

On the Use of Self-Similar Processes in Network Simulation

JOSÉ C. LÓPEZ-ARDAO, CÁNDIDO LÓPEZ-GARCÍA,
ANDRÉS SUÁREZ-GONZÁLEZ, MANUEL FERNÁNDEZ-VEIGA,
and RAÚL RODRÍGUEZ-RUBIO

University of Vigo, Spain

Several recent traffic measurement studies have convincingly shown the presence of self-similarity in modern high-speed networks, involving a very important revolution in the stochastic modeling of traffic. Thus the use of self-similar processes has opened new problems and research fields in network performance analysis, mainly in simulation studies, where the efficient synthetic generation of sample paths (traces) corresponding to self-similar traffic is one of the main topics. In this article, we justify the selection of interarrival time instead of counting processes for modeling arrivals. Also, we discuss the advantages and drawbacks of the most important self-similar processes when applied to traffic modeling in simulation studies, proposing the use of models based in F-ARIMA, mainly due to their flexibility to capture both long- and short-range correlations. However, F-ARIMA processes have been little used in simulation studies, mainly because the synthetic generation methods available in the literature are very inefficient compared with those for FGN. In order to solve this problem, we propose a new method that can generate high-quality traces corresponding to a F-ARIMA(p, d, q) process. A comparison with existing methods shows that the new method is significantly more efficient, and even slightly better than the best method for FGN.

Categories and Subject Descriptors: G.3 [Mathematics of Computing]: Probability and Statistics—*random number generation, stochastic processes*; I.6.m [Simulation and Modeling]: Miscellaneous; C.2.m [Computer-Communication Networks]: Miscellaneous

General Terms: Algorithms, Performance

Additional Key Words and Phrases: Fractional ARIMA, long-range dependence, self-similar processes, synthetic generation

1. INTRODUCTION

Traffic modeling has traditionally been based almost exclusively on the assumption of independence between the random variables that describe arrivals to a network. The fundamental reason for this assumption has been the analytical tractability. However, multimedia traffic is characterized by a high “burstiness” and a strong positive correlation [Fendick et al. 1989; Li and Hwang 1993], much more than in voice traffic [Willinger et al. 1996]. Moreover, several recent traffic measurement studies have convincingly shown the presence of self-similarity in modern high-speed networks [Leland et al. 1993; Paxson and Floyd 1995; Crovella and Bestavros 1997; Garrett and Willinger 1994; Beran et al.

Authors' address: E.T.S.I. Telecomunicación. Campus Universitario. 36200 Vigo (Spain); email: jardao@ait.uvigo.es.

Permission to make digital/hard copy of part or all of this work for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee.

© 2000 ACM 1049-3301/00/04-0125 \$5.00

1995], involving long-range correlations over arbitrarily large time-scales, a phenomenon usually referred to as long-range dependence (LRD).

All these findings have contributed to a very important revolution in the stochastic modeling of traffic, since the impact of the correlation on network performance may be drastic [Norros 1994; Likhanov et al. 1995; Erramilli et al. 1996]. Nevertheless, traditional models, like Markovian or autoregressive, ignore the correlations beyond a particular lag intending to keep the analytical tractability, and so it is said that they exhibit short-range dependence (SRD). But their validity is in doubt since modeling LRD through Markovian or autoregressive processes requires many parameters, and giving physically meaningful interpretations for the parameters becomes difficult.

Because of this, the rise of self-similar processes for network traffic modeling purposes has been essential, due to their capability to exhibit LRD over all time scales by making use of few parameters (parsimonious modeling).

Nevertheless, it is fair to say that some authors question the relevance of LRD when finite buffers are considered [Grossglauser and Bolot 1996; Heyman and Lakshman 1996; Ryu and Elwalid 1996], arguing that the impact of the correlation on performance becomes insignificant beyond certain lag, and so well-designed Markov traffic models are effective enough.

We must say that this discussion is not within the scope of this article, and we focus only on simulation studies with self-similar processes, where a wide range of research topics dealing with new problems arising in the application of self-similar models to performance analysis have been opened [Willinger 1995]. One of the most important problems in this field is the synthetic generation of sample paths (traces) of self-similar processes, since the existing real traces are not generally valid for simulation studies for the following reasons.

- The number of existing real traces is not enough to perform simulations with a large number of nodes, each generating independent traffic, or to study the wide range of different correlation structures (LRD and SRD) [Leland et al. 1993; Beran et al. 1995] and marginal distributions [Heyman et al. 1991; Garrett and Willinger 1994; Paxson and Floyd 1995; Krunz et al. 1995] found in real traffic.
- The traces are not long enough in most cases, and simulation experiments in the context of LRD will usually require very long independent sample paths to reach certain statistical requirements.

For all these reasons, synthetic generation methods are essential. So, the aim of this paper is twofold.

