Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computer Networks 48 (2005) 351–375
Computer Networks
www.elsevier.com/locate/comnet

Envelope process and computation of the equivalent bandwidth of multifractal flows

Cesar A.V. Melo, Nelson L.S. da Fonseca *

Institute of Computing, State University of Campinas, P.O. Box 6176, 13084-971 Campinas, SP, Brazil

Available online 13 January 2005

Abstract

Internet Protocol flows present high variability at small time scales as well as long range dependence, which can be captured by multifractal models. Estimating the bandwidth to support the Quality of Service required by these flows is the key to Traffic Engineering. This paper introduces a novel envelope process which is a minimalist yet accurate model for multifractal flows. The envelope process is an upper bound to the volume of arrivals from a multifractal Brownian motion. The envelope process accuracy was assessed using both real network traces and synthetically generated traces.

Moreover, the solution of a queue fed by multifractal flows is presented and an expression for the time at which the queue length reaches its maximum is derived. This time instant is used for the derivation of an efficient method for the computation of the equivalent bandwidth of multifractal flows. Furthermore, a policing mechanisms to assure the conformance of a flow to the multifractal envelope process is presented. It is also shown that a monofractal approach for modeling multifractal flows leads to overestimation of the bandwidth needed.

Keywords: Multifractal flows; Traffic engineering; Long-range dependence

1. Introduction

Since the seminal work of Leland et al. [1], several studies have shown that network traffic presents scale invariance, or ''scaling'', which is the absence of any specific time scale at which the ''burstiness'' of a traffic stream can be characterized. Instead, it is necessary to describe the traffic across different time scales. Since that time, self-similar or (mono) fractal processes have been used for modeling such network traffic.

* Corresponding author.
E-mail address: nfonseca@ic.unicamp.br (N.L.S. da Fonseca).

1389-1286/$ - see front matter
doi:10.1016/j.comnet.2004.11.010

Scaling of monofractal traffic is measured by a single constant value: the Hurst parameter, H. One of the most popular monofractal processes for traffic modeling is the fractal Brownian motion process (fBm), due to its parsimonious representation of the modeled traffic. The multifractal generalization of fractal Brownian motion is the multifractal Brownian motion (mBm), which is a Gaussian process capable of capturing both the high variability existing on small time scales and long-term correlations. Moreover, for small time scales its increment is, locally, an fBm realization.

In addition to long-term memory, Internet Protocol (IP) traffic presents a non-trivial scaling structure at small scales [2]. At these scales, traffic is highly variable and more complex; moreover, it follows less definitive scaling laws. For such traffic, the marginal distribution of counts is clearly non-Gaussian, calling for a representation beyond second-order statistics.

The concept of equivalent bandwidth is intimately connected with both network dimensioning and Quality of Service [3]. The equivalent bandwidth of a flow can be defined as the minimum required bandwidth such that QoS requirements of this flow are met. There has been a great interest in this concept, since it promises to bridge the gap between the design of statistical multiplexing networks and the familiar design of circuit-switched networks. Numerous studies investigate equivalent bandwidth [3–6]; some of them have involved traffic with long range dependence [7–9].

The computation of the equivalent bandwidth of a multifractal flow requires the solution of the queuing system fed by this flow. Solving queueing systems with (multi/mono) fractal input, however, is not a trivial task. Although the Large Deviation theory can be employed to overcome these difficulties [10,11], it generally implies a non-realistic assumption about buffer sizes. One way to avoid such an assumption is to use envelope processes, which set upper bounds to the accumulated amount of work arriving up to a certain time. These envelope processes are parsimonious representations of stochastic processes and allow simple solutions for queueing systems fed by (mono/multi) fractal processes yet do not incorporate assumptions about buffer size.

The present paper introduces a novel envelope process for modeling multiscaling traffic. The envelope process is an upper bound for the accumulated amount of traffic arriving up to a certain time from a multifractal Brownian motion process (mBm) [12]. The derivation of the present envelope process is based on the fact that an mBm can be locally modeled as an fBm. The envelope process has been validated using both synthetic and real network traffic. The traces of IP traffic used in the validation process were publicly available and were collected from various sites in academic and industrial environments during the years from 1995 to 2003, including periods before and after the expansion of the World-Wide-Web (WEB). The results show that although mBm is a steady state Gaussian process, the envelope process involves a tight bound for the amount of traffic arriving from real network streams, regardless of the time and place where the traces are collected.

This paper also proposes a method for the computation of the instant in time at which a queue fed by several multifractal flows reaches its maximum. Such results are used to calculate the probability of loss, as well as to determine the equivalent bandwidth of an aggregate of several multifractal flows. Moreover, an efficient algorithm for computing the equivalent bandwidth is presented.

An envelope process for monofractal traffic was introduced in [13] and was extensively validated in [9]. That process is a special case of the mBm envelope process introduced here. Modeling multifractal flows using monofractal models, however, has been found to lead to considerable overestimation of the bandwidth needed by these flows. While the time scale of interest derived in [9] considered a single busy period, that presented in this paper takes into account the evolution of the queue length and, consequently, considers various busy periods.

This paper is organized as follows: Section 2 reviews the definition of the multifractal Brownian motion process. Section 3 introduces an envelope process based on mBm and presents results from a validation procedure designed to evaluate the accuracy of this process. Section 4 presents the computation of the time scale at which a queue fed by a multifractal flow reaches its maximum. Section 5 introduces the computa-

tion of the equivalent bandwidth of an aggregate of multifractal flows. Section 6 extends the Fractal Leaky Bucket policing mechanism to monitor multifractal flows. Section 7 discusses related work, and Section 8 presents the conclusions.

2. Multifractal Brownian motion

Multifractal processes exhibit highly irregular patterns as a function of time. The local regularity of a sample path of such a process can be described using a local Holder exponent, which is a generalization of the Hurst parameter [14]. This exponent provides a measure of scaling and depends on both time and sample paths.

The Holder exponent is the largest value of $H(\cdot)$, $0 \leqslant H(\cdot) \leqslant 1$, such that

$$|X(t+\gamma) - X(t)| \leqslant k|\gamma|^{H(t)} \quad \text{for } \gamma \to 0. \tag{1}$$

For monofractal processes, the Holder exponent (Holder function) is a constant value (Hurst parameter H), whereas for multifractal processes it changes randomly with time. Let $H(\cdot) : (0,\infty) \to (0,1)$ be a Holder function. Multifractal Brownian motion is a continuous Gaussian process with non-stationary increments defined on $(0,\infty)$ as follows:

$$W_{H(t)} = \frac{1}{\Gamma(H(t)+1/2)} \left\{ \int_{-\infty}^{0} [(t-s)^{H(t)-1/2} - (-s)^{H(t)-1/2}]\, \mathrm{d}B(s) + \int_{0}^{t} (t-s)^{H(t)-1/2}\, \mathrm{d}B(s) \right\},$$

where $B(s)$ is the Brownian motion.

The multifractal Brownian motion process is a generalization of the fractal Brownian motion process, yet it exhibits local asymptotic self-similarity (lass), i.e.

$$\lim_{\rho \to 0^+} \left\{ \frac{W(t+\rho u) - W(t)}{\rho^{H(t)}} \right\}_{u \in R^+} = \{B_{H(t)}(u)\}_{u \in R^+}, \tag{2}$$

where $W(\cdot)$ is an mBm and $B_{H(t)}(u)$ is a realization of an fBm process with Hurst parameter H, given by $H(t)$.

The value of the Holder function is crucial for the characterization of multifractal traffic. An accurate estimator, introduced in [15], is used here.

3. Envelope process for multifractal traffic

The solution for a queueing system fed by an input process requires the knowledge of the amount of work arriving in that system. The envelope processes furnishing upper bounds for these quantities can be either deterministic or probabilistic. In deterministic envelopes, the amount of work arriving will never surpass the envelope value, whereas in probabilistic envelopes, this limit may be surpassed with a certain pre-defined probability. Probabilistic envelope processes provide tighter bounds, since they can take into consideration statistically probable fluctuations instead of the worst possible case for the process being modeled. Dimensioning based on deterministic envelope processes may lead to bandwidth waste, since what is provided must consider the maximum possible amount of work which might arrive at any time. Occasional spikes of arrivals can be ignored in probabilistic envelope processes. Even though a certain packet loss may occur when peak loads have not been taken into consideration, the probability of this loss is at most equal to the probability of violation, a parameter of the envelope process.

3.1. Definition of the envelope process

This section introduces a novel envelope process for the modeling of multifractal flows and provides experimental evidence of its high precision.

An upper bound for the accumulated amount of work arriving can be calculated as the mean amount of work plus an upper bound for accumulated increments. For calculations, this upper bound for mBm increments is equivalent to an upper bound for fBm increments, since in the neighborhood of any time instant t, an mBm can be approximated by an fBm with Hurst parameter H given by the value of the Holder function $H(\cdot)$ at the time instant t. For fBm increments, the upper bound can be computed as in [9]:

$$Z_H(t) \leqslant \kappa H t^{H-1}, \tag{3}$$

where $Z_H(t)$ is an fBm increment at time instant t.

Thus, an upper bound for the accumulate work arriving from an mBm process with a mean of $\bar{a}$, standard deviation σ and Holder function $H(\cdot)$ can be expressed as:

$$\hat{A}(t) = \int_0^t \bar{a} + \kappa\sigma H(x) x^{H(x)-1}\,\mathrm{d}x, \tag{4}$$

which is called the mBm envelope process.

This envelope reduces to the fBm envelope previously derived in [9] when $H(\cdot)$ is a constant value, i.e.,

$$\hat{A}(t) = at + \kappa\sigma t^H. \tag{5}$$

3.2. Assessment of envelope process accuracy

Simulation experiments using both synthetic traffic and real network traffic were conducted in order to assess the accuracy of the proposed envelope. Publicly available traces containing real network traffic were used in the validation process. These traces were collected in a period spanning eight years, from 1995 to 2003, a time span which included traffic prior to the expansion of the WEB. The traces collected in 2003 were obtained from the NLANR site (http://www.nlanr.net) and were obtained at aggregation points in high performance connection networks, such as vBNS and Internet2 ABILENE. Those gathered in 1999 and 2000 were collected at an OC-3 link between Auckland University and its Internet Service Provider [16]. The traces collected in 1995, the dec-pkt collection, were gathered on an Ethernet network which was the primary Internet Access point of the Palo Alto Digital research groups [17]. Table 1 shows the characteristics of these traces, all involving sampling precision in the order of microseconds.

