d}_H = \eta^{-1}(\bar{\Phi}^{-1}(\epsilon)/(C - m))$, where $\eta(t)$ is given by (3.1).
- 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 $n = 0, 1, \dots, N - 1$, where $N = \lceil T/\hat{d}_H \rceil$, that is, the integer part of T/\hat{d}_H .
- Define the DBAF as follows:

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

where $\xi(t)$ is given by (3.2).

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

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

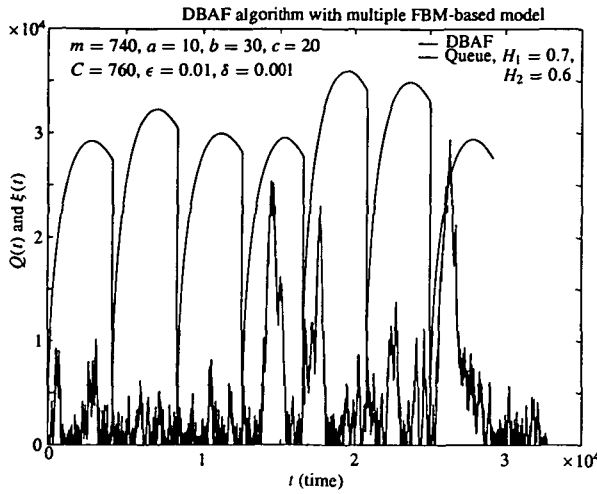


FIGURE 3. Concatenated DBAFs

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 $\delta = 0.001$ and $\epsilon = 0.01$. This generates a lower DBAF and some cell loss as can be seen. We wish to note that the graphs in this paper are but a few of many similar ones obtained by the authors for various parameter combinations.

4. An upper bound for the busy period using an envelope process

4.1. The envelope process In [10], the authors introduce a traffic model based on an FBM probabilistic envelope process

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

and use this process to determine approximations for the overflow probabilities of an ATM queueing system. The input traffic in their case is FB traffic. It is stated that the same framework can be applied to other arrival processes, as long as a “suitable envelope process” $\{\hat{A}(t), t \geq 0\}$ can be defined. “Suitable” means that $\mathbf{P}(A(t) > \hat{A}(t)) = \bar{\Phi}(k)$ must hold.

In what follows we solve the problem suggested by the authors of [10]. To this aim we introduce the envelope process associated with a multiple FBM-based traffic as follows:

$$\hat{A}(t) = mt + k \sqrt{\sum_{j=1}^M \sigma_j^2 t^{2H_j} + \tau^2 t}. \tag{4.1}$$

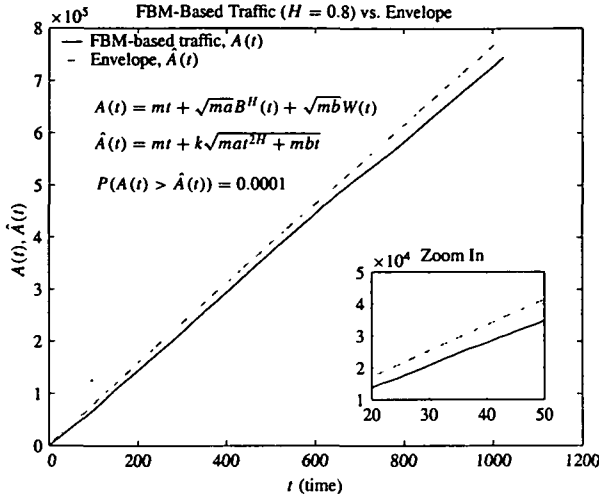


FIGURE 4. Traffic versus envelope process

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

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

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 (2.1) 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:

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

(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

$$\sum_{j=1}^M \sigma_j^2 (t + s)^{2H_j} \leq \sum_{j=1}^M \sigma_j^2 t^{2H_j} + \sum_{j=1}^M \sigma_j^2 s^{2H_j} + 2 \sqrt{\left(\sum_{j=1}^M \sigma_j^2 t^{2H_j} + \tau^2 t\right) \left(\sum_{j=1}^M \sigma_j^2 s^{2H_j} + \tau^2 s\right)}. \tag{4.2}$$

We will show that actually the following inequality holds:

$$\sqrt{\sum_{j=1}^M \sigma_j^2 (t + s)^{2H_j}} \leq \sqrt{\sum_{j=1}^M \sigma_j^2 t^{2H_j}} + \sqrt{\sum_{j=1}^M \sigma_j^2 s^{2H_j}}. \tag{4.3}$$

Obviously, (4.3) implies (4.2). Consider the space $X = \{1, 2, \dots, M\}$, the σ -algebra of all parts on the space X , and the weighted counting measure determined by $\mu(\{j\}) := \sigma_j^2, j = 1, \dots, M$. Set $f \in L^2(X), f(j) := t^{2H_j}$ for $j = 1, 2, \dots, M$, and similarly set $g \in L^2(X), g(j) := s^{2H_j}$, for $j = 1, 2, \dots, M$. The Minkowski inequality $\|f + g\|_2 \leq \|f\|_2 + \|g\|_2$ produces (4.3).

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 $\xi(t)$.