- First of all, we justify the selection of interarrival time instead of counting processes for modeling arrivals. Then, we discuss the advantages and drawbacks of the two types of commonly used self-similar processes: Gaussian (FGN and F-ARIMA) and Poisson ($M/G/\infty$) self-similar processes. After discussion, we propose to make use of Gaussian self-similar processes, choosing F-ARIMA as the most suitable one, because of its flexibility for characterizing both long- and short-range correlations.
- Finally, in relation to the synthetic generation of Gaussian F-ARIMA processes:

- we describe the most important methods, highlighting their poor efficiency compared with some methods for FGN. For this reason, F-ARIMA processes have generally not been used in simulation studies.
- In order to overcome this drawback, we propose a new generation method for Gaussian F-ARIMA(p, d, q) that produces high-quality traces much more efficiently than previous methods, and even slightly better than the best method for FGN.

The rest of the article is organized as follows. In Section 2, we review the main concepts related to LRD and self-similarity. In Section 3, we define the most commonly used self-similar processes in modeling network traffic (FGN, F-ARIMA and $M/G/\infty$). Section 4 presents the above-cited discussion about interarrival time versus counting processes, Gaussian versus Poisson processes, and, finally, F-ARIMA versus FGN. Next, in Section 5, we describe the most commonly used generation methods for F-ARIMA. Section 6 presents and describes a new generation method for F-ARIMA(p, d, q). Also, an extract of a performance study of this new method is included in this section. Finally, Section 7 summarizes the main conclusions.

2. LRD AND SELF-SIMILARITY

It is said that a process exhibits long-range dependence when its autocorrelation function is not summable (or, equivalently, its spectral density has a pole at the origin); that is, $\sum_{k=0}^{\infty} \rho(k) = \infty$, as in those processes whose autocorrelation function decays hyperbolically:¹

$$\rho(k) \sim k^{-\beta}, \quad k \rightarrow \infty, \quad 0 < \beta < 1. \quad (1)$$

In opposition, it is said that a process exhibits short-range dependence when its autocorrelation function is summable (or, equivalently, its spectral density is bounded at the origin), as in those processes whose autocorrelation function decays exponentially:

$$\rho(k) \sim \alpha^k, \quad k \rightarrow \infty, \quad 0 < \alpha < 1. \quad (2)$$

Let $X = \{X_k, k = 1, 2, 3, \dots\}$ be a stationary stochastic sequence and let $X^{(m)}$ be the corresponding aggregated sequence (with aggregation level m), obtained by averaging the original sequence X over nonoverlapping blocks of size m :

$$X_k^{(m)} = \frac{1}{m} \cdot \sum_{i=(k-1)m+1}^{k \cdot m} X_i, \quad k = 1, 2, 3, \dots$$

The stationary sequence X is called *exactly self-similar*, with self-similarity parameter H (the Hurst parameter [Hurst 1951]), if for all m , its finite-dimensional distributions are identical to those of the aggregated sequence $m^{1-H}X^{(m)}$, that is,²

$$X \stackrel{d}{=} m^{1-H}X^{(m)}. \quad (3)$$

¹The symbol \sim denotes an asymptotic relation.

²The symbol $\stackrel{d}{=}$ indicates identical finite-dimensional distributions.

The sequence X is called *asymptotically self-similar* if the previous condition holds as $m \rightarrow \infty$.

Another less strict definition involves the second-order moments exclusively [Cox 1984]. Thus the covariance stationary sequence X is called *exactly second-order self-similar* if the aggregated sequence $m^{1-H}X^{(m)}$ has the same variance and autocorrelation as X for all m , that is, if the aggregated processes possess the same nondegenerate correlation structure as the original stochastic process. If the process is Gaussian, self-similarity and second-order self-similarity are identical. Any of the following conditions is sufficient for second-order self-similarity.

- (1) The autocorrelation function is

$$\rho^{(m)}(k) = \rho(k) = \frac{1}{2} [(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}] \triangleq g^H(k), \quad \forall k > 0. \quad (4)$$

For $1/2 < H < 1$, we can see that [Cox 1984]

$$\rho(k) \sim H \cdot (2H - 1) \cdot k^{2H-2}; \quad (5)$$

that is, it decays hyperbolically as in (1), so the process exhibits long-range dependence.

- (2) The variances satisfy

$$\text{Var}(X^{(m)}) = \text{Var}(X) \cdot m^{2H-2}, \quad 1/2 < H < 1, \quad \forall m > 1. \quad (6)$$

- (3) The spectral density is

$$f(\lambda) = c \cdot |e^{j\lambda} - 1|^2 \cdot \sum_{i=-\infty}^{\infty} |2\pi i + \lambda|^{2H-1}, \quad \lambda \in [-\pi, \pi], \quad 1/2 < H < 1, \quad (7)$$

where c is a normalization constant such that $\int_{-\pi}^{\pi} f(\lambda) d\lambda = \text{Var}(X)$. The behavior of spectral density near the origin is obtained by Taylor expansion at zero:

$$f(\lambda) \sim |\lambda|^{1-2H}. \quad (8)$$

If (4) is satisfied asymptotically (as $m \rightarrow \infty$), the sequence is called *asymptotically second-order self-similar* [Cox 1984]:

$$\lim_{m \rightarrow \infty} \rho^{(m)}(k) = g(k), \quad \forall k > 0. \quad (9)$$

It has been shown that a second-order stationary process whose autocorrelation function decays hyperbolically is asymptotically second-order self-similar (see proof in Tsybakov and Georganas [1997]). For this reason, although long-range dependence and self-similarity are not equivalent concepts, they are often utilized without distinction.