Table 1
Real network traces used in trace-driven simulation

Trace	Date	Packets	Aggregation point
dec-pkt-1	03/08/1995 22:00	3,300,000	Digital IAP
dec-pkt-4	03/09/1995 14:00	5,700,000	Digital IAP
19991129-134258-1	11/29/1999 12:42	58,000,000	University of Auckland
20000125-143640-1	01/25/2000 14:36	7,000,000	University of Auckland
AIX-1049492523	04/04/2003 13:49	9376	NASA Ames to MAE-West
MEM-1053844177	05/24/2003 23:54	220,904	University of Memphis
MEM-1054459191	06/01/2003 02:54	266,708	University of Memphis
MRA-1057960474	07/11/2003 22:24	4,137,819	Merit Abilene
COS-1057970154	07/12/2003 00:49	1,247,518	Colorado State University

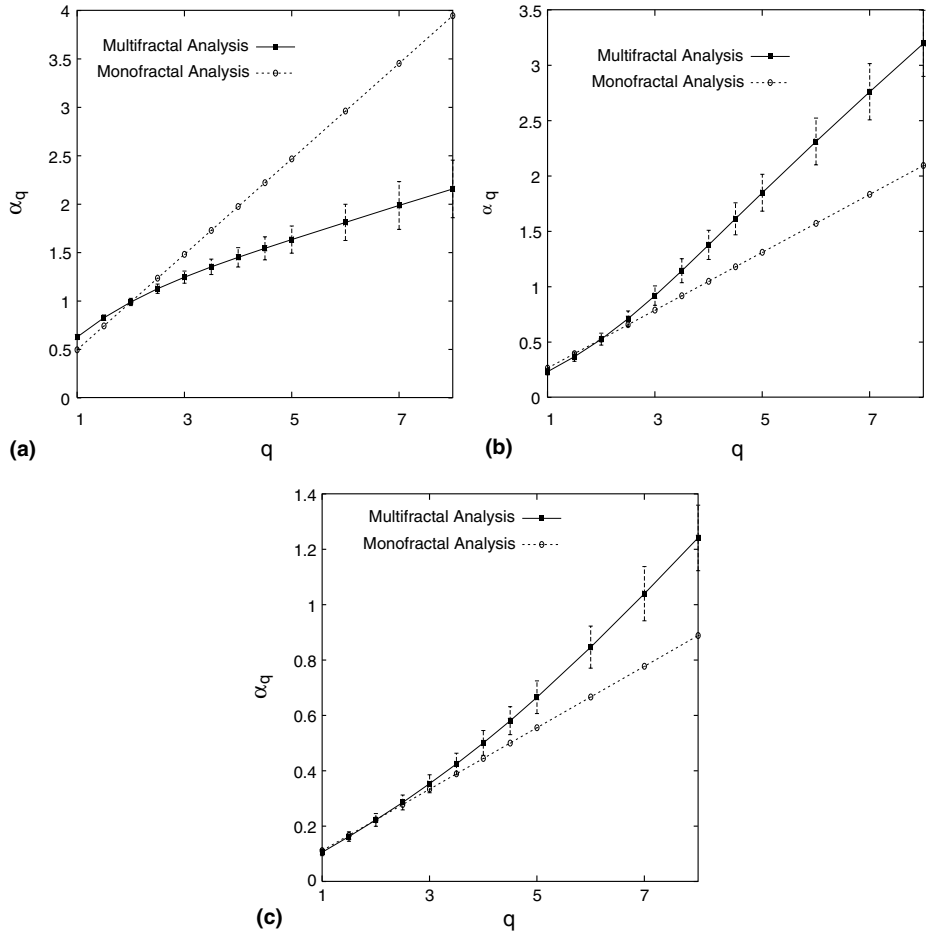

Fig. 1. Multiscale diagrams for the traces (a) MEM-1053844177, (b) 20000125-143640-1, and (c) dec-pkt-4.

Both monofractal and multifractal analysis of the traces in Table 1 were carried out using the A-V estimator [18], which has its code available in [19]. Monofractality is observed when the scaling exponent $\alpha_q = \varsigma(q) + q/2$ presents a linear behavior, $\varsigma(q) = Hq$, whereas multifractality is detected by non-linear behavior. Fig. 1 presents the multiscale analysis of the traces MEM-1053844177, dec-pkt-4 and 20000125-143640-1. Small time scales ($j1, j2$) defined in the range [1, 5] were used. Each entry $Y(k)$ represents the number of bytes observed at the points of aggregation during the kth time slot of duration γ, set to 1 ms, 10 ms and 100 ms for the traces MEM-1053844177, dec-pkt-4 and 20000125-143640-1, respectively. Multifractality can be identified by the presence of different slopes for different intervals. For instance, the trace MEM-1053844177 (Fig. 1b) has two distinct regions of q values, [0–3] and [3–8], indicating that multifractal modeling should be used.

Fig. 2 shows the raw data in the traces MEM-1053844177, dec-pkt-4 and 20000125-143640-1 as well the Holder functions at the same points computed according to the procedure in [15]. The highly changeable nature of the Holder function value is additional evidence of the multifractal nature of these traces.

Fig. 3 presents, both the envelope and the amount of work arriving. The mBm envelope process can be seen to provide a tight bound for the accumulated traffic arriving in real networks. This process was also validated using the other traces listed in Table 1, and a similar precision resulted.

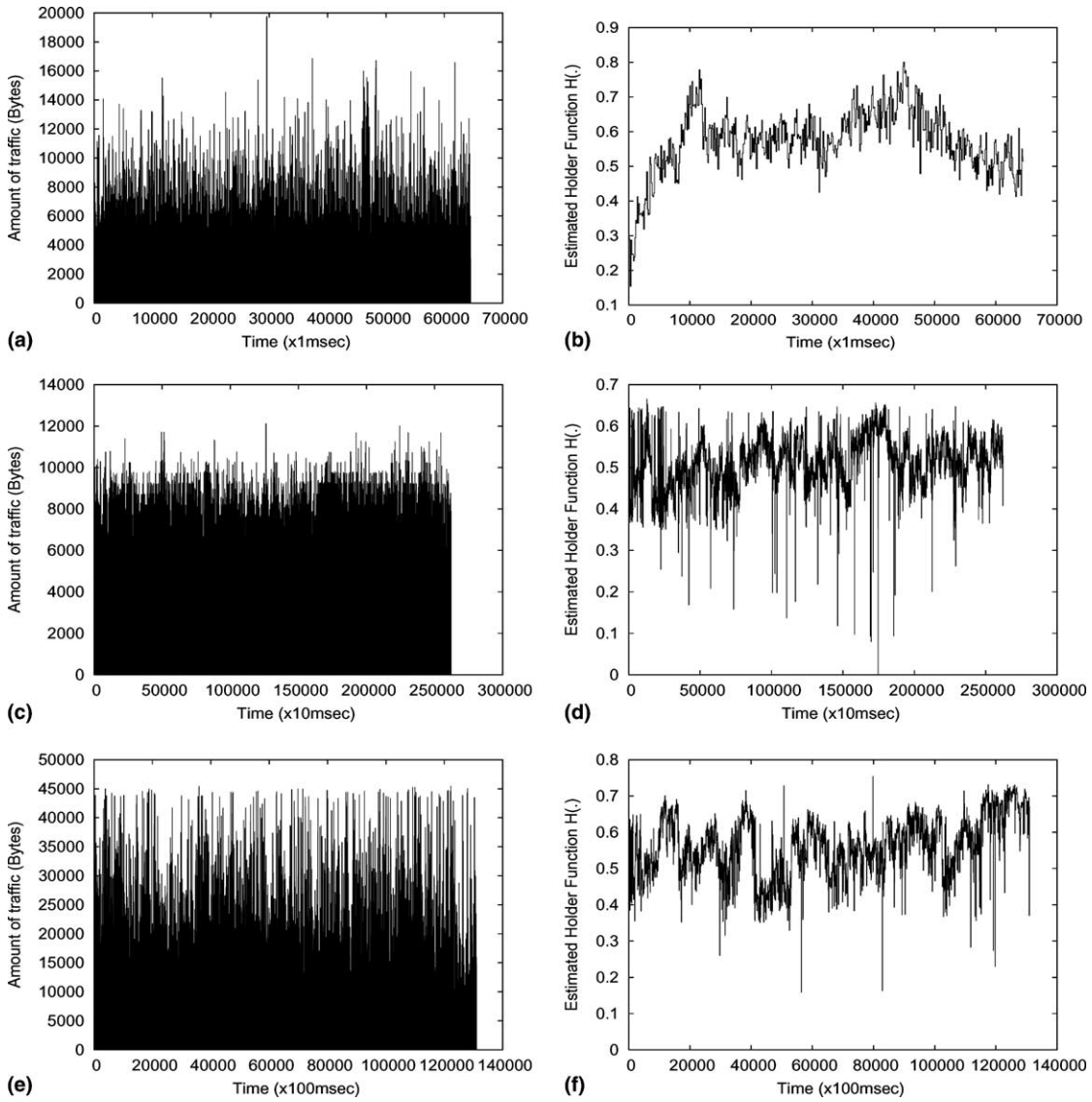

Fig. 2. Raw data and estimated Holder function of real network traffic: (a) raw data (MEM-1053844177), (b) Holder function (MEM-1053844177), (c) raw data (dec-pkt-4), (d) Holder function (dec-pkt-4), (e) raw data (20000125-143640-1) and (f) Holder function (20000125-143640-1).

To answer the question of whether real network traces can be modeled using a monofractal process, fBm envelope processes (Eq. (5)) were derived for the same traces used. The A-V estimator [18] was employed to evaluate the Hurst parameter value. Table 2 gives the parameters for the traces MEM-1053844177, dec-pkt-4 and 20000125-143640-1, whereas Fig. 4 shows the accumulated real traffic and monofractal envelope process. Since monofractal envelopes consider only the global burstiness value (the Hurst parameter), the monofractal fBm envelope process deviates greatly from the accumulated real traces, and consequently overestimates the dynamic (local) burstiness of what is actually multifractal traffic. Fig. 4 shows clearly that monofractal models do not capture the dynamics of multifractal flow.

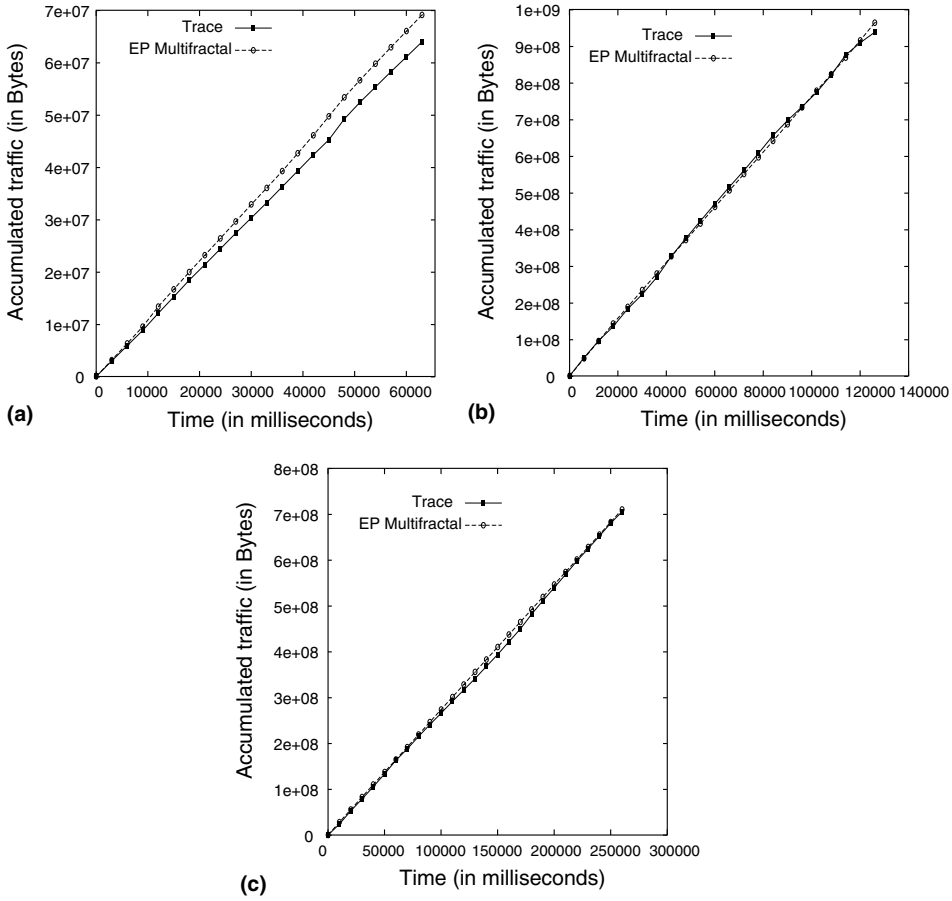

Fig. 3. Evaluation of mBm envelope process using real network traffic: (a) trace MEM-1053844177, (b) trace 20000125-143640-1, and (c) trace dec-pkt-4.

