SELF-SIMILAR LAN/WAN TRAFFIC MODEL: A SURVEY *

Anibal D. Angulo Miranda

Instituto Tecnológico de Aeronaútica (ITA)
Electronic Engineering – Department of Telecommunications
CTA, S.J. dos Campos 12228-900, SP, Brazil

ABSTRACT

Results that can be considerate as classicals from self-similar model paradigm are showed in this work. All measurements and results presented here were based on data traffic traces gathered in our network gateway using a simple yet reliable and efficient method. Several statistical methods, to infer an unbiased and well defined Hurst parameter, are also reviewed. Some important queuing performance measurements with both fractal and Markovian-type models are also presented.

1. INTRODUCTION

This work presents a survey, in an easy-to-read way, from those results obtained in the few past years. As it is vastly known, self-similarity has turned a "classical" and well established concept inside the network performance and traffic engineering areas. It is well-known that self-similarity matches, in a parsimonious way, the actual statistical data traffic behavior, and it seems that nobody could argue against it. Thus, self-similarity or its concomitant long-range dependence characteristic in computer networks present a fundamentally different set of problems to people doing analysis and/or design of networks, and many of the previous assumptions upon which systems have been built could not be longer valid in the presence of this fractal behavior. This feature has added a new dimension to the teletraffic modeling and analysis world. The aggregated IP datagrams traces to be analyzed were obtained in our network. The method to get it is simple yet reliable. Basically it uses the Simple Network Management Protocol (SNMP) and could be used either at network interface card (NIC layer two protocol to get Byte streams) or IP network level (layer three protocol to get datagrams).

The impact of this fractal characteristic to modern high-speed networks, like Asynchronous Transfer Mode (ATM) networks in a general sense, is reviewed. The paper is outlined as follows. Section 2 explains the method that has been used to gather data traffic from our network. Section 3 gives a briefing about the Hurst parameter. Section 4 shows a time-frequency analysis performed over the collected data. In order to estimate the Hurst parameter, Section 5 outlines six well-known methods. Section 6 explains the fractional Gaussian noise process. Some important performance measurement results from a simulation scenario built for our purposes are presented in Section 7, these performance tests have been done either with fractal and Markovian-type models in order to show their main differences. Finally, Section 8 presents the major conclusions of this work.

*This work is supported by the International Civil Aviation Organization (ICAO), grant project ITA-BRA006/95/802/805.

2. TRAFFIC SAMPLES GATHERING

Fig. 1 shows the method used to gather the aggregated Internet (IP) data traffic, as this figure shows, the Network Control Center (NCC) is the system manager that keeps tracking IP datagrams traffic through the gateway. The method is quite simple yet reliable. Basically, the SNMP software [6] was used. The agent information, obtained via *SetRequest/GetResponse* primitive functions to its Management Information Base (MIB) [6], was stored to be processed later on.

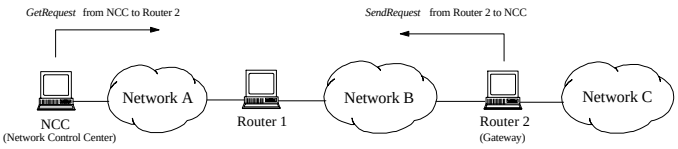


Figure 1. Method used to gather the aggregated Internet traffic.

Fig. 2 shows one time series of aggregated IP traffic collected using the method above. In order to infer self-similarity, many Hurst parameter estimation methods were applied over such huge time series sets.

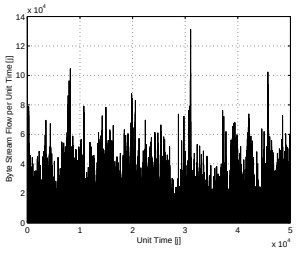


Figure 2. Time series constituted of aggregated Byte streams.

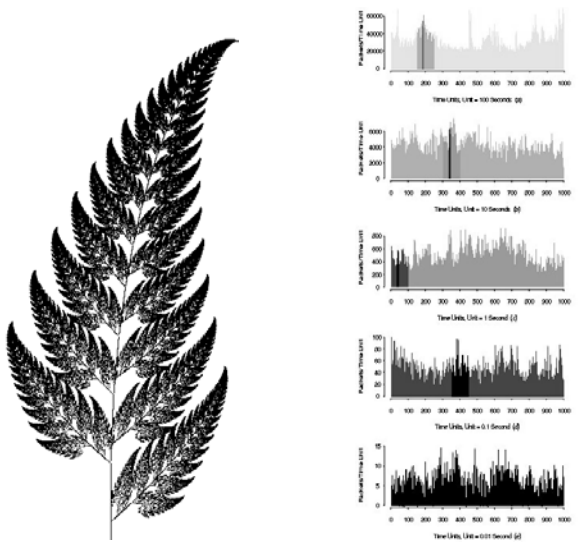
Among self-similar stochastic processes, there exit either deterministic and statistical ones. Deterministic self-similarity implies in a similar (or equal) structure whatever the observational scale is. Now, stochastic self-similarity appears only in a statistical context, that is, whatever the observational scale is, the statistical properties (at least up to 2^{nd} order) are quite similar and are degenerated in a very slow way, for this reason they are also called as *slow stochastic processes*. Fig. 3(a) shows a *fern leaf* that belongs to those deterministic self-similar objects, in this case, "*space*" constitutes the observational scale, as this figure shows, the whole fern leaf differs of their branches only by a linear scale factor. Fig. 3(b) shows the LAN traffic collected at Bellcore (now Telcordia) between August 1989 and February 1992 $[1-3]$, represents a statistical self-similar process, but now, "*time*" constitutes the observational scale. This scale independence characteristic is the main feature of fractal processes, as fig.3 shows, such processes looks almost the same for many increasing or decreasing observational scales. Therefore, self-similarity is a general intrinsic concept of those fractal processes.

3. THE HURST PARAMETER

The Hurst[1] parameter, denoted by H, represents the degree of self-similarity in any time series; it measures the variability around the expected value of studied data. Mathematically speaking, this parameter is a fractional number: $0 < H < 1$. Physically, this parameter has the following interpretation:

- If $0 < H < \frac{1}{2}$, this is a singular case [7], the process shows the *anti-persistence phenomena*, further the process reveals a negative auto-correlation function, sometimes called as *negative dependence*. Even though this region for practical purposes has not been used at all, some work was made on [36].
- If $H = \frac{1}{2}$, the process presents random characteristics and it is highly uncorrelated, inside the classical time series theory this value is very representative of those autoregressive and Markovian-type processes, all of them present a short-range dependence (SRD) time characteristic, that is, are memoryless processes. For this value of H, autocorrelation function shows a fast exponential decaying, its statistical behavior resembles a pure second order white noise turning such processes statistical predictable. SRD processes have a well-developed queuing theory but if applied to modern high speed networks this classical theory could gives false optimistic performance.
- If $\frac{1}{2} < H < 1$, the process is self-similar and shows the well-known *time persistence phenomena*, thus the process is highly positively correlated [7], its correlation function reveals an hyperbolic decaying, that is, this feature turns such process as having some sort of memory, in other words, it is a long-range dependence (LRD) process. Lately, processes with an H parameter lying into this interval are being used to model the statistical LAN/WAN traffic behavior. Therefore, self-similarity and LRD are physical and mathematical concepts to explain, in a parsimony way, the highly dynamics observed in LAN/WAN traffic over many time scales.