3. SELF-SIMILAR PROCESSES

Next, we describe the most commonly used self-similar processes: FGN, F-ARIMA, and $M/G/\infty$.

3.1 Fractional Gaussian Noise (FGN)

Let $B_H(t)$ be an FBM process.³ Then, the sequence of increments, $X_t = B_H(t) - B_H(t-1)$, is an exactly self-similar stationary Gaussian process with zero mean, referred to as fractional gaussian noise (FGN). In fact, the FGN process is the only stationary Gaussian process that is exactly self-similar [Samorodnitsky and Taqqu 1994].

The autocorrelation function and the spectral density for FGN are given, respectively, by expressions (4) and (7).

In spite of their rigid correlation structure, FGN processes are a good approximation of the aggregation of more complex LRD Gaussian processes [Samorodnitsky and Taqqu 1994].

3.2 Fractional ARIMA Processes

Fractional autoregressive integrated moving average (F-ARIMA) processes are a generalization of the widely used ARIMA models introduced by Box and Jenkins [Box et al. 1994].

An ARIMA(p, d, q) process, X_t , satisfies the equation

$$\phi_p(\mathcal{B}) \cdot (1 - \mathcal{B})^d \cdot X_t = \theta_q(\mathcal{B}) \cdot \epsilon_t, \quad (10)$$

where d is an integer, \mathcal{B} is the backshift operator ($\mathcal{B}^j \cdot X_t = X_{t-j}$), $\phi_p(\mathcal{B})$ is a polynomial of order p in \mathcal{B} , $\theta_q(\mathcal{B})$ is a polynomial of order q in \mathcal{B} , and ϵ_t is a renewal process with zero mean and variance σ_ϵ^2 .

Equation (10) can be generalized in a natural way by allowing d to assume any real value.

If d is an integer, then

$$(1 - \mathcal{B})^d = \sum_{k=0}^d \binom{d}{k} (-1)^k \mathcal{B}^k, \quad (11)$$

with the binomial coefficients

$$\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1) \cdot \Gamma(d-k+1)}. \quad (12)$$

³Fractional Brownian motion (FBM), introduced by Mandelbrot [Mandelbrot and Ness 1968] is a statistically self-similar continuous time process; namely, it verifies $Y_{at} \stackrel{d}{=} a^H Y_t$, $\forall a > 0$.

As the gamma function ($\Gamma(x)$) is also defined for all real numbers, and only the first $d + 1$ terms in (12) are nonzero if d is a positive integer,⁴ the expression (11) can be extended to any real number d by

$$(1 - \mathcal{B})^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k \mathcal{B}^k. \quad (13)$$

Thus, the expression (10) can also be extended to noninteger values of d , giving rise to fractional ARIMA processes, whose definition was proposed by Granger and Joyeux [1980] and Hosking [1981].

F-ARIMA processes have zero mean and variance σ_ϵ^2 , and they are stationary and invertible for $-\frac{1}{2} < d < \frac{1}{2}$, if all solutions of $\phi_p(\mathcal{B}) = 0$ and $\theta_q(\mathcal{B}) = 0$ lay outside the unit circle.

The spectral density of a F-ARIMA(p, d, q) process has the form

$$f(\lambda, \sigma_\epsilon^2, d, \phi_p, \theta_q) = \frac{\sigma_\epsilon^2}{2\pi} \left(2 \sin \left(\frac{\lambda}{2} \right) \right)^{-2d} \frac{|\theta_q(e^{j\lambda})|^2}{|\phi_p(e^{j\lambda})|^2}, \quad (14)$$

and its behavior at the origin ($\lambda \rightarrow 0$) is given by

$$f(\lambda) \sim f_{\text{ARMA}}(0) \cdot |\lambda|^{-2d}. \quad (15)$$

That is, for $0 < d < \frac{1}{2}$ spectral density has a pole at the origin and, therefore, the process exhibits long-range dependence. If we compare Equation (15) with (8), we can obtain the relation between d and H :

$$d = H - \frac{1}{2}.$$

Although it is not generally feasible to obtain the autocorrelation function for a F-ARIMA(p, d, q) process, for F-ARIMA($0, d, 0$) is of the form [Granger and Joyeux 1980]

$$\rho(k) = \frac{\Gamma(1-d)}{\Gamma(d)} \cdot \frac{\Gamma(k+d)}{\Gamma(k+1-d)} = \frac{\prod_{i=1}^k (d+i-1)}{\prod_{i=1}^k (i-d)}, \quad (16)$$

whose asymptotic behavior (as $k \rightarrow \infty$) is

$$\rho(k) \sim \frac{\Gamma(1-d)}{\Gamma(d)} \cdot k^{2d-1}. \quad (17)$$

So, for $0 < d < \frac{1}{2}$, the autocorrelation function exhibits a hyperbolic decay as expressed in (1) and, in view of (9), F-ARIMA($0, d, 0$) processes are asymptotically second-order self-similar. Cox [1984] extended this result showing that any F-ARIMA(p, d, q) process, with $0 < d < \frac{1}{2}$, is asymptotically second-order

⁴Gamma function has poles for negative integers so that the binomial coefficient is zero if $k > d$ and d is an integer.

self-similar. If ϵ_t is Gaussian, the process is also Gaussian and asymptotically self-similar.