Table 2
Parameters of fBm envelope process used in the experiments

Trace	Mean ($\bar{a}$)	Variance σ^2	H
MEM-1053844177	1013.8	3245708.7	0.78
20000125-143640-1	7354.2	37326523.3	0.87
dec-pkt-4	2711.0	3813455.9	0.82

The mBm generator introduced in [12] was used to generate synthetic data for validation. Up to 10^6 data samples from mBm processes were considered, and various Holder functions were employed. Fig. 5 shows the accumulated traffic generated synthetically, as well as the mBm envelope process. The Holder functions used were:

$$\begin{aligned} H(t) &= 1.9t^2 - 1.9t + 0.975, \quad t \in (0,1); \\ H(t) &= t/2.0 + 0.5, \quad t \in (0,1). \end{aligned} \tag{6}$$

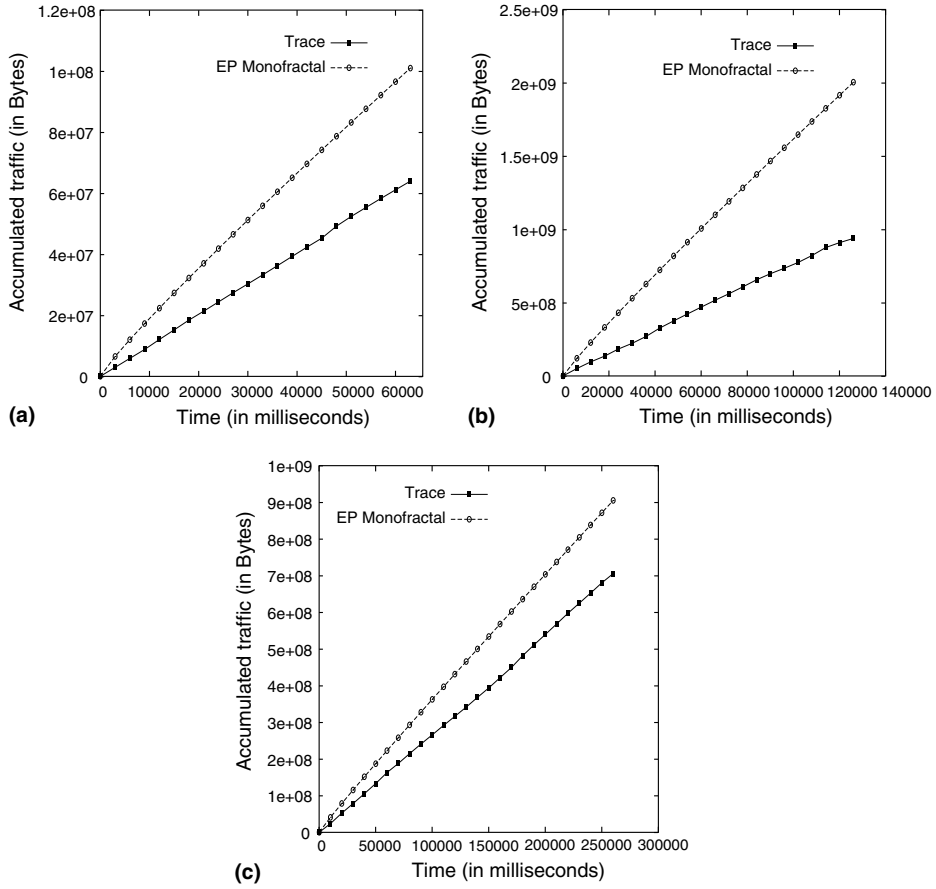

Fig. 4. Evaluation of the fBm envelope process using real network traffic: (a) trace MEM-1053844177, (b) trace 20000125-143640-1, and (c) trace dec-pkt-4.

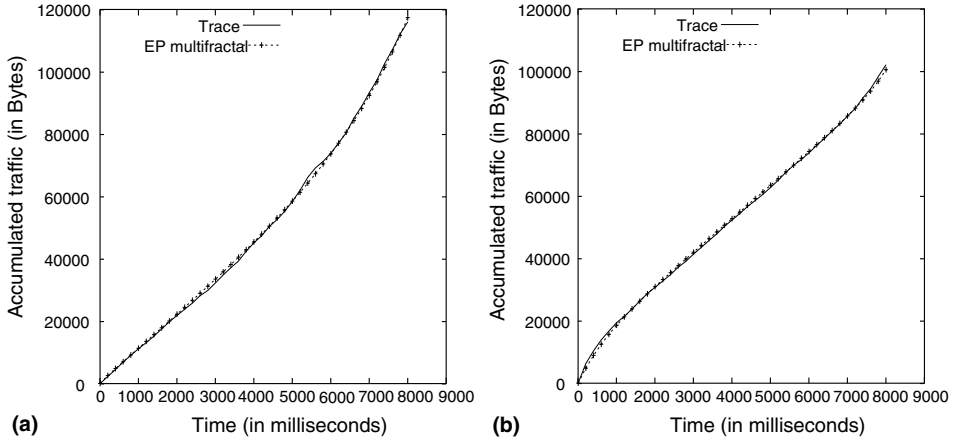

Fig. 5. Evaluation of the mBm envelope process using synthetic traces: (a) $H(\cdot)$ a linear function and (b) $H(\cdot)$ a quadratic function.

The mBm envelope process also provides a tight bound independent of the Holder function, although violations of the established bound are smaller than those established by pre-defined probability values.

As will be seen in Section 5, a closed analytical expression for Holder exponents is not necessary for the computation of the equivalent bandwidth of a traffic stream. However, such a closed expression can be useful for other purposes. The derivation of a polynomial approximation for the Holder function can be useful, since such approximations produce accurate results. In these experiments, the Holder function values for real network traces were computed using the procedure defined in [15]. Then, polynomial approximations for these Holder function values were derived. Polynomials of different degrees were then tested. Finally, the results of envelope processes using polynomial Holder functions were compared to actual traces from real network traces.

Fig. 6 shows the envelope processes derived using a Holder function approximated by a seven-degree polynomial for the traces MEM-1053844177, 20000125-143640-1 and dec-pkt-4. The precision observed was similar to that obtained using polynomials to the degree defined in the range [7,15]. For polynomials with a lower power, however, the bounds produced were loose. Precision did increase slightly for powers up to 15; but the use of a seventh power was found to provide a relatively good trade-off between accuracy and efficiency.

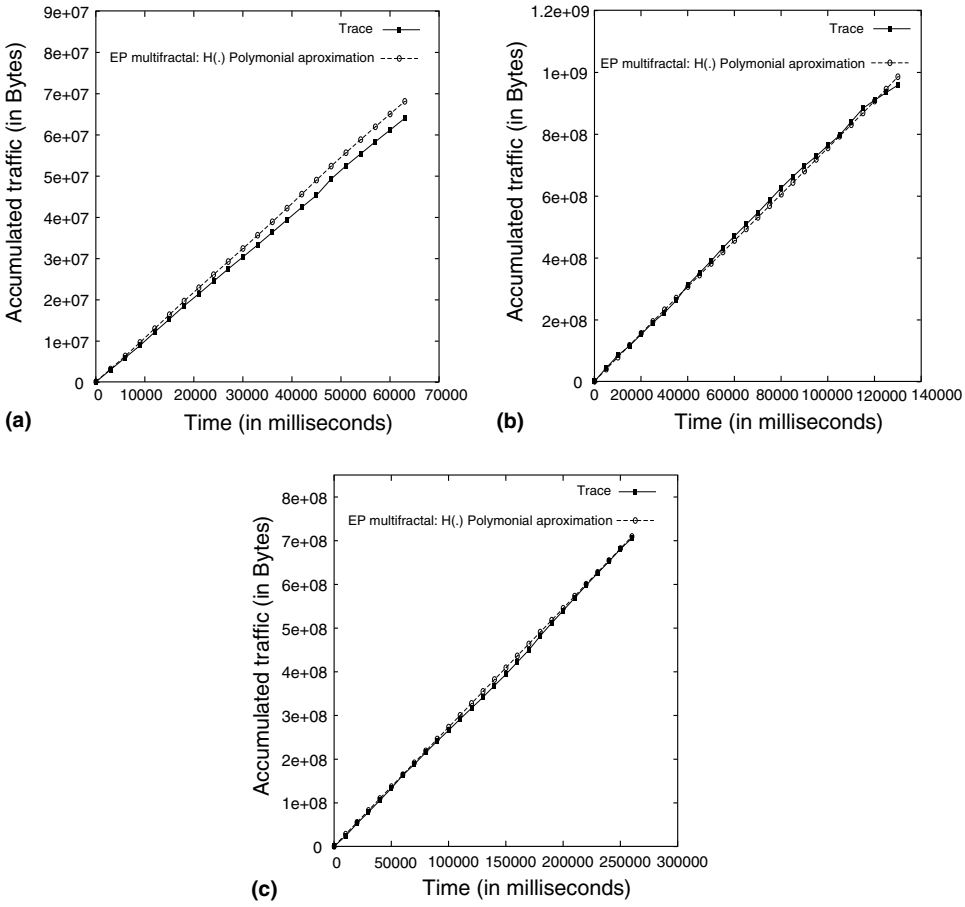

Fig. 6. Envelope processes derived using polynomial approximation for the Holder function: (a) trace MEM-1053844177, (b) trace 20000125-143640-1, and (c) trace dec-pkt-4.

4. Computation of the length of a queue fed by multifractal flows

In this section, an expression of the length of a queue fed by an mBm envelope process is presented, and the expression of the time at which the queue length reaches its maximum value is also derived.

4.1. Computation of queue length

The queue length at time t of a First-In First-Out (FIFO) queue with constant rate is given by the following equation [20]:

$$Q(t) = A(t) - S(t), \tag{7}$$

where $A(t)$ and $S(t)$ are the amount of work arriving and the amount of work served up to time t. Moreover, $S(t)$ can be computed as

$$S(t) = Ct + \min\left\{0, \inf_{t\geqslant 0}\{A(t) - Ct\}\right\} = Ct + A(t^{\bullet}) - Ct^{\bullet}, \tag{8}$$

where $A(0) = 0$ and $t^{\bullet} = \arg\inf_{t\geqslant 0}\{A(t) - Ct\}$, the time instant at which the largest idle period in the interval $[0, t]$ ends.

Let $r^{\bullet} = t/t^{\bullet}$, $S(\cdot)$ can be expressed as

$$S(t) = Ct + A(t/r^{\bullet}) - Ct/r^{\bullet}. \tag{9}$$

Thus, $Q(t)$ can also be expressed as

$$Q(t) = A(t) - S(t) = A(t) - A(t/r^{\bullet}) - Ct(1 - 1/r^{\bullet}). \tag{10}$$

By bounding the behavior of the arrival process with the mBm envelope process, it is possible to transform the problem of obtaining a solution for the stochastic system $Q(\cdot)$ into the easier problem of finding the solution for a deterministic system, $\hat{Q}(\cdot)$.