[1] In the mid-1960, hydrologist Harold Edwin Hurst found an unexpected behavior (in many geophysical time series throughout the world, overall when studied river Nyle's annual stream-flow volumes) differing from the classical theory behavior. Nowadays, this discovery is known as the *Hurst effect* or *Hurst phenomena*. For example, river Rhine in Germany, has an $H \approx 0.5$.



(a) (b)

Figure 3. Self-similar objects: (a) A (deterministic) fern leaf and (b) (statistical) LAN traffic trace. *Copyrigth of "Transactions On Networking IEEE/ACM".*

Methods to estimate the Hurst parameter take full advantage of the so-called *aggregated process*, as explained below.

3.1 The Aggregated Process

An appropriate definition, inside the context of classical time series theory [3, 8], involves an aggregated stationary sequence with an aggregation level equals m, given by $X_j^{(m)}$. This new time series is obtained by blocking the original stationary sequence, X_j, over j non-overlapping blocks of size m, each block is averaged (X_j is the number of IP packets arriving in the *j-th* time interval). Basically, aggregation and averaging tend to smooth the structure of the original time series, X_j. This new sequence, for $m \geq 1$, stills stationary and it is known as the aggregated process, mathematically is given by

$$X_j^{(m)} = \frac{1}{m}\Big[\sum_{i=(j-1)m+1}^{jm} X_i \Big] \qquad (1)$$

Note that the aggregated process concept is quite different from that of aggregated traffic concept, the latter is the original time series, X_j, the former has to be further processed to become the aggregated process, $X_j^{(m)}$, itself.

3.2 Aggregated Traffic and Gaussian Processes

A mathematical model that has been used [3, 12, 13] to characterize aggregated IP datagrams traffic is the fractional Gaussian noise[2] (fGn) model, which is derived of the fractional Brownian motion (fBm), this process is the only exactly second-order self-similar stochastic Gaussian process. The self-similar characteristic observed in aggregated LAN/WAN traffic can be emulated in a parsimony way through this model keeping the familiar Gaussian distributions. In order to generate fBm traces the *random midpoint displacement* algorithm [10, 11] was used. To this end, one needs only three parameters, namely, the mean value m, the variance value between samples given by $\sigma^2 t^{2H}$ and the Hurst parameter, H, for $H = 0.5$ one has the ordinary Brownian (*random walk*) process. All those features makes the fBm one parsimonious model. Although it was observed that aggregated traffic is less Gaussian than expected [31], for practical issues the fBm increments [12] can be a good model to work with.

A process that has been derived from any self-similar process through increments is known as a fractional noise. Furthermore, if its probability density function (PDF) has a characteristic exponent, $\alpha \approx 2$, then one is faced with the fGn [7, 8, 27]. A stable random variable X with index α is called α-*stable*, this parameter deals with the PDF tail behavior, it controls the burtiness of the random variable. Basically, regardless of the behavior of the distribution for small values of the random variable, if the asymptotic shape of the distribution is hyperbolic it is heavy-tailed, mathematically, if $P(X > x) \propto x^{-\alpha}$ as $x \to \infty$ and $0 < \alpha < 2$, then the marginal distribution has a heavy-tailed behavior. The simplest heavy-tailed distribution is the *Pareto* distribution, this distribution is hyperbolic over its entire range and it is useful to represent LAN/WAN traffic network distribution functions regarding the sojourn time sessions and file sizes on Web servers [3, 16, 17].

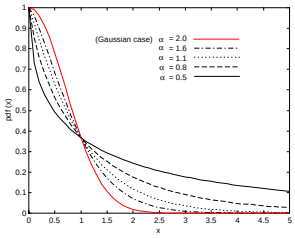


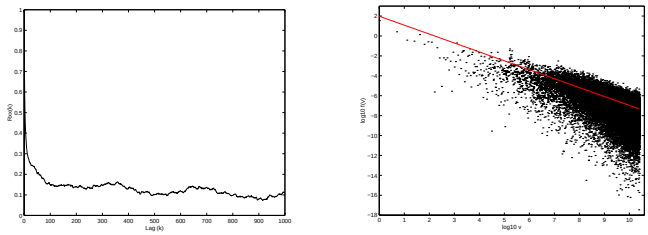
Figure 4. Symmetric stable distributions with varying α, as $\alpha \to 0$ the distributions tails get heavier.

Fig.4 shows the heaviness of the tails of such distributions where a set of symmetric standard stable distributions is given with α varying from 2 (the Gaussian case) to 0.5 (non-Gaussian cases). The mathematical relationship between α and H, is given by $H = (3 - \alpha)/2$ [3, 32].

[2]The fGn has a colored noise spectral behavior $(1/f^{\alpha})$, the color chosen for this process was the *brown* color, note that the fGn has nothing to do with such a color. It was so in honor to the research work made by English botanist Robert Brown (1827).

4. TIME AND FREQUENCY ANALYSIS

Fig.5(a) reveals the intrinsically LRD of gathered IP traffic. The hyperbolic decaying rather than classical exponential (that SRD processes present) is clear. This behavior is also known as the *persistence phenomena*. When a spectral analysis is performed, due reciprocal relationship between time and frequency, such LRD characteristic is reflected at lowest frequencies, that is, the spectral behavior blows up following a power law, as $\nu \to 0$, thus the process has a singularity (*a pole*) near the origen. Moreover, it could be seen as an $1/f^{\alpha}$ noise. Fig.5(b) shows this typical spectral behavior of mono fractal processes like some colored noises (i.e., pink, red, brown) and music [3, 15, 18] among many others physical processes.



(a) (b)

Figure 5. Gathered IP data traffic analysis: (a) Autocorrelation function (b) Spectral analysis.

5. ESTIMATION METHODS

Here, we outline six commonly and vastly used methods to estimate the self-similar Hurst parameter. The average value obtained by these methods will be used to generate synthetically fGn traces to emulate real LAN/WAN traffic traces.