Finally note that, for large lags, the correlation structure of the simplest LRD F-ARIMA process, F-ARIMA(0, d, 0), is very similar to that of any F-ARIMA(p, d, q) process, without regard to the value of p and q. That is, F-ARIMA processes allow us to model SRD and LRD independently.

3.3 $M/G/\infty$ Processes

Another interesting self-similar process is the occupancy process of an $M/G/\infty$ queueing model, referred to as an $M/G/\infty$ process. In this queueing model, customers arrive according to a Poisson process with rate λ , to an infinite server group, and service times are independent and identically distributed according to a continuous-time general distribution S , with finite mean.

Cox and Isham [1980] showed that the process X_t (number of customers, or busy servers, in the system at t) is a Poisson process with mean $\lambda E(S)$ whose autocovariance function depends on the distribution function $F_S(x)$, and is obtained from

$$\gamma(k) = \lambda \cdot E((S - k)^+) = \lambda \int_k^\infty [1 - F_S(x)] dx \quad k = 0, 1, \dots, \quad (18)$$

where $(x)^+ = x$ if $x \geq 0$, and 0 otherwise.

Generally, the process X_t is not stationary although, under certain conditions, the system admits a stationary and ergodic solution [Parulekar 1997], X_t^* ,

$$\{X_{t+k}, t = 0, 1, \dots\} \xrightarrow{d} \{X_t^*, t = 0, 1, \dots\} \quad \text{as } k \rightarrow \infty,$$

for any initial distribution of the random variable “initial number of customers in the system,” X_{0-} , and for any distribution of the random variables that describe the residual life, $\{\hat{S}_j, j = 1, \dots, X_{0-}\}$.

Furthermore, if the initial probability distribution of the corresponding Markov chain is the stationary distribution and the residual life distribution is that given by Kleinrock [1975],

$$f_{\hat{S}}(x) = \frac{1 - F_S(x)}{E(S)}, \quad (19)$$

then the chain will be strict-sense stationary. So, X_{0-} must be a Poisson random variable with rate $\lambda \cdot E(S)$ [Parulekar and Makowski 1996; Parulekar 1997]. On the other hand, if $E(S)$ is finite, it can be demonstrated that

$$X_t \text{ exhibits LRD} \iff \text{Var}(S) = \infty$$

as happens in heavy-tailed service distributions.

Considering that X_t is a discrete-time process, the model simulation suggests making use of a discrete-time clock. Consequently, it would seem more appropriate to utilize discrete-time service distributions too. However, it is impossible to calculate the mean for heavy-tailed discrete-time distributions, although it exists and is finite. This obstacle can be overcome by using continuous-time

distributions, although discretization should be done very carefully. For example, in Paxson and Floyd [1995] a Pareto distribution with infinite variance ($1 < \alpha < 2$) is proposed, yielding a process with an autocorrelation function that shows a hyperbolic decay,

$$\rho(k) \sim k^{1-\alpha},$$

and so, the $M/G/\infty$ process is second-order asymptotically self-similar with Hurst parameter $H = (3 - \alpha)/2$.

4. WHAT SELF-SIMILAR PROCESS MUST I CHOOSE?

When we face the simulation of a communications network making use of self-similar traffic models, we have to make initially two very important decisions.

- With regard to arrivals modeling, we must choose the most suitable type of process for the traffic model: interarrival time or counting processes.
- Then, we should carefully select the most suitable self-similar process according to the objectives, namely, a process that has enough flexibility for modeling the most important traffic characteristics with few parameters.

With respect to the first issue, and given that self-similarity was found in counting processes of real traffic, it is obvious why this approach is commonly selected. However, it has two great drawbacks.

- We must take into account the subjectivity in choosing the time scale, since it could have an important impact on some performance measurements [Paxson and Floyd 1995; Leland et al. 1993]. Consequently, simulation studies could depend on the particular time-scale chosen, and so extrapolations and comparisons become very difficult.
- On the other hand, given that the network simulator will generally use interarrival times, we must deal with the added difficulty of distributing the arrivals into each interval; that is, convert arrival counts into interarrival times. For example, Lau et al. [1995] highlight this problem and propose using fractal interpolation in order not to destroy the correlation structure for time scales shorter than the modeled one in the counting process, as would happen if arrivals were equally spaced, or distributed uniformly (as a local Poisson process). Nevertheless, this approach makes sense only if the number of arrivals per interval is large. If at any point this number is small enough, the interpolation becomes problematic, and perhaps incorrect, inasmuch as the real arrival processes at the finest time scales may no longer be self-similar [Paxson and Floyd 1995; Leland et al. 1993], or even exhibit a multifractal behavior [Feldmann et al. 1999].