The upper bound for the amount of service provided up to the time instant t, the service envelope, can be computed as

$$\hat{S}(t) = Ct + \hat{A}(t/\hat{r}^{\bullet}) - Ct/\hat{r}^{\bullet}. \tag{11}$$

The upper bound for the queue length $Q(\cdot)$, $\hat{Q}(\cdot)$, can then be calculated by considering both the arrival and service envelopes. $\hat{Q}(\cdot)$ is given by

$$\begin{aligned}\hat{Q}(t) &= \hat{A}(t) - \hat{S}(t)\\ &= \hat{A}(t) - \hat{A}(t/\hat{r}^{\bullet}) - Ct(1 - 1/\hat{r}^{\bullet})\\ &= \int_0^t \bar{a} + \kappa\sigma H(x)x^{H(x)-1}\,\mathrm{d}x - \int_0^{t/\hat{r}^{\bullet}} \bar{a} + \kappa\sigma H(x)x^{H(x)-1}\,\mathrm{d}x - Ct(1 - 1/\hat{r}^{\bullet}).\end{aligned} \tag{12}$$

4.2. Time scale of interest of a queue fed by a single multifractal flow

The time scale of interest, at which the queue length reaches its maximum value, $t^{\star}$, is found by solving the following equation:

$$\hat{q}_{\max} = \max_{t\geqslant 0}\{\hat{Q}(t)\}, \tag{13}$$

which yields

$$\hat{t}^{\star} = \left\lceil \frac{\kappa\sigma\left(H(\hat{t}^{\star})(\hat{t}^{\star})^{H(\hat{t}^{\star})} - \hat{r}^{\bullet}H(\hat{t}^{\star}/\hat{r}^{\bullet})(\hat{t}^{\star}/\hat{r}^{\bullet})^{H(\hat{t}^{\star}/\hat{r}^{\bullet})}\right)}{(C-\bar{a})(1-1/\hat{r}^{\bullet})} \right\rceil. \tag{14}$$

The time scale of interest defines the point in time at which, in a probabilistic sense, the unfinished work in the queuing system is at its maximum, indicating that the average arrival rate has dropped below the link capacity so that the queue size will start decreasing. At this point in time, the source rate still exceeds the link capacity but afterward the probability that the average arrival rate will exceed the link capacity is negligible. Such a time scale is the most important for estimating the equivalent bandwidth of a stream.

Note that to evaluate $\hat{t}^{\star}$ it is necessary to know the actual value of $H(\cdot)$ at specific time instants, rather than the analytical expression of $H(\cdot)$. In other words, $\hat{t}^{\star}$ can be computed by using the sample mean, sample variance and specific values of $H(\cdot)$ which are measured from the traffic stream, making the framework introduced in this paper appropriate for the real time estimation of the equivalent bandwidth of a traffic stream.

Simulation experiments were conducted using both synthetic and real network traces to verify the accuracy of Eqs. (12) and (14). The evolution of the queue length was recorded for various levels of utilization. The time scale of interest, $t^{\star}$, for the trace MEM-1053844177 aggregated at a time scale of 1 ms was computed and compared to that found in the experiments. Fig. 7a and b show this evaluation for utilization values of 0.7 and 0.9. For that arrival process, the mean and variance are $\bar{a} = 1013.83$ and $\sigma^2 = 3245708.75$, respectively. The probability of violation was defined as 10^{-3}, with the Holder function $H(\cdot)$ estimated using the method described in [15].

The precision of the estimation of $\hat{t}^{\star}$ increases with the load. For $\rho = 0.7$ the estimated value deviates 3% from the measured values (45,698 ms) while for $\rho = 0.9$ it deviates 0.69% from the measured value (48,130 ms). Maximum deviation was found to be 3.5% for the traces in Table 1. The deviation in simulation experiments with synthetic traces was negligible.

When the amount of work is bounded by a monofractal envelope process (Eq. (5)) the time scale of interest, $\hat{t}^{\star}$, is given by the following:

$$\hat{t}^{\star} = \left\lceil \frac{\kappa\sigma H(1-(\hat{r}^{\bullet})^{(-H)})}{(C-\bar{a})(1-1/\hat{r}^{\bullet})} \right\rceil^{\frac{1}{1-H}}. \tag{15}$$

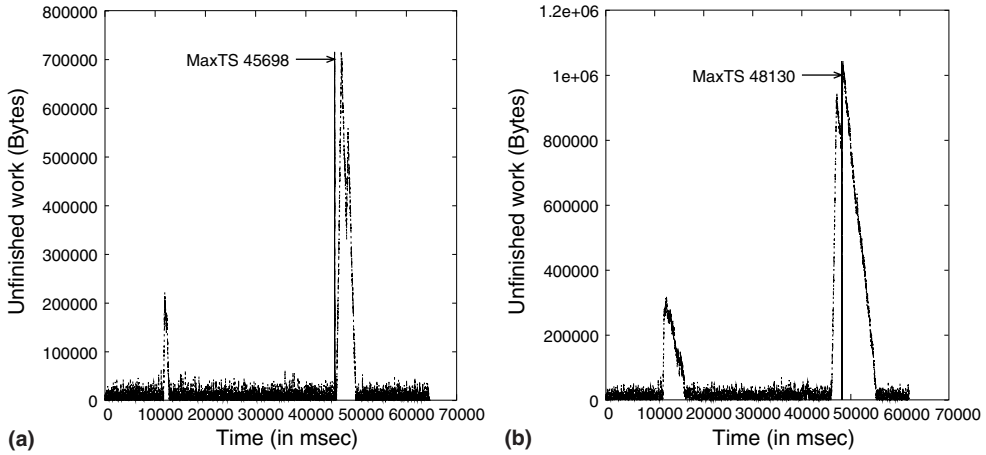

Fig. 7. Evaluation of queue length for different utilization levels and the estimated MaxTS for a system with different utilization levels: (a) $\rho = 0.70$ and (b) $\rho = 0.90$.

The Hurst parameter, H, plays an important role in this equation, since the $t^{\bigstar}$ value increases exponentially with $[\frac{1}{1-H}]$. Hence, misleading results can be expected when multifractal traffic is modeled using the monofractal envelope process, since the Hurst parameter, H, overestimates the variation of the Holder exponent.

Simulation experiments to estimate the time scale of interest were carried out for the amount of work in trace MEM-1053844177 which is bounded by a monofractal envelope process. The estimated value of $\hat{t}^{\bigstar}$ for utilization levels 0.7 and 0.9 were 231,543 ms and 3,242,741 ms, respectively; these results reinforce the fact that the monofractal envelope process provides a loose upper bound for multifractal flows.

4.3. Time scale of interest of a queue fed by several multifractal flows

An expression for the envelope process for modeling the aggregation of several flows is also needed to estimate the time scale of interest. The amount of traffic in this aggregated flow can be computed using the *local asymptotically self-similar* (lass) property. In [21] it was shown that the aggregate of N fBm processes with mean $\bar{a}_i$ and variance σ_i^2 is an fBm process with mean of $\bar{a} = \sum_{i=1}^{N} \bar{a}_i$ and of $\sigma^2 = \sum_{i=1}^{N} \sigma_i^2$. Since, an mBm process can be represented locally by a single fBm process resulting from the aggregation of several fBm processes, the mBm envelope process can also be approximated for small time scales by using an overall fBm envelope process resulting from the aggregation of N fBm envelope processes.

The derivation of $\hat{A}^N(\cdot)$ is as follows: assume N independent flows defined by the following parameters: mean $\bar{a}_i$, variance σ_i^2 and Holder exponents $H_i(t)$. Let the aggregate process be denoted by $W(\cdot)$, with the envelope process for each flow given by $\hat{A}_i(t)$. The aggregate envelope process $\hat{A}^N(\cdot)$ for the cumulative work of $W(\cdot)$ in the interval $[0, t]$ is then given by

$$\hat{A}^N(t) = \sum_{i=1}^{N} \hat{A}_i(t) = \int_0^t \sum_{i=1}^{N} \bar{a}_i + \kappa \left(\sum_{i=1}^{N} \sigma_i^2 H_i(x) x^{2H_i(x)-1} \right) \left(\sum_{i=1}^{N} \sigma_i^2 x^{2H_i(x)} \right)^{-1/2} \mathrm{d}x, \tag{16}$$

where $\hat{A}_i(t)$ is the envelope process for the ith flow.

$\hat{A}^N(\cdot)$ can now be inserted into Eq. (11) to obtain an envelope process for the amount of service provided up to time instant t, which gives the following:

$$\hat{S}^N(t) = Ct + \hat{A}^N(t/\hat{r}^{\bullet}) - Ct/\hat{r}^{\bullet}. \tag{17}$$

Moreover, an upper bound for the queue length $Q^N(\cdot)$, $\hat{Q}^N(\cdot)$, can be computed by considering both the arrival and service envelopes. $\hat{Q}^N(\cdot)$ is given by

$$\begin{aligned}
\hat{Q}^N(t) &= \hat{A}^N(t) - \hat{S}^N(t) \\
&= \hat{A}^N(t) - \hat{A}^N(t/\hat{r}^{\bullet}) - Ct(1 - 1/\hat{r}^{\bullet}) \\
&= \int_0^t \sum_{i=1}^{N} \bar{a}_i + \kappa \left(\sum_{i=1}^{N} \sigma_i^2 H_i(x) x^{2H_i(x)-1} \right) \left(\sum_{i=1}^{N} \sigma_i^2 x^{2H_i(x)} \right)^{-1/2} \mathrm{d}x \\
&\quad - \int_0^{t/\hat{r}^{\bullet}} \sum_{i=1}^{N} \bar{a}_i + \kappa \left(\sum_{i=1}^{N} \sigma_i^2 H_i(x) x^{2H_i(x)-1} \right) \left(\sum_{i=1}^{N} \sigma_i^2 x^{2H_i(x)} \right)^{-1/2} \mathrm{d}x - Ct(1 - 1/\hat{r}^{\bullet}).
\end{aligned} \tag{18}$$

The time scale of interest here is the time instant t at which $\hat{Q}^N(\cdot)$ reaches its maximum value, i.e.

$$\hat{q}_{\max}^N = \max_{t \geqslant 0} \{\hat{Q}^N(t)\} \tag{19}$$

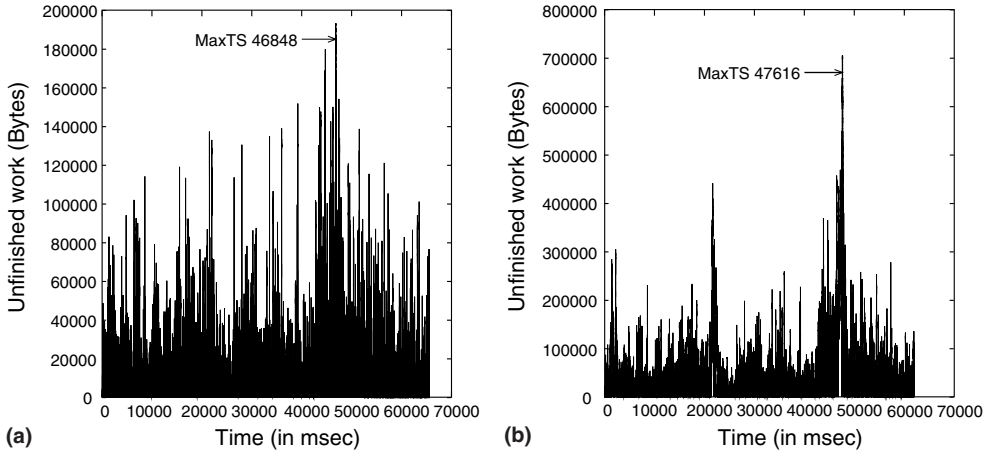

Fig. 8. Evaluation of queue length for a system fed by several flows and the MaxTS for a system under different utilization levels: (a) $\rho = 0.70$ and (b) $\rho = 0.90$.

or

$$\kappa\left[\left(\sum_{i=1}^{N}\sigma_i^2 H_i(t)t^{2H_i(t)-1}\right)\left(\sum_{i=1}^{N}\sigma_i^2 t^{2H_i(t)}\right)^{-1/2} + \left(\sum_{i=1}^{N}\sigma_i^2 H_i(t/\hat{r}^{\bullet})(t/\hat{r}^{\bullet})^{2H_i(t/\hat{r}^{\bullet})-1}\right)\left(\sum_{i=1}^{N}\sigma_i^2 (t/\hat{r}^{\bullet})^{2H_i(t/\hat{r}^{\bullet})}\right)^{-1/2}\right] - \left(C - \sum_{i=1}^{N}\bar{a}_i\right)(1 - 1/\hat{r}^{\bullet}) = 0. \quad (20)$$

This equation can be solved numerically using classical root-finding methods such as the Newton–Raphson [22] which requires low computational time. To verify the accuracy of Eq. (20), various simulation experiments were conducted using the traces presented in Table 1. A queue was fed by the aggregate of these flows, and the evolution of the queue length was recorded for different levels of utilization. Fig. 8 shows the results obtained from simulation experiments using traces MRA-1057960474, MEM-1053844177, COS-1057970154 and AIX-1049492523. The estimated and measured time scale values differ by less than 1%. For a utilization level of 0.8, the estimated maximum time scale value was 46,848 ms, whereas the measured value was 46,805 ms. For a utilization level of 0.9, the estimated and measured maximum time scale values were 47,616 ms and 47,607 ms, respectively. These results reinforce those found for a single flow (Eq. (14)), thus suggesting that precision increases with load.