5.1 The R/S Statistic

This method [3, 8, 9], infers the degree of self-similarity (via the Hurst effect) of any stochastic process. Its relative robustness against changes in the marginal distribution of the underlying process turns it particularly attractive. Briefly, for a given set of observations, X_k; $k = 1, \ldots, n$, this set is subdivided into K non-overlapping blocks as in eq.(1). One computes the rescaled adjusted range $R(t_i, n)/S(t_i, n)$ for a number of values n, where $t_i = \lfloor n/K \rfloor (i-1) + 1$ are the starting points of the blocks which satisfy $(t_i - 1) + d \leq n$. Therefore

$$R(t_i, n) = max\{0, W(t_i, 1), \cdots, W(t_i, n)\} - \\ - min\{0, W(t_i, 1), \cdots, W(t_i, n)\} \qquad (2)$$

Where

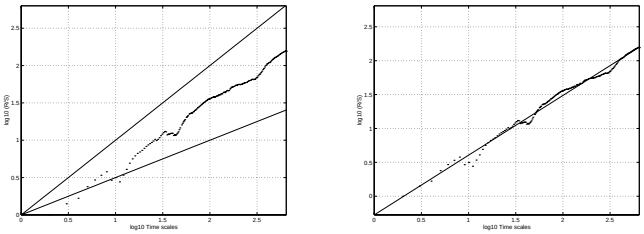
$$W(t_i, k) = \sum_{j=1}^{k} X_{t_i+j-1} - k\Big(\frac{1}{n}\sum_{j=1}^{k} X_{t_i+j-1}\Big) \qquad (3)$$

For each value of n one obtains a number of R/S samples. For small values of n there are K samples. The number decreases

for large values of n because of limiting condition on the t_i values mentioned above. According to this method, many natural occurring time series appear to be well represented by the following asymptotical law[3], as $n \to \infty$

$$E\Big[\frac{R(n)}{S(n)}\Big] = c\,n^H \tag{4}$$

Where n is the sample size, c is a positive constant not depending of n and $S(n)$ is the sample variance of $X_j^{(m)}$ (eq.(5)). Recall that, in order to show self-similarity, H, must to be well-defined in the interval $0.5 < H < 1$. On the other hand, observations X_k from a SRD process obey the relationship given by the law proportional to $n^{0.5}$, this discrepancy is currently referred as the *Hurst effect* or *Hurst phenomena*. Therefore, the law given by Einstein (1905) as $n^{0.5}$, was generalized by Hurst (1954) to n^H with $0 < H < 1$ (see Section 3).



(a) (b)

Figure 6. (a) The R/S statistic (b) Linear regression to estimate the Hurst parameter.

Plotting $\log E[R(t_i,n)/S(t_i,n)]$ vs. $\log(n)$ results in what is known a *pox diagram*. Next, a least squares line is fitted to the points of the R/S plot. The R/S samples corresponding to the smallest values of n are dominated by short-range correlations. Samples of large values of n are statistically insignificant if the number of samples per n is less than, say, 5. The slope of the regression line for these R/S samples is an estimate for H (eq.(4)). Fig.6(a) makes possible to show a well-defined Hurst parameter value. Fig.6(b) shows the linear regression line fitted, given the following equation, $y = 0.89906x - 0.32998$, therefore $H = 0.899$.

5.2 The Variance-time Analysis

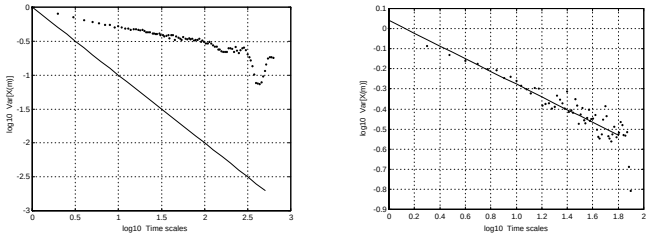
Regarding time series analysis [8, 26], the sample variance, $S(n) = Var[X_j^{(m)}]$, of the aggregated process (eq.(1)) can be written as

$$Var[X_j^{(m)}] = \frac{\sigma^2}{m}\big[1 + 2\sum_{k=1}^{m}\big(1 - \frac{k}{m}\big)r(k)\big] \tag{5}$$

Where $r(k)$ is the auto-correlation function of the aggregated process. For LRD processes $\sum r(k) \to \infty$ is verified, such feature has never been observed in SRD processes. All SRD processes, have an exponential decreasing. Regarding eq.(5), for SRD processes, the autocorrelation function vanishes at larger time scales and in this case $Var[X_j^{(m)}] = \sigma^2/m$, that is, it vanishes as fast as m^{-1}, this is a well-known classical time series theory result [8]. On the other hand, LRD processes have an hyperbolically decaying in their auto-correlation functions (Fig.5(a)), thus $Var[X_j^{(m)}]$ obeys the asymptotically relationship [3, 8, 26], as $m \to \infty$

$$Var[X_j^{(m)}] \approx c\,m^{-\beta} \tag{6}$$

Here, m indicates the aggregation level of the averaged process, c is a constant not depending of m, and $0 < \beta < 1$. Therefore, $Var[X_j^{(m)}]$ decays linearly, as $m \to \infty$, in a log-log plot, with a slope flatter than -1 as depicted in Fig.7(a). This figure is known as a *variance-time plot*, such plots are useful to determine the degree of the asymptotic LRD behavior. The variance-time plots are obtained plotting $\log Var[X_j^{(m)}]$ vs. $\log(m)$. In order to estimate β, the slope of the least squares fitted line is used, from which the Hurst parameter can be determined using $H = 1 - \frac{|\beta|}{2}$.



(a) (b)

Figure 7. (a) The variance-time plot (b) Linear regression to estimate the Hurst parameter.

Usually the points with small values of m (the transient period) and the few largest values of m (not accurate for finite sequence) are ignored. Fig.7(b) shows the estimation procedure, it gives, $y = -0.31679x + 5.7962 \Rightarrow \beta = -0.31679$, which in turn implies in $H = 0.8416$.

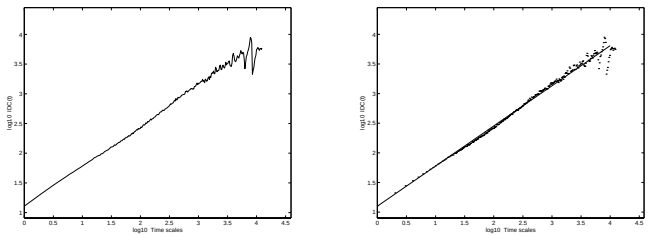
5.3 The Index of Dispersion (for Counts) Analysis

The index of dispersion for counts (IDC) is given by the variance of the number of arrivals during a interval of time t divided by the expected value of that same quantity. Fig.8(a) shows the IDC analysis, this figure reveals the monotonically linear behavior over many time scales, such behavior obeys the following empirical mathematical law [3]

$$IDC(t) = \frac{Var\Big(\sum_{j=1}^{t} X_j^{(m)}\Big)}{E\Big(\sum_{j=1}^{t} X_j^{(m)}\Big)} \approx c\,t^{2H-1} \tag{7}$$

Where c is a finite positive constant independent of t. As eq.(7) shows, a log-log diagram results in a straight line with a slope equals $2H - 1$. Therefore, making a linear regression one computes the following equation: $y = 0.67741x + 1.1002$, equating $2H - 1 = 0.67741$ one obtains $H = 0.838$.