In order to avoid all these problems, it would be desirable to use interarrival time models in such a way that the corresponding counting processes still exhibit LRD. In this regard, an empirical study in López-Ardao [1999] shows that the interarrival time processes and the corresponding counting processes exhibit a duality in their LRD structure; namely, the estimated value of H is practically the same. This behavior is also observed when we construct interarrival time processes starting from counting processes, even using different

methods of distributing arrivals. In this study we have used different marginal distributions and short-range correlation structures. Consequently, we propose modeling arrivals by interarrival time processes and so, henceforth, we always refer to these processes.

Next we must choose a self-similar process that is able to model the most important traffic characteristics parsimoniously. In our view, such a process should at least have a well-known marginal distribution and allow us to set its mean and variance. For this reason, some LRD processes with unknown marginal distribution, such as chaotic maps [Erramilli et al. 1995] or the shifting level process [Mandelbrot 1967; Grasse et al. 1997], are seldom utilized. Instead, FGN and Gaussian F-ARIMA processes, both with Gaussian marginal distribution, and $M/G/\infty$ processes, with Poisson marginal distribution, are commonly used.

Since the queueing performance of a network depends not only on the correlation structure, but also on the marginal distribution of the input processes [Grossglauser and Bolot 1996; Ryu and Elwalid 1996], mainly due to its tail behavior [López-Ardao et al. 1998; López-Ardao 1999], it would be very important to be able to modify the marginal distribution of the process without altering its correlation structure.

In this regard, FGN and Gaussian F-ARIMA processes enjoy the attractive advantage that their marginal distribution can be changed by a sufficiently regular transformation⁵ (parameters included) without modifying their long-range correlation structure [Huang et al. 1995b]. Nevertheless, the magnitude of the short-range correlations generally decreases slightly.

Instead, the non-Gaussian marginal distribution of $M/G/\infty$ processes can become a drawback from the point of view of its modeling flexibility. In addition, given that it is inherently a counting process, and so an integer-valued process, its use as an interarrival time process could make sense only in slotted-time simulations.

However, $M/G/\infty$ processes have the advantage that the synthetic generator can be embedded in the code of the simulator, and the samples can be generated in a sequential way as long as the simulator needs them (a method usually referred to as online generation).

However, the available methods for synthesizing FGN or F-ARIMA are offline; namely, for each source of the self-similar traffic a large enough trace must be generated before a simulation run. In this case, the finite length of traces limits the set of usable mean estimation techniques: the independent replication technique is the only one flexible enough. Moreover, if we are interested in the simulation of models with several traffic sources (as, e.g., in communication networks of several nodes) it may happen that all needed traces for a simulation run do not fit into core memory, so they have to be stored in disk. In this case there are two bottlenecks on computational efficiency in addition to CPU power: core memory size and disk access. To reduce this computational cost, in Suárez et al [1999] the authors propose a concurrent methodology based on sharing resources between running processes for efficiently organizing and dealing with this simulation study.

⁵The distribution change will be done by a transformation $h(x) = F_Y^{-1}(F_X(x))$, where $F_X(x)$ is the original marginal distribution and $F_Y(x)$ is the target marginal distribution.

Despite the drawbacks of the $M/G/\infty$ process, the possibility of online generation makes it attractive for use as an interarrival time model in slotted-time simulators, or as a message (frame or packet) size process. And, of course, they could be utilized as counting processes if considered suitable. On the other hand, the $M/G/\infty$ model can be an acceptable explanation for the presence of self-similarity in WAN traffic [Paxson and Floyd 1995], where, by means of a multilevel approach, we can assume Poisson processes at the session time scale, and heavy-tailed distributions at the packet (frame) time scale.

For all these reasons, at present we are working on the design of an efficient online generator for $M/G/\infty$ that allows modeling of SRD and LRD simultaneously.

However, from now on, we consider the $M/G/\infty$ process not within the scope of this article, and focus on Gaussian self-similar processes.

It is obvious that Gaussian processes can also be utilized as counting processes⁶ but, in light of the above discussion, we focus on their use as interarrival time processes.

4.1 Gaussian Self-Similar Processes

In this section we compare both Gaussian self-similar processes, FGN and Gaussian F-ARIMA, with regard to the ability to model different correlation structures. We also discuss their advantages and disadvantages for simulation purposes.

FGN processes have only three parameters (μ , σ , and H), and their correlation structure is determined by a single parameter H , that only permits us to model their LRD structure in such a way that their SRD structure remains fixed and can not be modified independently. So, FGN processes have a correlation structure that is too rigid to capture the wide range of low-lag correlation structures encountered in practice (VBR video, e.g.) [Leland et al. 1993]. This fact is a strong limitation of FGN since although the presence of LRD can have a dramatic effect on performance, SRD can also have a significant impact, involving the need to incorporate SRD into traffic models [Erramilli et al. 1996; López-Ardao 1999].