A special case of statistical multiplexing is the multiplexing of homogeneous flows, in which the envelope process of the aggregated traffic $A^N(\cdot)$, $\hat{A}^N(t)$, is defined as:

$$\hat{A}^N(t) = \sum_{i=1}^{N}\hat{A}_i(t) = \int_0^t N\bar{a} + N^{1/2}\kappa\sigma H(x)x^{H(x)-1}\,\mathrm{d}x, \quad (21)$$

where the upper bound for the queue length $Q^N(\cdot)$, $\hat{Q}^N(\cdot)$, is defined by

$$\begin{aligned}\hat{Q}^N(t) &= \hat{A}^N(t) - \hat{S}^N(t)\\ &= \hat{A}^N(t) - \hat{A}^N(t/\hat{r}^{\bullet}) - Ct(1 - 1/\hat{r}^{\bullet})\\ &= \int_0^t N\bar{a} + N^{1/2}\kappa\sigma H(x)x^{H(x)-1}\,\mathrm{d}x - \int_0^{t/\hat{r}^{\bullet}} N\bar{a} + N^{1/2}\kappa\sigma H(x)x^{H(x)-1}\,\mathrm{d}x - Ct(1 - 1/\hat{r}^{\bullet}).\end{aligned} \quad (22)$$

The time scale of interest for these homogeneous flows is given by the following equation:

$$\hat{t}^{\star} = N^{-1/2}\left[\frac{\kappa\sigma\left(H(\hat{t}^{\star})(\hat{t}^{\star})^{H(\hat{t}^{\star})} - \hat{r}^{\bullet}H(\hat{t}^{\star}/\hat{r}^{\bullet})(\hat{t}^{\star}/\hat{r}^{\bullet})^{H(\hat{t}^{\star}/\hat{r}^{\bullet})}\right)}{(c-\bar{a})(1-1/\hat{r}^{\bullet})}\right] = N^{-1/2}\hat{t}_i^{\star}, \tag{23}$$

where $\hat{t}_i^{\star}$ is the time scale of interest for a single flow (Eq. (14)), with link capacity normalized according to the number of flows, N, i.e., $c = C/N$.

5. Multifractal flow aggregate equivalent bandwidth

In this section, a method for computing the bandwidth necessary to support requirements for buffer overflow is proposed, as well as one for determining the maximum probabilistic delay for an aggregate of heterogeneous flows. The problem in this section can be stated as follows:

Given a set of flows with mean $\bar{a}_i$, standard deviation σ_i and the Holder exponent expressed by $H_i(t)$, what is the link capacity needed so that the maximum queue size will be bounded by $\hat{q}^N_{\max}$ with probability ϵ?

To answer this question, it is necessary to find the effective bandwidth value, $\hat{C}$, which should satisfy the following relationship:

$$\max_{t>0}\{\hat{Q}^N(t)\} - \hat{q}^N_{\max} = 0, \tag{24}$$

$$\max_{t>0}\{\hat{A}^N(t) - \hat{A}^N(t/\hat{r}^{\bullet}) - Ct(1-1/\hat{r}^{\bullet})\} - \hat{q}^N_{\max} = 0, \tag{25}$$

where $\hat{Q}^N(\cdot)$ is an upper bound for the queue length (Eq. (18)). $\hat{C}$ can be computed by traditional numerical methods, such as the Newton–Raphson or Quasi-Newton methods [22]. However, in such methods, convergence to a solution is highly dependent on the initial values given, which makes the approach inappropriate for real-time implementation, since convergence may be greatly delayed. An heuristic method based on knowledge of the specific queueing system was developed and is presented in Fig. 9.

This heuristic is based on the fact that the equivalent bandwidth is lower bounded by the mean arrival rate $\bar{a}$ and upper bounded by the peak rate. Since the latter may not be known, an Unbounded Binary Search (UBS) is conducted in the interval $[\bar{a}, C)$, where C is the channel capacity; this UBS is subject to the restriction given in Eq. (25).

A UBS is carried out in two steps. In the first step, an effective bandwidth candidate value $\hat{C}$ is defined as the mean value increased by a small increment $\bar{a} + \epsilon$; this value is then inserted into Eq. (25). If Eq. (25) is satisfied, the desired value of the equivalent bandwidth has been found. Otherwise, the value of $\hat{C}$ is repeatedly doubled until $\hat{q}^N_{\max} > \max_{t>0}\{\hat{Q}^N(t)\}$. With the $\hat{C}$ value derived in the first step, a Binary Search is conducted on the interval $(\hat{C}/2, \hat{C}]$. Theorem 1 establishes the convergence of this algorithm.

Theorem 1. *The equivalent bandwidth of an aggregate of N flows can be computed in at most*

$$n = \mathrm{O}(\log(C)) \tag{26}$$

iterations, where C is the channel capacity in bits/sec.

The proof for this theorem is given in Appendix A. The execution times for the algorithm were recorded for different combinations of traces. The machine used had a 1 GHz AMD processor, 256 MB of memory and the LINUX RedHat 8.0 operating system. The execution time never exceeded 250 ms, which is quite encouraging for the implementation of such an algorithm in real time.

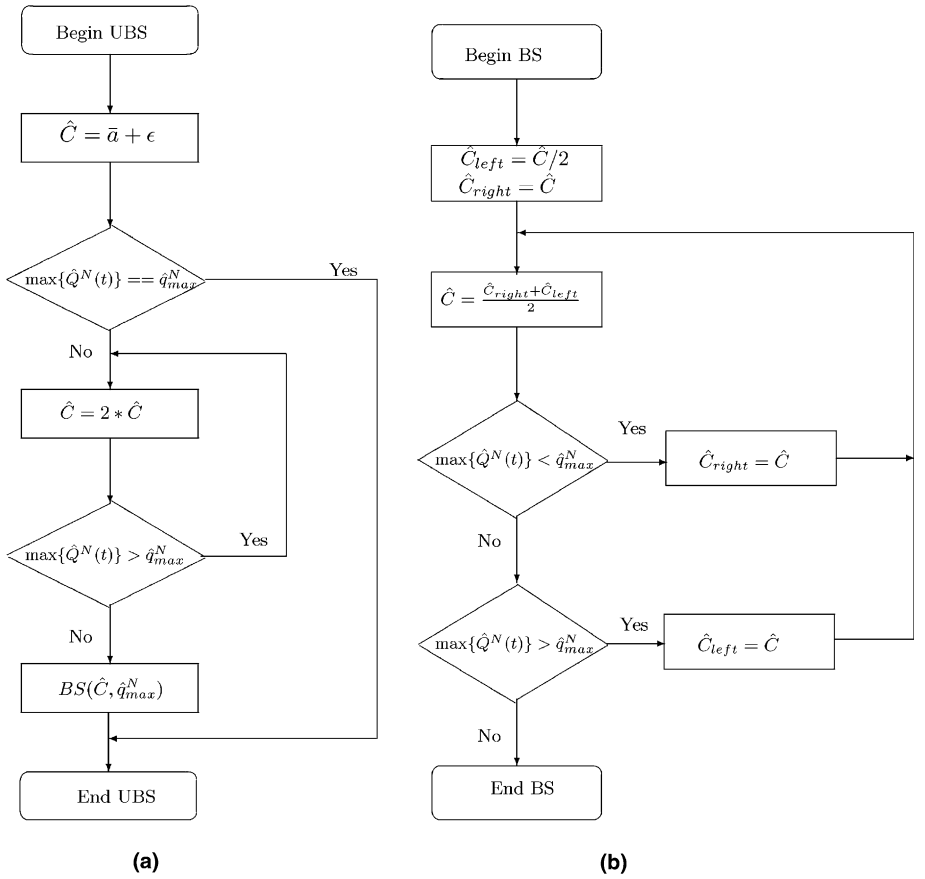

Fig. 9. The algorithm for computation of the equivalent bandwidth of a flow: (a) Step 1: Initialization and (b) Step 2: Binary research.

The whole advantage of statistical multiplexing is the efficient use of resources achieved by interleaving packets from different streams, which allows the support of a greater number of flows than circuit switching does. This benefit can be evaluated by calculating the gain measure, $G(n)$, defined as the ratio between n times the equivalent bandwidth of a flow and the equivalent bandwidth for the aggregate of n homogeneous flows. $G(n)$ is given by

$$G(n) = \frac{\sum_{i=1}^{n} EB_i}{EB(n)} = \frac{\dfrac{\left[\sum_{i=1}^{n} \int_0^{\hat{t}_i^{\star}} \bar{a}_i + \kappa\sigma_i H_i(x) x^{H_i(x)-1}\,\mathrm{d}x - \int_0^{\hat{t}_i^{\star}/\hat{r}_i^{\bullet}} \bar{a}_i + \kappa\sigma_i H_i(x) x^{H_i(x)-1}\,\mathrm{d}x - K\right]}{\left[\hat{t}_i^{\star}(1-1/\hat{r}_i^{\bullet})\right]}}{\dfrac{\left[\int_0^{\hat{t}^{\star\star}} \sum_{i=1}^{n} \bar{a}_i + \kappa \dfrac{\left(\sum_{i=1}^{n} \sigma_i^2 H_i(x) x^{2H_i(x)-1}\right)}{\left(\sum_{i=1}^{n} \sigma_i^2 x^{2H_i(x)}\right)^{\frac{1}{2}}}\,\mathrm{d}x - \int_0^{\hat{t}^{\star\star}/\hat{r}_n^{\bullet}} \sum_{i=1}^{n} \bar{a}_i + \kappa \dfrac{\left(\sum_{i=1}^{n} \sigma_i^2 H_i(x) x^{2H_i(x)-1}\right)}{\left(\sum_{i=1}^{n} \sigma_i^2 x^{2H_i(x)}\right)^{\frac{1}{2}}}\,\mathrm{d}x - K'\right]}{\left[\hat{t}^{\star\star}(1-1/\hat{r}_n^{\bullet})\right]}}, \tag{27}$$

where EB_i is the equivalent bandwidth of ith flow and $EB(n)$ is the equivalent bandwidth of an aggregate of n flows; $\hat{t}_i^{\star}$ and $\hat{t}^{\star\star}$ are the time scales resulting from the effective bandwidth computations of the ith flow and multiple flows, respectively. K is the buffer size at the multiplexer, and $K' = K/n$.