[3]This relationship is the actual definition of fractional Brownian movement, if $H = 0.5$ the process becomes the ordinary Brownian motion. In 1905, Einstein used the entropy of an ideal gas to explain the Brownian motion of a pollen grain suspended in water.



(a) (b)

Figure 8. (a) The IDC analysis (b) Linear regression to estimate the Hurst parameter.

Nowadays, according with Section 3.2, peaked heavy tailed processes, that is, processes having an $IDC > 1$ are useful to represent the time duration of any point-to-point LAN/WAN connection [4, 20] (i.e. the communication between any pair of end workstations). In other words, the sojourn time in ON (busy) state and/or OFF (idle) state must to be drawn from a random variable with a heavy-tail marginal distribution [16, 21, 32]. It has been mathematically proven that when n end-to-end (ON-OFF) source connections like telnet, ftp, http[4] etc. are aggregate, in the limit when $n \to \infty$, we can fairly approach an fGn [21] process, which will be the buffer input process of any edge router.

5.4 The Periodogram Analysis

This method [3, 8] reveals the periodogram behavior near the origin. Using the estimate periodogram based on Fourier technic, we have

$$I(\nu) = \frac{1}{2\pi N}\left|\sum_{j=1}^{N} X_j e^{ij\nu}\right|^2 \tag{8}$$

Where X_j is the time series of length N and $0 < \nu < \pi$ is the frequency. If the time series intrinsically has long-range dependence its spectral density will be proportional to $|\nu|^{1-2H}$, as $\nu \to 0$ (eq.(9)). A log-log regression thus provides an estimate of H. Fig.5(b) shows such analysis, it is clear the power law $(1/f^{\alpha})$ topology, as $\nu \to 0$. The straight-line slope is an estimate of $1 - 2H$, hence $H = 0.876$.

5.5 The Maximum Likelihood Estimator of Whittle

The maximum likelihood estimator (MLE) of Whittle is used to estimate the self-similar, H parameter, [23]. This MLE provides asymptotically consistent and normally distributed estimators of the unknown parameters of both Gaussian and non-Gaussian time series. One important point that arises with this method is to associate a topological structure for the power spectral function to those time series obtained. If time series from Fig.2 is considerate as an fGn, then its spectral behavior [14, 24] is given asymptotically, as $\nu \to 0$, by

$$f(\nu) \to C_H |\nu|^{1-2H} \tag{9}$$

[4]High layers (end-to-end) Internet communication protocols.

Where C_H[5] is a constant [24] and ν is the frequency. The Whittle estimator is the value of vector η which minimizes $Q(\eta)$. Thus, given a data sample of size N, the estimation process essentially involves minimizing the discrete version of the MLE of Whittle

$$Q(\eta) = \sum_{j=1}^{M} \frac{I(\nu_j)}{f(\nu_j, \eta)} \tag{10}$$

Dealing with fGn implies that η is simply the parameter H, [14]. M is the integer part of $(N-1)/2$, ν_j are the usable Fourier frequencies $(\nu_j = 2\pi j/N)$, $j = 1, 2, \ldots, [(N-1)/2]$ and $I(\nu_j)$ is the periodogram function as given by eq.(8). Replacing $f(\nu)$ (eq.(9)), by its estimate and taking logarithms, the function to be minimized is $R(H) = Q(H) - 1$, hence

$$R(H) = \log\left[\frac{1}{M}\sum_{j=1}^{M}\frac{I(\nu_j)}{\nu_j^{1-2H}}\right] - (2H-1)\frac{1}{M}\sum_{j=1}^{M}\log(\nu_j) \tag{11}$$

After a properly data normalization, Fig.9 shows the minimization procedure. Recall that such time series was forced to obey the fGn asymptotically spectral behavior, as $\nu \to 0$. Confidence intervals can be found as $1.96 \cdot \sigma$, where σ is

$$\sigma^2 = 4\pi\frac{R(\widehat{H})}{N_\nu} \tag{12}$$

As depicted in Fig.9, the result using the MLE of Whittle minimized eq.(11) for $H = 0.861 \pm 0.051$. One variation of the MLE of Whittle is the so-called MLE of Whittle local, this is a semi-parametric estimator in that it only specifies the parametric form of the spectral density when $\nu \to 0$, this estimator is also based on the periodogram (eq.(8)) [14, 24].

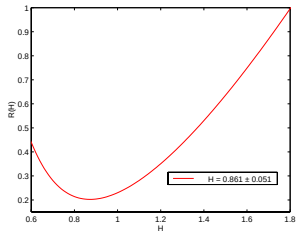


Figure 9. The MLE of Whittle applied on IP traffic collected.

5.6 The Wavelet-based Estimator

Wavelet analysis is proving to be a very powerful tool for characterizing behavior, especially self-similar behavior, over a wide range of time scales. Wavelets basis functions are self-similar themselves, thus a natural approach for analyzing and synthesizing random fractals. As stated in [33, 34], this method is based on what is known the Multi Resolution Analysis (MRA), or its concomitant Discrete Wavelet transform (DWT), which consists of splitting a time series (signal) into a (low-pass) approximation and (high-pass) details, that is, the key idea of the MRA consists in examining the loss

[5]$C_H = [2/(2H+1) + 1/(H+2) - 2/(H+1)]$.

of information (details) when going from one approximation to the next one, and so on. Basically, the MRA consists in rewriting the information of any dataset as a collection of details at different resolutions and a low-resolution approximation. These are associated with the coefficients a_x and d_x, respectively. Denoting as $x(t)$ the time series from Fig.2, then $x(t) = \text{approx}_j(t) + \sum_{j=1}^{j=J} \text{detail}_j(t)$ [33], therefore

$$x(t) = \sum_k a_x(J,k)\phi_{J,k}(t) + \sum_{j=1}^{J}\sum_k d_x(J,k)\Psi_{J,k}(t) \qquad (13)$$