Instead, F-ARIMA(p , d , q) processes are much more attractive for traffic modeling purposes, since they are capable of modeling LRD (by means of the parameter d) and SRD (by means of the AR(p) and MA(q) components) independently. In fact, the simple addition of one AR(1) or MA(1) component to a F-ARIMA(0, d , 0) process creates processes that are much more flexible than FGN [Hosking 1981; Leland et al. 1993].

On the other hand, the fact that F-ARIMA processes are only asymptotically self-similar is not important given that simulation experiments in the context of LRD will usually require very long sample paths to reach certain statistical requirements.

For example, a simulation study of an FDDI network in López-Ardao [1999] shows that mean waiting time for FGN is practically identical to that for F-ARIMA(1, d , 0) if its SRD structure is adequately fitted to FGN by means of the AR(1) component. In Figures 1 and 2 we can see that the autocorrelation

⁶In the case of FGN, this involves the cumulating arrival process being FBM. Not in vain, FBM-based models exploit this fact [Norros 1994].

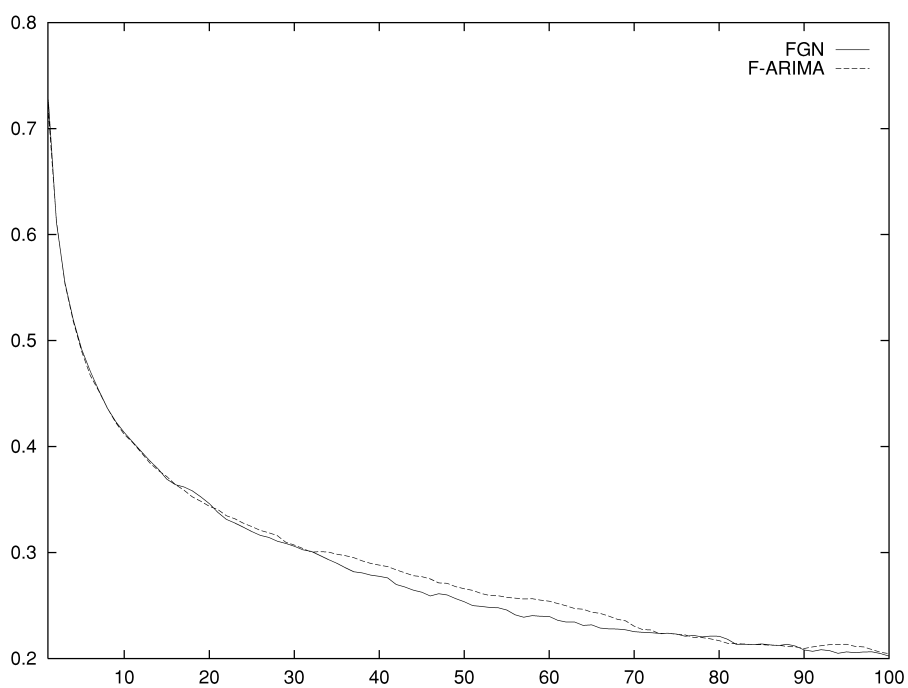


Fig. 1. Short-term correlations of synthesized traces of F-ARIMA(1, d, 0) with $\alpha_1 = 0.09$ and FGN. $H = 0.9$.

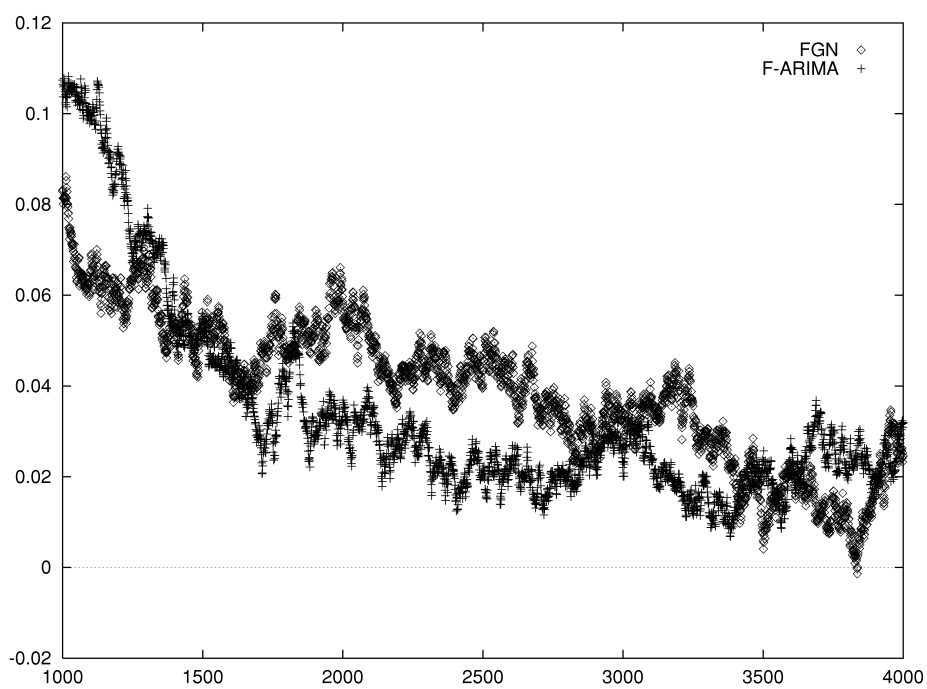


Fig. 2. Long-term correlations of synthesized traces of F-ARIMA(1, d, 0) with $\alpha_1 = 0.09$ and FGN. $H = 0.9$.

function of a synthesized trace of F-ARIMA(1, d, 0) with $\alpha_1 = 0.09$ fits almost perfectly, both short- and long-term, the autocorrelation function of a trace of FGN.