A special case is the multiplexing of homogeneous flows. Here, the gain is defined as the ratio between n times the equivalent bandwidth of a flow and the equivalent bandwidth for the aggregate of n homogeneous flows. $G(n)$ is given by

$$G(n) = \frac{nEB(1)}{EB(n)} = \frac{\dfrac{\left[\int_0^{\hat{t}^\star} \bar{a}+\kappa\sigma H(x)x^{H(x)-1}\,\mathrm{d}x - \int_0^{\hat{t}^\star/\hat{r}_1^\bullet} \bar{a}+\kappa\sigma H(x)x^{H(x)-1}\,\mathrm{d}x - K\right]}{\left[\hat{t}^\star(1-1/\hat{r}_1^\bullet)\right]}}{\dfrac{\left[\int_0^{\hat{t}^{\star\star}} \bar{a}+n^{\frac{-1}{2}}\kappa\sigma H(x)x^{H(x)-1}\,\mathrm{d}x - \int_0^{\hat{t}^{\star\star}/\hat{r}_n^\bullet} \bar{a}+n^{\frac{-1}{2}}\kappa\sigma H(x)x^{H(x)-1}\,\mathrm{d}x - K'\right]}{\left[\hat{t}^{\star\star}(1-1/\hat{r}_n^\bullet)\right]}}, \tag{28}$$

where $EB(1)$ is the equivalent bandwidth of a single flow and $EB(n)$ is the equivalent bandwidth of an aggregate of n flows, with $\hat{t}^\star$ and $\hat{t}^{\star\star}$ being the time scales given by the effective bandwidth computations of single and multiple flows, respectively. K is the buffer size at the multiplexer, and $K' = K/n$.

Fig. 10 shows the gain for traces with $H(\cdot)$, defined in Table 3 for different values of variance. The mean arrival rate is $\bar{a} = 1000$ and the variance for the flow labeled "low" is $\sigma^2 = 10{,}000$. For the curves labeled "average" and "high", the variance values are $10\sigma^2$ and $100\sigma^2$, respectively. The gain can thus be seen to increase with variance. For instance, for Holder exponents expressed as a quadratic function, the maximum gain is 1.35 for streams with low variance, whereas it is greater than 3.5 for streams with high variance. The gain is also influenced by the Holder exponent values. What is actually relevant is the mean value of the Holder exponent up to the time scale of interest, i.e., $\int_0^{\hat{t}^{\star\star}} H(x)\,\mathrm{d}x/\hat{t}^{\star\star}$. This relationship can be observed by comparing Fig. 10a and b. Traces with Holder exponents expressed by a quadratic function yield greater gains than those with exponents expressed by cubic function even when the mean and variance are the same. For flows with average variance, values of the maximum gain is 1.9 for traces with exponents expressed by a quadratic function, but only 1.2 for those with exponents expressed by a cubic function.

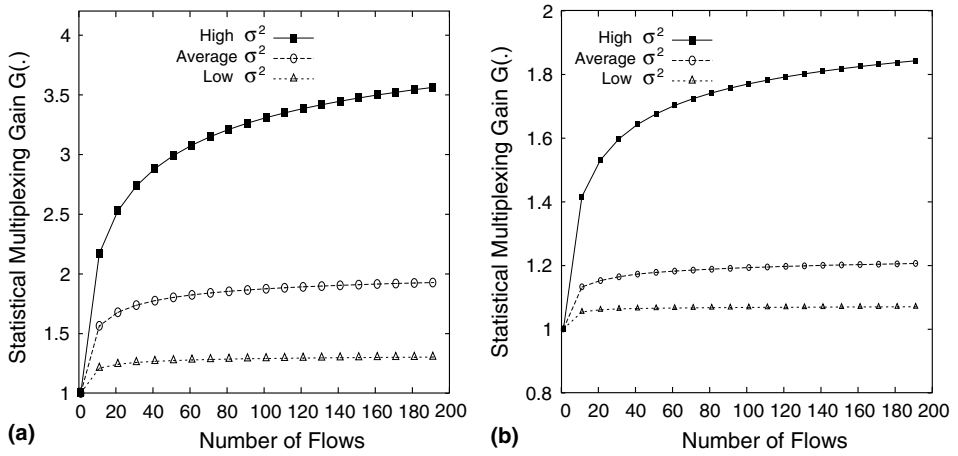

Fig. 10. Multiplexing gain obtained using synthetic data: (a) quadratic $H(\cdot)$ and (b) cubic $H(\cdot)$.

Table 3
Holder function of the process used for the generation of synthetic data

Flow	$H(\cdot)$
1	$\frac{19}{10}t^2 - \frac{19}{10}t + 0.985$
2	$\frac{49}{10}t^3 - \frac{79}{10}t^2 + \frac{33}{10}t + 0.51$
3	$\frac{21}{10}t^4 + \frac{11}{10}t^3 - \frac{1}{10}t^2 + \frac{8}{10}t + 0.51$
4	$\frac{\sin(t)}{10} + 0.61$
5	$\frac{1}{2}t + 0.5$

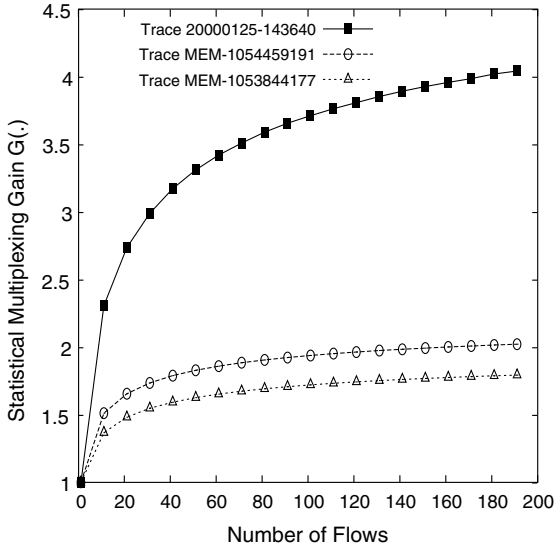

Fig. 11. Multiplexing gain obtained using real traffic traces.

Statistical multiplexing gain was also investigated using real network traces. Fig. 11 shows experiments using the traces MEM-1053844177, MEM-1054459191, and 20000125-143640-1 (Table 1). The gain for the trace 20000125-143640-1 is greater than that for the other traces. Again, this can be seen to be the result of the fact that it has a greater variance and a higher mean $H(\cdot)$ value up to the time the queue reaches its maximum.

To determine whether a monofractal approach can be used for resource dimensioning in networks with multifractal flows, a monofractal envelope (Eq. (5)) was derived for a trace with Holder exponents expressed by a quadratic function. Fig. 12 shows the gain considering both an mBm envelope process,

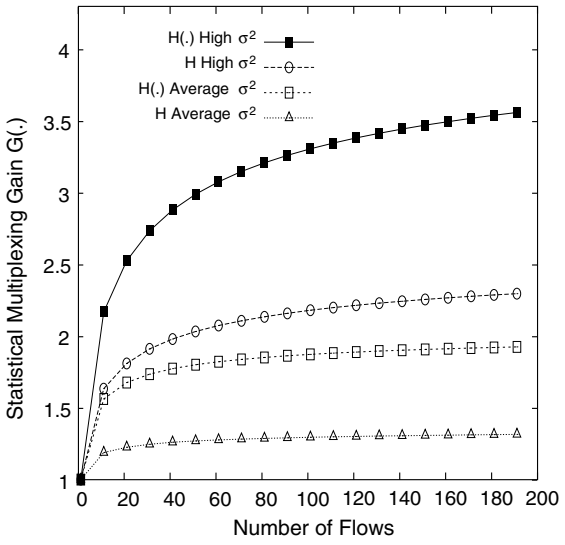

Fig. 12. Comparison between the statistical multiplexing gains obtained by the use of monofractal and multifractal approaches.

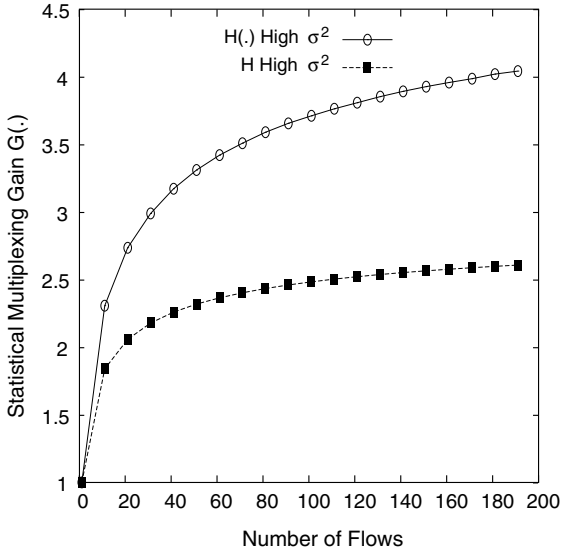

Fig. 13. Comparison between the statistical multiplexing gains obtained by the use of monofractal and multifractal approaches for real network traces.

denoted $H(\cdot)$, and the monofractal envelope process, denoted H, for both high and average variance values. The Hurst parameter H is defined as the mean value of the $H(\cdot)$ function up the instant of time $\hat{t}^{\star}$. It can be seen that the gains obtained by these two types of envelopes differ. For average variance values, the gain obtained by the use of the mBm envelope process is slightly one unit larger than the gain obtained by the use of the fBm envelope process, whereas it is almost two units for high variance values. This difference results from the fact that the Hurst parameter overestimates the local behavior of Holder exponents, leading to overprovisioning of resources and, consequently, a lower gain.

The appropriateness of monofractal modeling was also evaluated using real network traces. In Fig. 13, an example involving the trace 20000125-143640-1 is shown. Again, one can see that a monofractal model leads to overprovisioning of resources and a low gain (H High σ^2 curve).

6. Policing multifractal flows

Once a flow is admitted into a network domain, it must be policed in order to enforce that the stream of bytes generated is in accordance with the declared traffic descriptors at the admission time. An ideal policing mechanism would allow packets into the network if and only if the flow is well-behaved. Otherwise, incoming packets should be dropped or marked as low priority.

The Leaky Bucket policing mechanism has been adopted by several network standards. However, it assumes that the accumulated policed traffic grows linearly as a function of the variance, which does not happen in multifractal traffic. To cope with this deficiency, the bucket sizes needs to be quite large which is undesirable, since bursts of violating traffic can be misleadingly considered to be in conformance with the negotiated contract and be marked as having high priority. In [9], this problem was illustrated using monofractal traffic.

Simulation experiments with real network traffic have also been carried out to show that the Leaky Bucket demands an unrealistic bucket size for the policing of multifractal flows. The leaky rate was set

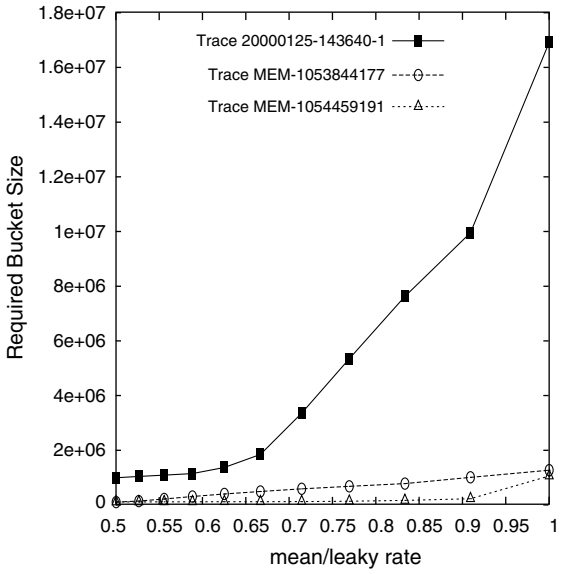

Fig. 14. Required bucket size for LB mechanism operating with different mean arrival to leaky rate ratio.

to the mean arrival rate, and the bucket size was calculated using an Unbounded Binary Search so that a target violation probability value was obtained for the flow being policed.