The term "$\text{approx}_j(t)$" essentially being coarser and coarser approximation of $x(t)$ means that ϕ_0 needs to be a low-pass function. The term "$\text{detail}_j(t)$", is like a "differential" information, indicates rather that Ψ_0 is a band-pass function, and therefore a small wave, hence a *wavelet*. The parameters $\phi_{j,k}(t) = 2^{-j/2}\phi_0(2^{-j}t - k)$ and $\Psi_{j,k}(t) = 2^{-j/2}\Psi_0(2^{-j}t - k)$ are the set of shifted and dilated functions of the scaling function ϕ_0 and the mother-wavelet Ψ_0. Therefore, given a scaling function ϕ_0 and a mother-wavelet Ψ_0, the DWT is a mapping from $L^2(\mathcal{R}) \rightarrow l^2(\mathcal{Z})$ given by $a_x(J,k)$, $k \in Z$ and $d_x(j,k)$, $j = 1,\ldots,J$; $k \in Z$ as defined by the inner products $a_x(J,k) = \langle x, \phi_{j,k}(t)\rangle$ and $d_x(j,k) = \langle x, \Psi_{j,k}(t)\rangle$. Both coefficients, a_x and d_x are located on a dyadic, the details given by these coefficients are important. When going from high resolution to lower resolution, the MRA gives rise to details at larger time scales. This can be interpreted in the frequency domain as band-pass filtering, going from high to low frequencies with constant relative bandwidth. On the other hand, spectral estimators (based on periodograms) may easily get strongly biased due to the fact that constant bandwidth mismatches the power-spectrum to be analyzed [35]. In contrast, the wavelet constant relative bandwidth manages to provide a perfect match. Given that the wavelet method has some connections to the variance plot. Recalling the variance expression when studying aggregated series $X_j^{(m)}$ over dyadic blocks of size 2^j, it means that: $Var[\text{detail}_j] \approx 2^{j(2H-2)}$. This estimator could be very biased (not very reliable). In the wavelet framework is necessary to study differences of aggregated series. Regarding the simplest case one computes the difference between points in non-overlapping blocks of size 2 as defined by the Haar wavelet. Consider now Y^{j+1} as being the series made by difference Y^j, note that Y^0 is the data series at highest time resolution.

$$Y_k^{j+1} = 2^{-\frac{1}{2}} \cdot (Y_{2k}^j - Y_{2k-1}^j) \qquad (14)$$

Eq.(14) has $k = 1,2,\ldots,N/2j$ and $j = 1,2,\ldots$. The variance of Y^j decay according to a similar power-law as above

$$[Y^j] \approx 2^{j(2H-1)} \qquad (15)$$

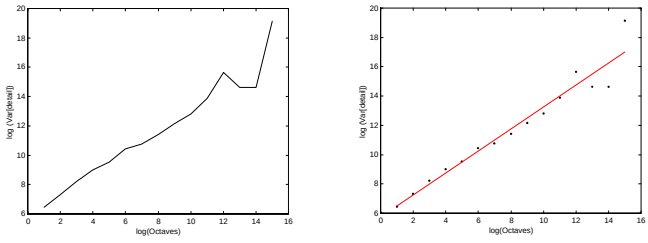
It turns out that the variance is equal to the second order moment. Since the expectation of Y is zero. In the frequency domain, the variance, $E[(Y^j)^2] = Var[Y^j]$, is equivalent to the signal energy in a frequency band depending on j. In resume, this method computes the DWT, averages the squares of the coefficients of the transform, and then performs a linear regression on the logarithm of the average, versus the log of j, the scale parameter of the transform. The result should be directly proportional to H, such linear relationship implies LRD. Then, taking logarithms in eq.(15) one obtains

$$\log_2[Var(\text{detail}_j)] = (2H-1)j + c \qquad (16)$$

Where c is a finite constant. This estimator has been proven to be unbiased under very general conditions and efficient under Gaussian assumptions. Then the wavelet estimator is very efficient. It is also possible to evaluate measurement statistics such as confidence interval, and weighted regression. Under Gaussian and quasi-decorrelation of the wavelet coefficient hypothesis and in the asymptotic limit, a closed-form for the variance of the estimate of H can be obtained and is given by [33] $\sigma_H^2 = Var\, H[j_1, j_2]$, which in turn can be written as

$$\sigma_H^2 = \left(\frac{2}{n_{j_1} ln^2 2}\right) \cdot \left(\frac{1 - 2^j}{1 - 2^{-(J+1)}(J^2+4) + 2^{-2J}}\right) \qquad (17)$$

Here, $n_{j_1} = n/2^{j_1}$ is the number of available coefficients at the lowest scale involved in the linear fit, and $J = j_1 - j_2 + 1$ is the number of scales (octaves) involved in the linear fit. Fig.10(a) shows the wavelet-based method applied over the collected IP datagrams and Fig.10(b) shows the linear regression equation adjusted over obtained points. This wavelet-based analysis gave the following Hurst value: $H = 0.876 \pm 0.044$.



(a) (b)

Figure 10. (a) The wavelet-based method (b) Linear regression to estimate the Hurst parameter.

So far, a neatly well-defined Hurst parameter has been obtained from IP datagrams traffic. Results are summarizes in Table 1. Next step is to generate synthetic fGn traces following real data traffic behavior. To this end, the averaged Hurst parameter value, $H = 0.862$, is used.

6. FRACTIONAL GAUSSIAN NOISE SAMPLES

The fGn process used as arrival traffic is represented by the following relationship [12]

$$A_t = mt + \sqrt{am}\, Y_t \qquad (18)$$

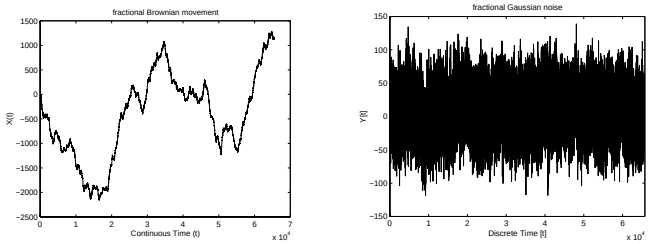
Where "m" is the mean rate and "a" is the *peakedness* factor defined as the ratio between the variance and the mean

Table 1. Hurst parameter estimation results.