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

$$\zeta(t) = k \sqrt{\sum_{j=1}^M \sigma_j^2 t^{2H_j} + \tau^2 t} - (C - m)t. \tag{4.4}$$

For $k = \bar{\Phi}^{-1}(\epsilon)$, that is, $\zeta = \xi$ the minimal DBAF, 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 $\lim_{t \rightarrow \infty} (\hat{A}(t)/t) = m < C$.

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 contradicts the definition of d . Thus $\hat{A}(d) = Cd$, which means that

$$k \sqrt{\sum_{j=1}^M \sigma_j^2 d^{2H_j} + \tau^2 d} - (C - m)d = 0. \tag{4.5}$$

If we set $\zeta(t)$ as in (4.4), then by the considerations about the function $\xi(t)$ defined in Section 3, it follows that there is a unique $d > 0$ satisfying (4.5). 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 $\eta(t)$ is given by (3.1). Then

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

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.

4.2. An upper bound for the queue size using the envelope process We have seen that the stationary queueing 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 queueing 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 (4.1), 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

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

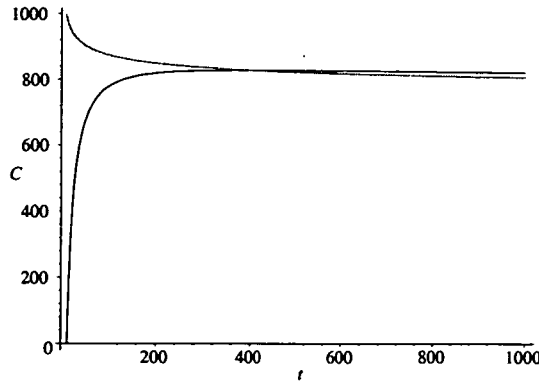


FIGURE 5. Geometric solution of (4.6)–(4.7).

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

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

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

$$(m - C) \sqrt{\sum_{j=1}^M \sigma_j^2 t^{2H_j} + \tau^2 t} + \frac{1}{2}k \left(2 \sum_{j=1}^M \sigma_j^2 H_j t^{2H_j-1} + \tau^2 \right) = 0. \tag{4.6}$$

Thus t_0 and C will satisfy the system formed by the equations (4.6) and (4.7):

$$(m - C)t_0 + k \sqrt{\sum_{j=1}^M \sigma_j^2 t_0^{2H_j} + \tau^2 t_0} = x. \tag{4.7}$$

For a traffic model with parameters identical to those in Figure 1 we solve the system (4.6)–(4.7) 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 Figure 5 were produced using MATLAB.

References

- [1] J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, "Long-range dependence in variable-bit-rate video traffic", *IEEE Trans. Communications* **43** (1995) 1566–1579.
- [2] C. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks", *IEEE Trans. Aut. Control* **39** (1994) 913–931.
- [3] M. E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: Evidence and possible causes", *IEEE/ACM Trans. Networking* **5** (1997) 835–846.
- [4] A. Dembo and O. Zeitouni, *Theory of large deviation techniques and applications* (Jones and Bartlett, Boston, 1993).
- [5] N. G. Duffield and N. O'Connell, "Large deviations and overflow probabilities for the general single-server queue, with applications", *Math. Proc. Cambridge Philos. Soc.* **118** (1995) 363–374.
- [6] A. Erramilli, O. Narayan and W. Willinger, "Experimental queueing analysis with long-range dependent packet traffic", *IEEE/ACM Trans. Networking* **4** (1996) 209–223.
- [7] M. W. Garrett and W. Willinger, "Analysis, modeling and generation of self-similar VBR video traffic", in *Proceedings of ACM Sigcom'94, London*, (ACM Press, New York, 1994), 269–280.
- [8] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the self-similar nature of Ethernet traffic", in *Proceedings of ACM Sigcom'93, San Francisco*, (ACM Press, New York, 1993), 183–193.
- [9] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the self-similar nature of Ethernet traffic (Extended version)", *IEEE/ACM Trans. Networking* **2** (1994) 1–15.
- [10] G. Mayor and J. Silvester, "Time scale analysis of an ATM queueing system with long-range dependent traffic", in *Proceedings of INFOCOM'97*, (IEEE, Washington, 1997) 205–212.
- [11] K. Meier-Hellstern, P. E. Wirth, Y.-L. Yan and D. A. Hoeflin, "Traffic models for ISDN data users: office automation application", in *Proceedings of 13th ITC, Copenhagen, Denmark*, (Elsevier Science Publ., Amsterdam, 1991), 167–172.
- [12] I. Norros, "A storage model with self-similar input", *Queueing Systems Theory Appl.* **16** (1994) 387–396.
- [13] V. Paxson and S. Floyd, "Wide-area traffic: the failure of Poisson modeling", in *Proceedings of ACM Sigcom'94, London*, (ACM Press, New York, 1994), 257–268.
- [14] J. Roberts, U. Mocchi and J. Virtamo (eds.), *Broadband network traffic* (Springer, Berlin, 1996).
- [15] M. Schwartz, *Broadband integrated networks* (Prentice-Hall, Englewood-Cliffs, NJ, 1996).
- [16] W. Willinger, M. S. Taqqu, W. E. Leland and D. V. Wilson, "Self-similarity in high-speed packet traffic: Analysis and modelling of Ethernet traffic measurements", *Statist. Sci.* **10** (1995) 67–85.
- [17] W. Willinger, M. S. Taqqu, R. Sherman and V. Wilson, "Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level", *IEEE/ACM Trans. Networking* **5** (1997) 71–86.