Fig. 14 shows the bucket size as a function of the ratio between the mean arrival rate and the leaky rate for the traces MEM-1053844177, MEM-1054459191 and 20000125-143640-1. Note that huge bucket sizes are demanded by all flows when the leaky rate is set to the mean arrival rate. Even when the leaky rate is twice the mean arrival rate, the bucket size is still too big. In this case, the calculated bucket sizes were 75,120, 77,866 and 967,820 bytes, respectively, for the traces MEM-1053844177, MEM-1054459191 and 20000125-143640-1. Moreover, a high leaky rate value (twice the mean rate) is also undesirable since it has been shown that when the leaky rate is set to such values, violating traffic is marked as in conformance [23].

A policing mechanism for monofractal traffic was introduced in [9]. In this paper such a mechanism has been extended to handle multifractal flows. The mechanism is called a multifractal leaky bucket (MFLB).

The amount of work accepted by MFLB is given by

$$\hat{L}(t) = \int_0^t \bar{a} + \Psi H(x) x^{H(x)-1}\,\mathrm{d}x + S, \tag{29}$$

where $\bar{a}$ is the average arrival rate of the flow. Ψ is given by $\kappa\sigma$, where κ is a constant and σ is the standard deviation of the flow.

The multifractal leaky bucket (MFLB) works as follows. A time window with the duration of τ time units is defined. The arrival rate during this time window is compared to the declared mean value ($\bar{a}\tau$). If it exceeds the mean value, the work which has arrived during this time window is compared to what is allowed by the MFLB envelope process (Eq. (29)) during the same period. If this work surpasses what is allowed, all packets in excess are marked. The time window is then increased by τ time units, i.e., a time window with duration of 2τ is now considered. This new window begins at the time the arrival has violated the declared mean arrival rate. If the work arriving during this new time window also exceeds the allowed volume of work for this period, the number of packets with a total volume minus the volume already marked are also marked.

As long as the mean number of arrivals surpasses the declared mean value, the sampling interval (time window) is increased in units of τ time units. Whenever the mean number of arrivals drops below the declared value, the time window is shrunk back to τ time units, and the mean arrival rate continues to be checked. Note that MFLB works with a specific envelope process which can be redefined dynamically if necessary.

A mathematical description of the dynamics of the MFLB is as follows. Let $C(\tilde{t}+n\tau)$ define the cumulative number of packets arriving during the interval $[\tilde{t}, \tilde{t}+n\tau]$:

$$C(\tilde{t}+n\tau) = A(\tilde{t}+n\tau) - A(\tilde{t}), \tag{30}$$

where $A(t)$ is the number of arrivals up to time t.

MFLB checks whether $C(\tilde{t}+n\tau)$ exceeds the allowed mean number of packets arriving during the interval $n\tau$, i.e, $\bar{a}n\tau$. If it does, the MFLB verifies if the number of arrivals exceeds the number of arrivals allowed by the envelope process where $\lambda(\tilde{t}+n\tau) = \hat{L}(\tilde{t}+n\tau) - \hat{L}(\tilde{t})$. If this limit is also surpassed, the excess not already marked in previous windows is marked, i.e., $C(\tilde{t}+n\tau) - \lambda(\tilde{t}+n\tau) - C(\tilde{t}+(n-1)\tau) + \lambda(\tilde{t}+(n-1)\tau)$ packets are marked. Then the MFLB increases the time window ($n = n + 1$) and repeats the whole process. Whenever the mean arrival rate drops below the declared value $\bar{a}$, the time window is shrunk back to τ time units (Fig. 15).

Fig. 16 presents the violation probabilities produced by an ideal mechanism, as well as by the MFLB and the LB with leaky rates of 1.5 and 2 times the mean arrival rate. Under an ideal policing mechanism, the violation probability jumps from a very low value to 1 as soon as the source starts transmitting with a mean arrival rate value above the declared value. The violation probability for a multifractal leaky bucket follows a pattern which is similar to the behavior of such an ideal mechanism, except that it reacts to violation at a rate of 1.10 the nominal value. This delay in reaction is quite acceptable for a non-ideal mechanism. On the

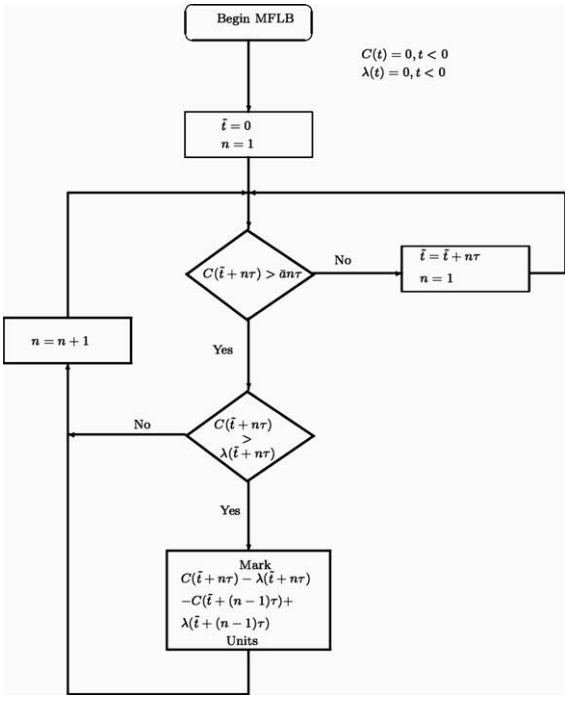

Fig. 15. Multifractal leaky bucket.

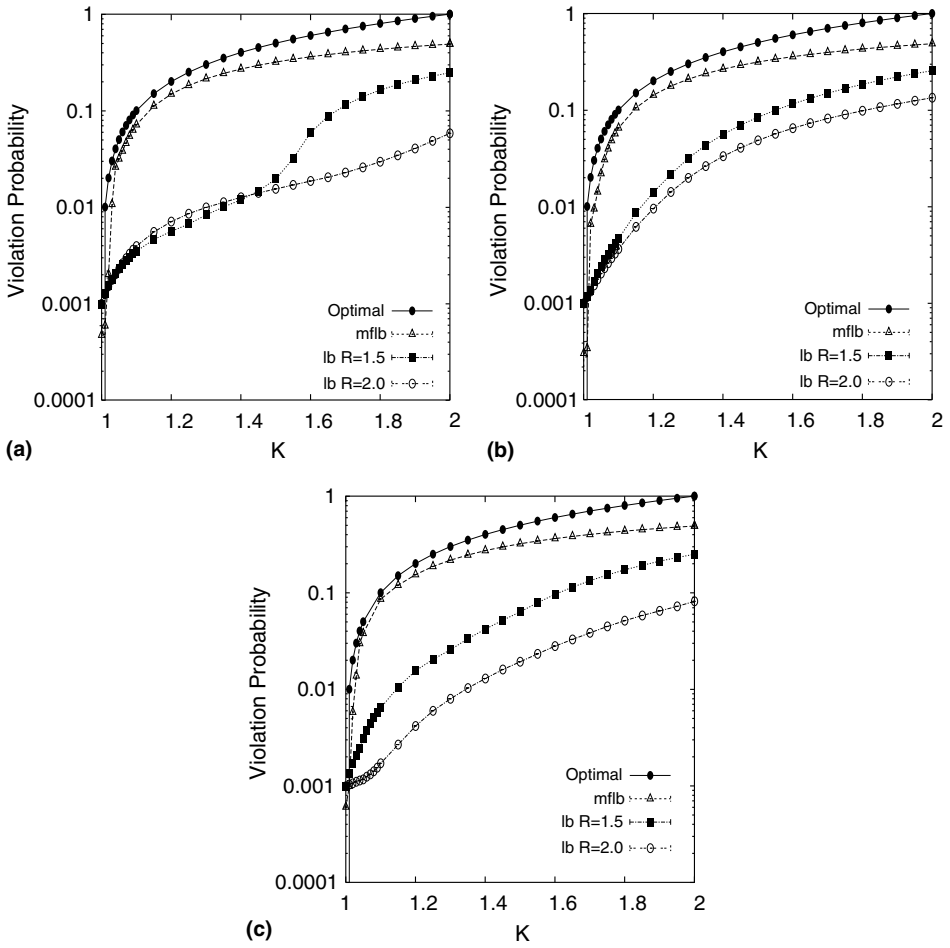

Fig. 16. A comparison between MFLB and LB for the detection of violating flows: (a) trace MEM-1053844177, (b) trace 20000125-143640-1, and (c) trace dec-pkt-4.

other hand, the leaky bucket makes no significant distinction between a violating source and a non-violating source, nor does it tend to 1 when arrival rate is high.

The multifractal leaky bucket mechanism compares the cumulative number of arrivals to the number of arrivals allowed by the MFLB envelope process at various points in time. The sensitivity of these results to the duration of the time window is evaluated by investigating the behavior with time windows 1, 100, 1000 and 5000 times longer than the window size used in the previous example. The outcomes are found not to be dependent on the duration of the time window. Fig. 17 shows the violation probability as a function of the arrival rate for different window durations. These findings reinforce the robustness of the multifractal leaky bucket as a policing mechanism for multifractal flows.

7. Related work

Erramilli et al. [2] have proposed that traffic should be modeled by random cascades for time scales smaller than a certain cutoff value and be represented by an fBm for larger scales. They show that for

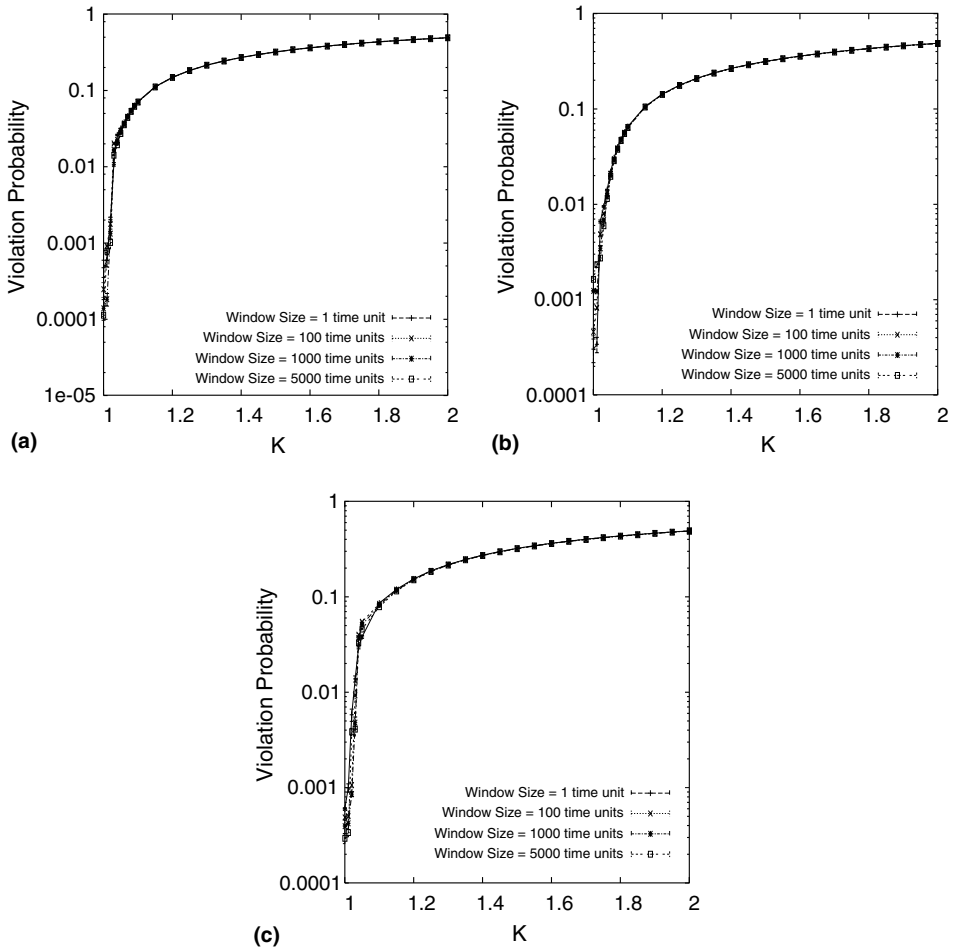

Fig. 17. Sensitivity analysis of MFLB: (a) trace MEM-1053844177, (b) trace 20000125-143640-1, and (c) trace dec-pkt-4.