Statistical Method	H
R/S Statistic	0.8992
Variance-time analysis	0.8416
Index of dispersion (for counts) analysis	0.8385
Periodogram-based analysis	0.8768
MLE of Whittle	0.861 ± 0.051
Wavelet-based estimator	0.876 ± 0.044
Averaged H value :	0,8621

value it is also known as the variance coefficient. This factor can be seen as the IDC of the inter-arrival process[6]. Therefore, A_t has its amount of traffic, m, that is offered to the network (*buffer*) in the interval $(0, t]$ while H and a characterize the type of traffic mix. The incremental process given by $Y_t = X_{t+1} - X_t$ in eq.(18) is a discrete time stationary process with X_{t+1} and X_t being samples of the fBm. As eq.(18) shows, A_t consists of one deterministic part, namely mt, that represents the average incoming traffic and one stochastic part, $\sqrt{am}Y_t$, that accounts for the random fluctuations in the traffic, recall that Y_t is bearing by H. This model works quite well if the number of the sources is huge and with not too high individual peak rates. Unfortunately it does not work equally well when there are only a few active sources or when there is high diversity among the speeds from the different sources. Fig.11(a) shows a sample trace of a non-stationary continuous time fBm process generated via the *Random Midpoint Displacement* (RMD) algorithm [10, 11], with $H = 0.862$ a mean value $m = 0$ and a variance between samples given by $\sigma^2 t^{2H}$. Fig.11(b) shows the derived fGn process[7]. The fGn asymptotical auto-correlation function [3, 7, 8, 24], as $k \to \infty$, is given by

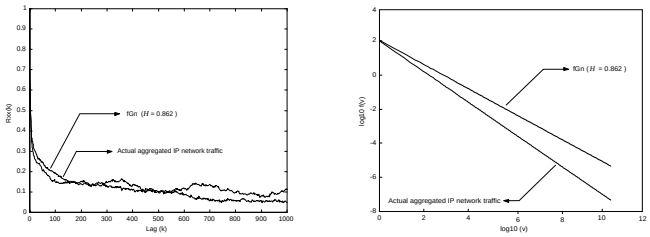
$$r(k) \to H(2H-1)k^{2H-2} \qquad (19)$$



(a) (b)

Figure 11. (a) fractional Brownian motion (fBm) sample trace (b) fractional Gaussian noise (fGn) sample trace.

Regarding the moving average and harmonic integral representation of the fGn rather than a direct Fourier transform for $r(k)$ instead [7, 27], one obtains the spectral density function (refer to eq.(9)) as an $1/f^{\alpha}$ noise. Fig.12(a) compares the auto-correlation functions from both fGn and aggregated IP network traffic processes and Fig.12(b) compares the spectral behavior, near the origin, of the same processes. From the above results it is possible to conclude that the time and frequencies characteristics, of the fGn are very closer to those actual aggregated IP network traffic traces, even though aggregate IP traffic seems to be less Gaussian than expected [31]. However, for practical purposes this model can be used to describe the highly dynamic, over many time scales, of the aggregated IP datagrams traffic collected.



(a) (b)

Figure 12. Actual IP datagrams traffic vs. fGn: (a) auto-correlation functions (b) Spectral behavior.

7. PERFORMANCE ANALYSIS

The scenario to obtain some measurements as cell loss rate (CLR), effective bandwidth and queuing performance via simulation is shown in Fig.13. This figure shows that the router and the ATM switch were represented by two queues. The ATM switch using the Enhanced Proportional Rate Control Algorithm (EPRCA) controls the Source End System (SES) bit rate as intended for Available Bit Rate (ABR) service [19].

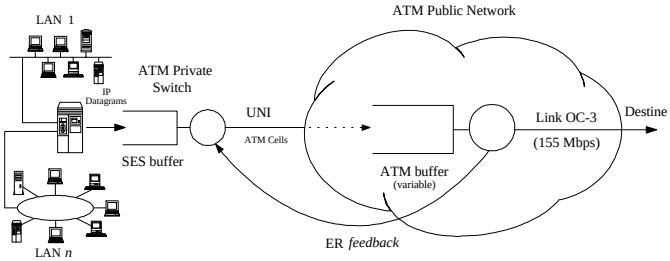


Figure 13. The simulation scenario.

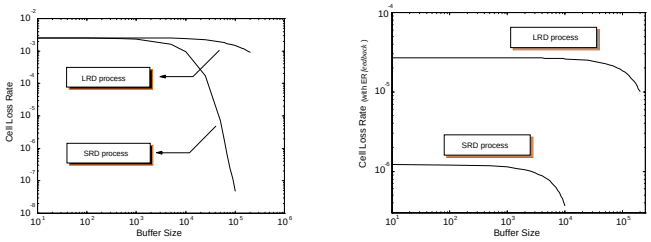
7.1 Cell Loss Rate

Every IP packet was segmented and enveloping in ATM cells utilizing the AAL5 protocol [22]. Fig.14(a) and Fig.14(b) show the CLR as a function of the ATM switch output buffer size (in cell units) without and with the presence of EPRCA respectively, more about this congestion control algorithm can be obtained in [19] and references therein.

Comparing the results obtained in the presence of Markovian and self-similar traffic, it is clear that in both cases the use of EPRCA algorithm reduce the CLR, but for self-similar traffic

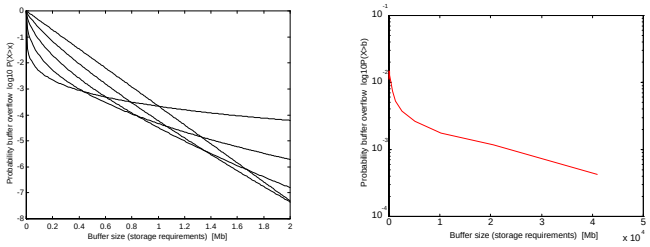
[6] For LRD processes $a \gg 1$ is mandatory.

[7] In order to use this model to simulate LAN/WAN traffic intensity, one has to truncate samples of the original fGn to eliminate negative values.



(a) (b)

Figure 14. Cell loss rate: (a) without explicit rate (ER) control (b) with ER control.



(a) (b)

Figure 15. Probability buffer overflow: (a) theoretical Norro's formula (b) simulation.

the CLR is still high even with large buffer size. Therefore, a solution should not be to get high buffer sizes into an ATM switch, but new congestion control avoidance algorithms instead, that takes into account the long-range dependencies of source/destination end systems.

7.2 Effective Bandwidth

The effective bandwidth of a given source is the desired bandwidth to meet the QoS requirements of the source. We will see that effective bandwidth and probability buffer overflow are intimately related concepts. As stated in [12], "(Fig.15(a)) shows the residual distribution function of the Weibull distribution with different values of H, m and a, the sharp knee between the cell and burst scale components of the queue length distribution, is replaced by continuous flattening, corresponding to the intuitive idea of burstiness in all time scales. In particular, the curves don't have non-zero asymptotic slopes when Hurst parameter is $H > 0.5$. Note that for $H = 0.5$ this probability buffer overflow behavior reduces to an exponential distribution function". Fig.15(b) shows the simulation result in agree with Norros formula.

The probability buffer overflow under fractal input processes, has a very complex analytical evaluation and only asymptotic of a lower bound of the complementary probability have been provided by many researches to date [12, 13, 28]. The so-called lower bound for approximate queue length distribution [12, 37] is given by

$$P(X > x) \approx \exp\left[-\frac{(C-m)^{2H}}{2\kappa(H)^2 a\, m} x^{2-2H} \right] \qquad (20)$$

Where C is the channel capacity, m is the mean rate of the aggregated IP data traffic, $\kappa(H) = H^H(1-H)^{1-H}$, a is the peakedness factor and x is the buffer size. CAC algorithms will be enormous beneficiate of all breakthroughs done over this subject. Eq.(20) gives a very important feature for modern high speed networks based on some QoS and LRD processes. In this case, if one solves eq.(20) for C given the QoS parameter as a constraining such as $P(X > x) = \xi$, then

$$C_{eff} = m + \left[-2\,\kappa(H)^2\, a\, m\, \ln(\xi)\, x^{-2(1-H)} \right]^{\frac{1}{2H}} \qquad (21)$$

Where C_{eff} is the expected or effective bandwidth that is needed in order to keep the probability buffer overflow small.