Due to the large number of long runs that must be used in an exhaustive simulation study in the context of LRD, the time spent in synthesizing a very large trace is critical. For this reason, despite their flexibility, F-ARIMA processes have been little used in simulation studies, since the commonly used generation methods for FGN are significantly more efficient than those for F-ARIMA.

In order to overcome this drawback, in this article we propose a new generation method for Gaussian F-ARIMA(p, d, q) that produces high-quality traces much more efficiently than the usual methods, even if compared with the best method for FGN.

5. SYNTHETIC GENERATION OF F-ARIMA

In this section we describe the most important methods for synthesizing Gaussian F-ARIMA traces that have been proposed.

5.1 Exact Methods

Several exact methods exist for generating Gaussian processes with a given autocorrelation (or autocovariance) function and, therefore, suitable for F-ARIMA(0, d, 0) processes (also for FGN), whose autocorrelation function is given by expression (16).

The most direct method consists of multiplying a vector of N (sample size) i.i.d. standard normal random variables by a real lower triangular matrix, obtained from a Cholesky decomposition of the covariance matrix. Since the traces needed in the context of LRD are usually very long, the covariance matrix can be extremely huge, involving a large amount of computer memory, seldom available, and a prohibitive computational cost (on the order of N^2).

Hosking [1984] proposed another general exact method that can be utilized for generating a sample path corresponding to a Gaussian process, $\{X_k | k = 0, \dots, N-1\}$, with zero mean, variance v_0 , and a given autocorrelation function, $\rho(k)$.

The algorithm is as follows.

Starting from $N_0 = 0$, $D_0 = 1$, and from an initial value X_0 , drawn from a standard normal distribution $N(0, v_0)$, perform the following operations for $k = 1, \dots, N-1$.

1. $N_k = \rho(k) - \sum_{j=1}^{k-1} \phi_{k-1,j} \cdot \rho(k-j)$.
2. $D_k = D_{k-1} - N_{k-1}^2 / D_{k-1}$.
3. $\phi_{kk} = N_k / D_k$.
4. For $j = 1, \dots, k-1$ calculate $\phi_{kj} = \phi_{k-1,j} - \phi_{kk} \cdot \phi_{k-1,k-j}$.
5. $m_k = \sum_{j=1}^k \phi_{kj} \cdot X_{k-j}$ and $v_k = (1 - \phi_{kk}^2) \cdot v_{k-1}$.
6. Obtain the sample X_k from a standard normal distribution $N(m_k, v_k)$.

This method has often been used in the literature for synthesizing Gaussian F-ARIMA(0, d, 0) traces, but, as far as we know, it has not been applied for FGN. For instance, in Garrett and Willinger [1994] this method is utilized for generating a F-ARIMA(0, d, 0) trace that exclusively captures the LRD structure of a

VBR video sequence. However, given that the method is suitable for synthesizing any Gaussian causal process with given autocorrelations [Hosking 1984], in Huang et al. [1995b] it is used with the empirical autocorrelations of a real video trace.

Nevertheless, as the calculation of each sample depends on the previous samples, the great disadvantage of this method is its extremely high computational cost (on the order of N^2), that makes it absolutely prohibitive for very long traces. For example, in Leland et al. [1994] the authors report that synthesizing 100,000 samples of F-ARIMA(0, d, 0) requires 10 hours of CPU time on a Sun SPARCstation-2.

In the case of F-ARIMA(0, d, 0), given that $\phi_{kk} = d/(k - d)$ [Hosking 1981] in Step 3 of the algorithm, there is no need to calculate either $\rho(k)$ or to perform Steps 1 and 2, so that computation time decreases about 40% in relation to the general case.

5.2 Approximate Methods

The computational cost can be reduced significantly if we make use of approximate methods that, however, require the consideration of an additional criterion: the goodness of the approximation, evaluated by the quality of the synthesized traces.

Next, we describe and discuss the three most important approximate methods for synthesizing Gaussian F-ARIMA(0, d, 0) traces: Davies–Harte, Haslett–Raftery, and aggregation of AR(1) processes.

In order to synthesize a F-ARIMA(p, d, q) trace we must make use of an ARMA(p, q) filter with F-ARIMA(0, d, 0) traces as innovations.

5.2.1 Davies–Harte Method. This general method for synthesizing Gaussian processes [Davies and Harte 1987] is based on the existing relation between the spectral density and the modulus of the Fourier transform of the sequence to be generated. The spectral density is obtained by means of a discrete Fourier transform (DFT) of the autocovariances.