IP traffic, the cutoff scale is of the order of one Round Trip Time (RTT), while for VBR videos, it is typically approximately one frame in duration. Erramili et al. showed that much more accurate results can be obtained by using their model rather than using purely monofractal models.

Other models based on multiplicative cascades have also been proposed. These models map a given sample onto a binary multiscale tree [24]. Each node of the tree corresponds to the aggregate of the traffic mapped onto its descendants. Thus, nodes at higher levels of the tree correspond to coarser time scales, whereas nodes at lower levels correspond to finer time scales. The multipliers (weights) assigned to each descendant of a node can be set to represent a specific marginal distribution and scaling. In one of these, the Multifractal Wavelet Model (MWM) [25], multipliers are multiplicative innovations, generating approximately a log-normal marginal distribution. Such models require the setting of $2 + \log_2 N$ parameters, where N is the sample size. The major drawback of MWM is the number of parameters to be fitted and the need to construct a multiscaling binary tree, which is not suitable for on-line characterization. A maximum time scale for queues fed by MWM has also been derived.

Recent investigation [26] on small time scales of Internet traffic suggests that monofractal behavior is observed for these scales. It is claimed that correlations for small time scales are caused mainly by flows

with bursts of densely clustered packets, rather than the acknowledgement mechanism of TCP. However, the investigations based on publicly available traces used for this paper clearly show multifractal behavior for these scales.

Another line of work related to what is proposed in this paper is traffic control based on measurements. In [27], the time scale for which losses most probably occur, the (dominant) time scale of interest is computed using a hybrid measurement/analytical approach. The time scale is determined by observing a virtual queue with a smaller capacity than that of the one under study. In [28], a maximum rate envelope process was used to characterize the arrival as well as the service rate, these envelopes are derived via measurement. The maximum time scale can be derived by using these two measures.

8. Conclusions

This paper has introduced a novel traffic model, called the mBm envelope process, for the modeling of multifractal flows. This model has been validated using both synthetic and real network traffic and has been shown to be quite accurate. An algorithm for computing the equivalent bandwidth of a multifractal stream has also been presented. Such an algorithm can be used for real time estimation in admission control mechanisms. It has been shown that the mean value of the Holder function computed up to the maximum time scale of interest determines the gain obtained by statistical multiplexing of multifractal flows. This paper has also presented a special policing mechanism, an extension of the mechanism introduced in [9], appropriate for multifractal traffic.

Currently, a comparison between the modeling capacity of the envelope process and of the models based on wavelets is under study. It is suggested that the effectiveness of the envelope process when used in measurement-based frameworks be assessed.

Acknowledgments

This work was partially sponsored by FAPESP (Process 00/09772-6) and by CNPq (Process 305076/2003-5).

Appendix A

In this appendix, a proof for the theorem which limits the computational complexity of the algorithm for flow equivalent bandwidth computation is provided.

Theorem 1. *The equivalent bandwidth of an aggregate of N flows can be computed in at most*

$$n = \mathrm{O}(\log(C)) \tag{A.1}$$

iterations, where C is the channel capacity in bytes/s.

Proof. In the first step, the estimated bandwidth value is doubled at each iteration. The initial value is $\bar{a} + \epsilon$ and the maximum possible value is C. Thus, the first step will require a maximum number of iterations corresponding to the following:

$$n = \lceil \log(C - (\bar{a} + \epsilon)) \rceil = \mathrm{O}(\log(C)). \tag{A.2}$$

In the second step, a Binary Search is made in the interval $(\hat{C}/2, \hat{C}]$. It is known that a Binary Search in an interval $[a, b]$ takes at most

$$k = \lceil \log(m) \rceil \tag{A.3}$$

iterations [29], where m is the width of the interval. In this case

$$m = \hat{C}/2.$$

Thus,

$$n = \lceil \log(\hat{C}/2) \rceil = \lceil \log(\hat{C}) - 1 \rceil = \mathrm{O}(\log(C)).$$

The complexity of the algorithm is

$$\mathrm{O}(\log(C)) + \mathrm{O}(\log(C)) = \mathrm{O}(\log(C)). \qquad \square \tag{A.4}$$

References

[1] W. Leland, M. Taqqu, W. Willinger, D. Wilson, On the self-similar nature of ethernet traffic (extended version), IEEE/ACM Trans. Network. 2 (1) (1994) 1–15.

[2] A. Erramilli, O. Narayan, A. Neidhart, I. Saniee, Multi-scaling models of TCP/IP and sub-frame VBR video traffic, J. Commun. Networks 3 (4) (2001) 383–395.

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

[4] N.G. Duffield, 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.

[5] G. de Veciana, G. Kesidis, J.C. Walrand, Resource management in wide-area ATM networks using effective bandwidths, IEEE J. Select. Areas Commun. 13 (6) (1995) 1081–1090.

[6] F.P. Kelly, Notes on effective bandwidths, in: Stochastic Networks: Theory and Applications, Oxford University Press, Oxford, 1996, pp. 141–168.

[7] I. Norros, On the use of fractional brownian motion in the theory of connectionless networks, IEEE J. Select. Areas Commun. 13 (6) (1995) 953–962.

[8] C. Courcoubetis, V.A. Siris, G. Stamoulis, Application of the many sources asymptotic and effective bandwidths to traffic engineering, Telecommun. Syst. 12 (2–3) (1999) 167–191.

[9] N.L. Fonseca, G.S. Mayor, C.A.V. Neto, On the equivalent bandwidth of self-similar source, ACM Trans. Model. Comp. Simul. 10 (2) (2000) 104–124.

[10] C. Courcoubetis, V.A. Siris, Procedures and tools for analysis of network traffic measurements, Perform. Eval. 48 (1–4) (2002) 5–23.

[11] S. Molnar, T.D. Dang, I. Maricza, On the queue tail asymptotics for general multifractal traffic, in: Proceedings of the Second International IFIP-TC6 Networking Conference on Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; and Mobile and Wireless Communications, Springer, Berlin, 2002, pp. 105–116.

[12] R. Peltier, J.L.Vehel, Multifractional brownian motion: definition and preliminary results, Technical report, INRIA, 1995.

[13] G.S. Mayor, Performance Modeling and network management for self-similar traffic, Ph.D. thesis, University of Southern California, 1997.

[14] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, D. Veitch, The multiscale nature of network traffic: discovery, analysis and modelling, IEEE Signal Process. Mag. 19 (3) (2002) 28–46.

[15] J.E. Cavanaugh, Y. Wang, J.W. Davis, Self-similar processes and their wavelet analysis, in: Handbook of Statistics, Vol. 21: Stochastic Processes: Modeling and Simulation, Elsevier, Amsterdam, 2003.

[16] Wand Network Research Group, Internet Site, July 2003, The University of Waikato. Available from: <http://wand.cs.waikato.ac.nz>.

[17] Lawrence Berkeley National Laboratory, Internet Site, July 2003. Available from: <http://ita.ee.lbl.gov/html/contrib/dec-pkt.html>.

[18] P. Abry, P. Flandrin, M.S. Taqqu, D. Veitch, Wavelets for the analysis, estimation and synthesis of scaling data, in: Self-Similar Network Traffic and Performance Evaluation, New York, 2000.

[19] D. Veitch, D. Veitch homepage, July 2003. Available from: <http://www.cubinlab.ee.mu.oz.au/~darryl/>.

[20] V.E. Benes, General Stochastic Process in the Theory of Queues, Addison-Wesley, Reading, MA, 1962.
[21] I. Norros, A storage model with self-similar input, Queue. Syst. 18 (1994) 387–396.
[22] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, Berlin, 1999.
[23] K. Sohraby, M. Sidi, On the performance of bursty and modulated sources subject to leaky bucket rate-based access control schemes, IEEE Trans. Commun. 42 (3) (1994) 477–487.
[24] R.H. Riedi, M.S. Crouse, V.J. Ribeiro, R.G. Baraniuk, A multifractal wavelet model with application to network traffic, IEEE Trans. Inform. Theory 45 (4) (1999) 992–1018.
[25] V.J. Ribeiro, R.H. Riedi, M.S. Crouse, R.G. Baraniuk, Multiscale queuing analysis of long-range-dependent network traffic, in: Proceeding of IEEE Infocom, 2000, pp. 1026–1035.
[26] Z.L. Zhang, V. Ribeiro, S. Moon, C. Diot, Small-time scaling behaviors of internet backbone traffic: an empirical study, in: Proceeding of IEEE Infocom, San Francisco, March 2003.
[27] D.Y. Eun, N.B. Shroff, A measurement-analytic approach for QoS estimation in a network based on the dominant time scale, IEEE/ACM Trans. Network. 11 (2) (2003) 222–235.
[28] J. Qiu, E. Knightly, Measurement-based admission control using aggregate traffic envelopes, IEEE/ACM Trans. Network. 9 (2) (2001) 199–210.
[29] U. Manber, Introduction to Algorithms: A Creative Approach, Addison-Wesley Longman, Reading, MA, 1989.

Cesar Augusto Viana Melo received his B.Sc. degree in Computer Science from the University of Amazonas and his M.Sc. with honours (1999) and the Ph.D. (2004) degrees in Computer Science from the State University of Campinas, Campinas, Brazil. His M.Sc. dissertation was awarded at the 2000 Latin America Thesis and Dissertation Contest sponsored CLEI/UNESCO. His areas of interest are traffic modeling and traffic control.

Nelson Luís Saldanha da Fonseca received the Electrical Engineering degree (1984) and the M.Sc. in CS (1987) degree from Pontificial Catholic University at Rio de Janeiro and M.Sc. (1993) and Ph.D. (1994) degrees in Computer Engineering from the University of Southern California. He is currently an Associate Professor at the Institute of Computing, State University of Campinas. He has published 100+ papers and contributed to three books. He received the Elsevier Computer Networks Editor-of-Year-2001 Award. He is an active member of the IEEE Communications Society (ComSoc). He served as ComSoc Director of Online Services (2002–2003) and Editor for the Global Communications Newsletter (1999–2002). He is currently the Editor for Computer Networks, ComSoc e-news, an Area Editor for IEEE Transactions on Multimedia, a Technical Editor for IEEE Communications Magazine, and editor for IEEE Communications Surveys and Tutorials. He co-chaired the IEEE Globecom 2004 General Symposium, the IEEE ICC2004 QoS and Performance Modeling Symposium, the Communication QoS, the IEEE ICC03 Reliability and Performance Modeling Symposium), the International Teletraffic Congress 2001, the IEEE Globecom99 Multimedia Technology & Services Symposium and IEEE Workshop on Computer Aided Modeling Design and Analysis of Communications Links and Networks (2000 and 1998). His research interests include Traffic Modeling and Control, Multimedia Service and Quality of Service Provisioning.