Current CAC algorithms are based on eq.(21) (with some variants) as a first step toward more complex algorithms, taking into account the fractal behavior.

Eq.(21) can be interpreted as follows: the bandwidth, C_{eff}, required to support a stream of traffic should be equal (at least) to its mean bit rate (this is the deterministic part of the equation) plus sufficient spare capacity to absorb traffic burst above the mean (the stochastic part). Note that under asymptotically second-order (Gaussian) self-similar arrival process assumptions the probability buffer overflow decays algebracally as $x^{-2(1-H)}$. But if a non-Gaussian arrival process is considerate, which corrsponds to $\alpha \neq 2$, then the algebraic decay term takes the form $x^{-\alpha(1-H)}$ leading to more realistic and complex frameworks.

Fig.16 depicts the practical use of eq.(21) as a link dimensioning formula; it is interesting to considerer its sensitivity on a and H, this figure shows a family curves with various values of a and H for $m = 2$ Mbps, $\xi = 10^{-3}$ for two buffer sizes. The upper three curves (right side of the figure) correspond to the buffer size 100 Kbytes, the lower three curves to the buffer size 1 Mbyte.

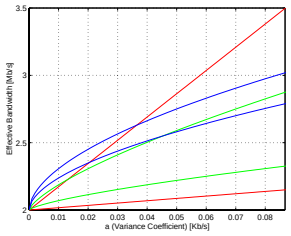


Figure 16. Required link capacity as a function of the peakedness factor a.

As Fig.16 shows, for small a values, the capacity required for $H > 0.5$ will be larger than for $H = 0.5$; eventually, as a increases, because of the slower growth in the required capacity, traffic with $H > 0.5$ may require less capacity (*multiplexing gain*). This is referred to as a *cross-over* [39]. Thus, when the buffer is small, the link requirement depends much less on H than when the buffer is larger [12]. It is very difficult for a SRD traffic to fill up a large buffer, as explained below.

7.3 Queuing Performance

Fig.17 shows another system performance characteristic simulation. First running was intended for SRD processes, in this case classical queuing theory results have emerged as expected, i.e., when a Markovian-type traffic was injected to the buffer, either the $M/M/1$ or $M/D/1$ queuing models were obtained. When LRD processes were used things changed in a sharp and clear way as expected, with such processes as queuing input, even with low system utilization the ATM switch experiments a rapid increase in its buffer size [29], this result suggests that is very probably for self-similar traffic to fill up a buffer[8].

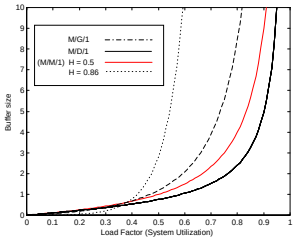


Figure 17. Simple ATM switch queue performance analysis with no feedback control.

The fact that larger buffers are needed for self-similar heavy-traffic conditions explains the observations that, in real network environments, ATM switches do not often meet their specifications, whose derivation is based on non self-similar or short-range dependent models. For this end, the ATM Forum [22] needs to review their congestion avoidance control algorithms and call admission control policies at the user network interface (UNI) point (Fig.13), where the user parameter control (UPC) plays a main role and probably at the network-to-network interface (NNI). This is crucial for the complete success of the ATM or whatever real Broadband-Integrated Services Digital Network (B-ISDN) technology with a guaranteed QoS.

8. SUMMARY

The strong timely correlation in the aggregated IP data traffic trace analyzed was so evident. Therefore, through several statistical methods either in frequency or time domain, an unbiased and well-defined Hurst parameter, $H \approx 0.86$, was obtained. Despite some arguments claiming that the MLE of Whittle is very biased and the Wavelet-based method is not, results have shown that both values are in direct agree each other (see Table 1). Other methods gave compatible results too. Therefore, this permitted us to generate synthetically traces obeying as close as possible the highly dynamic behavior of our real data traffic.

As has been mentioned, this paper is a sort of survey showing classical results that have been obtained in the past years regarding the impact of self-similar traffic in ATM networks, via simulation. Even though, the aggregated IP datagrams traffic is not as Gaussian as expected, it was found that fGn can emulate very well such behavior. Therefore, the use of this mathematical model, as a discrete inter arrival process like packets or cells into a buffer, can be for practical issues, a very good start point to work with. Regarding that IP datagrams will be one of the most important traffic flows in ATM networks, is likely that ATM cells will show the same fractal characteristic, it means that edge ATM switches could receive data traffic with a Hurst parameter between $0.5 < H < 1$ and to forward such data traffic with $H \approx 0.5$ through the network will be a great challenge, dimensioning even more [25, 30, 38]. Results have shown that intrinsic self-similar characteristic of data traffic could give a low QoS for end users. The meaning of self-similarity for dimensioning is still investigated.

So far, fortunately, much work has been done and many researchers all over the world still (hardly) working over those fascinating topics, namely, fractals and high-speed networks toward a comprehensive successful union.

[8] This is critical for VBR video coded as its bit rate is expected not to be network controlled.

References

[1] Fowler, H. J. and W. E. Leland. 1991. "Local Area Network Traffic Characteristics, with Implications for Broadband Network Congestion Management.", *IEEE Journal on Selected Areas in Communications*, No 7, Sep.:1139-1149.

[2] Leland, W. E. and D. V. Wilson. 1991. "High Time-Resolution Measurement and Analysis of LAN Traffic: Implications for LAN Interconnection." In *Proceedings INFOCOM'91* (Bal Harbour, FL, April) IEEE, Piscataway, NJ, 1360-1366.

[3] Leland, W. E.; M. S. Taqqu; W. Willinger; D. V. Wilson. 1994. "On the Self-Similar Nature of Ethernet Traffic." *IEEE/ACM Trans. on Networking*, No 1, Feb.:1-15.

[4] Paxson, V. and S. Floyd. 1995. "Wide Area Traffic: The Failure of Poisson Modeling." *IEEE/ACM Trans. on Networking 3*, No 3, Jun.:226-244.

[5] Beran, J.; R. Sherman; M. S. Taqqu; W. Willinger. 1995. "Long-Range Dependence in Variable-Bit-Rate Video Traffic." *IEEE Trans. on Communications*, Vol. 43, April:1566-1579.

[6] Stallings W. 1993. *SNMP, SNMPv2, and CMIP The practical Guide to Network-Management Standards.* Addison-Wesley, Reading, Massachusetts.

[7] Samorodnitsky G. and M. S. Taqqu. 1994. *Stable Non-Gaussian Random Processes.* Chapman & Hall, One Penn Plaza, New York, NY.

[8] Beran J. 1994. *Statistics for Long-Memory Processes.* Chapman & Hall, One Penn Plaza, New York, NY.

[9] Hosking R. M. J. 1981. "Fractional differencing.", *Biometrica*, Vol. 68, No 1, p. 165-76.

[10] Lan W. C.; A. Erramilli; J. L. Wang; W. Willinger. 1995. "Self-Similar Traffic Generation: The Random Midpoint Displacement Algorithm and Its Properties." In *Proceedings of the 1995 ICC* (Seattle, WA). IEEE, Piscataway, NJ, 466-472.

[11] Mannersalo P. and I. Norros. 1997. "RMD-mn simulator: a fast and simple way to simulate Gaussian processes." Technical Report COST257TD(97). Finland: VTT Information Technology.

[12] Norros I. 1995. "On the Use of Fractional Brownian Motion in the Theory of Connectionless Traffic." *IEEE Journal on Selected Areas of Communications*, Vol. 13, p. 953-962.

[13] Norros I. 1994. "A Storage Model with Self-Similar Input." *Queuing Systems And Their Applications*, Vol. 16, p. 387-396.

[14] Taqqu M. S. and V. Teverovsky. 1995. "Robustness of Whittle-type Estimators for Time Series with Long-Range Dependence." Preprint of Stochastic Models.

[15] Taqqu, M. S.; W. Willinger; V. Teverovsky. 1996. "Is Network Traffic Self-Similar or Multifractal?" *Fractals.*

[16] Resnick, S. I. 1995. "Heavy Tail Modeling and Teletraffic Data." *Annals of Statistics*, Vol. 25, p. 1805-1869.

[17] Crovella, M. E. and A. Bestavros. 1997. "Self-Similarity in World Wide Web Traffic Evidence and Possible Causes." *IEEE/ACM Trans. On Networking 5*, No. 6, Dec.:835-846.

[18] Brillinger D. R. and A. R. Irizarry. 1997."An Investigation of the Second and Higher-order Spectra of Music." Technical Report. Dept. of Statistics, University of Berkeley California, CA.

[19] Jain, R. 1995. "ABR Service on ATM Networks: What is it?" Technical Report, Computer and Information Science Ohio State University.

[20] Huang, C.; M. Devetsikiotis; I. Lambadaris; A. Kaye. 1995. "Modeling and Simulation of Self-Similar Variable Bit Rate Compressed Video: A Unified Approach." *Computer Communication Review*, Vol. 25, p. 114-125.

[21] Taqqu, M. S.; W. Willinger; R. Sherman. 1997. "Proof of a Fundamental Result in Self-Similar Traffic Modeling." *Computer Communication Review.*

[22] The ATM Forum. 1996. "Technical Committee Traffic Management Specification Version 4.0". Tech. Report. (af-tm-0056.000).

[23] Fox, R. and M. S. Taqqu. 1986. "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series." *Annals of Statistics*, Vol. 14, p. 517-532.

[24] Taqqu, M. S.; V. Teverovsky; W. Willinger. 1995. "Estimators for Long-Range Dependence: an Empirical Study." , Vol. 3, No 4, p. 785-798.

[25] Stallings, W. 1997. "Viewpoints: Self-similarity upsets data traffic assumptions." *IEEE Spectrum.*

[26] Adas, A. and A. Mukherjee. 1995. "On Resource Management and QoS Guarantees For Long Range Dependent Traffic." In *Proceedings INFOCOMM'95* (San Francisco, CA), IEEE, Piscataway, NJ.

[27] Angulo, D. A. M. and A. Anzaloni. 1999. "Synthetic Fractional Gaussian Processes to Emulate Actual Data Traffic Behavior of Current High Speed Networks." In *Proceedings of the 1999 Aplied Telecommunication Symposium* (San Diego, CA, April 11-15). SCS.

[28] Narayan, O. 1998. "Exact Asymptotic Queue Length Distribution for Fractional Brownian Traffic." *Advances in Performance Analysis.*

[29] Erramilli, A.; O. Narayan; W. Willinger. 1997. "Frontiers in Queuing: models and applications in science and engineering." Cap. 9: Fractal Queuing Models, CRC Press.

[30] John, B. P.; S. V. Iakovos; S. Jorge-A. 1998. "Shaping Aggregate LAN Flows for Transmission over ABR Connections." *European Communication Networks.*

[31] Bates, S. and S. Mclaughlin. 1998. "Testing the Gaussian Assumption for Self-similar Teletraffic Models." Technical Report. Dept. of Electrical and Electronic Engineering, University of Edinburgh.

[32] Likhanov, N.; B. Tsybakov; N. D. Georganas. 1995. "Analysis of an ATM Buffer with Self-Similar ("Fractal") Input Traffic." In *Proceedings INFOCOMM'95* (San Francisco, CA), IEEE, Piscataway, NJ.

[33] Abry, P. and D. Veitch. 1998. "Wavelet Analysis of Long Range Dependent Traffic." *IEEE Transaction on Information Theory*, Vol. 44, No 1, p. 2-15.

[34] Popescu, A. 1999. "Traffic Self-Similarity." Technical Report, University of Karlskrona/Ronneby, Sweden.

[35] Taqqu M. S. and V. Teverovsky. 1998. "On estimating long-range dependence in finite and infinite variance series." *A Practical Guide to Heavy Tails: Statistical Techniques and Applications.* Robert J. Adler, Raisa E. Feldman and Murad S. Taqqu, editors. Birkhäuser, Boston, p. 177-217.

[36] Qiong Li and D. L. Mills. 1999. "Investigating the Scaling Behavior, Crossover and Anti-persistence of Internet Packet Delay Dynamics." In *Globecom'99* (Rio de Janeiro, Brazil, Dec. 5-9). IEEE, Piscataway, NJ.

[37] Duffield, N. G. and N. O'connell. 1995. "Large Deviations and Overflow Probabilities for the General Single-Server Queue, with Applications." *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 118, p. 363-375. Sciences, Dublin City University, DIAS-APG-93-30.

[38] Willinger, W. and V. Paxson, V. 1998. "Where Mathematics Meets the Internet." *Notice of the AMS*, Vol. 45, No 8, Sep.: 961-971.

[39] Krishnan K. R.; Neidhardt A. L.; Erramilli A. "Scaling Analysis in Traffic Management of Self-similar Processes", In *Proceedings ITC-15* (Washington D.C., April 1997).