The algorithm for generating a sequence $\{X_t, t = 0, 1, \dots, N - 1\}$ is the following.

1. Calculate the spectral density $\{f_k\}$ sampled at each frequency $\lambda_k = 2\pi k/N$, with $k = 0, \dots, 2N - 1$, by means of the DFT (of $2N$ points) of the autocovariances sequence, $\gamma(0), \gamma(1), \dots, \gamma(N - 1), \gamma(N), \gamma(N - 1), \dots, \gamma(1)$.
2. Define a set of complex random variables $Z_k = \sqrt{f_k} \cdot (a_k + jb_k)$, $k = 0, \dots, 2N - 1$, where
 - $Z_0 = 0$;
 - for $1 \leq k \leq N - 1$, $\{a_k\}$ and $\{b_k\}$ are two independent sequences of normal i.i.d. random variables with zero mean and variance 1;
 - a_N is a normal random variable with zero mean and variance 2, and $b_N = 0$;
 - for $N + 1 \leq k \leq 2N - 1$, $Z_k = Z_{2N-k}^*$.
3. Calculate X_t , for $t = 0, \dots, N - 1$, as the inverse-DFT (of $2N$ points) of the sequence Z_k and then divide by $2\sqrt{N}$.

In this way, if we set $\gamma(0) = 1$, we obtain a Gaussian sequence with zero mean and variance 1, whose autocorrelation function will be given by $\rho(k) = \gamma(k)$.

The advantage of this method lies in the use of FFT algorithms for calculating both DFTs, direct and inverse. So, the method is fast enough, with a computational cost that is only proportional to the sample size, and it is a good choice for generating FGN or Gaussian F-ARIMA(0, d, 0), since their autocovariances are known.

5.2.2 Haslett–Raftery Method. The method proposed in Haslett and Raftery [1989] is a slight but very efficient modification of the Hosking method when applied for F-ARIMA(0, d, 0). The modification consists of an approximation of the sum in Step 5, whose calculation is the most computationally intensive, that is based on another approximation proposed by Hosking [1981] for F-ARIMA(0, d, 0):

$$\phi_{kj} \sim -\pi_j \quad \text{as } k \rightarrow \infty$$

and

$$\pi_j \sim \frac{k^{-d-1}}{(-d-1)!}.$$

Haslett and Raftery propose taking this asymptotic relation as exact for $j > M$ and assume that π_j is constant for $M < j \leq k-1$ and equal to the approximated mean value. In this way, for $k \leq M+1$, the expression for m_k in Step 5 is calculated as usual in the Hosking algorithm, and for $k \geq M+2$:

$$\begin{aligned} m_k &= \sum_{j=1}^k \phi_{kj} \cdot X_{k-j} = \phi_{kk} \cdot X_0 + \sum_{j=1}^M \phi_{kj} \cdot X_{k-j} - \sum_{j=M+1}^{k-1} \pi_j \cdot X_{k-j} \\ &\approx \phi_{kk} \cdot X_0 + \sum_{j=1}^M \phi_{kj} \cdot X_{k-j} - M \cdot \pi_M \cdot \frac{1 - (M/j)^d}{k-1-M} \cdot \sum_{s=1}^{k-M-1} X_{M+s}. \end{aligned}$$

Haslett and Raftery recommend using $M = 100$, a value that shows good performance over a wide range of values of d and N.

This approximation makes the computational cost proportional to the sample size, and therefore, the algorithm is much faster than Hosking’s method (about 70 times), although this method is still nearly 6 times slower than the Davies–Harte method, with similar quality of the synthesized traces.

This algorithm is included in the commonly used commercial statistical software S-Plus [Statistical Sciences, Inc. 1991].

5.2.3 Aggregation of AR(1) Processes. This method exploits a convergence result obtained by Granger [1980] who showed that the aggregation of n AR(1) processes, where the AR(1) parameters are chosen from a Beta distribution on $[0, 1]$ with shape parameters r and s , converges to a F-ARIMA(0, d, 0) process, with $d = 1 - s/2$, as n goes to infinity. Moreover, if the innovations of all AR(1) processes are Gaussian, then the aggregate process is also Gaussian.

The main difficulty of this method is that one must trade off computational efficiency (low n) against the degree of agreement with a true self-similar process (high n). Nevertheless, its advantage is that it is well suited for parallel

computers. For instance, in Willinger et al. [1995] it is said that generating a synthetic trace of length 100,000 on a MasPar MP-1216, a massively parallel computer with 16,384 processors, takes about three to five minutes. In any case, this computational cost is still significantly higher than that for the Davies–Harte method, which requires only eight seconds on an ordinary Pentium-166 processor.

5.3 Discussion

In summary, the Davies–Harte method is the most suitable one for generating approximate Gaussian F-ARIMA(0, d, 0). Nevertheless, it is striking that this method has hardly been cited or used in the literature (as far as we know, this method is only proposed in Beran [1994], where the code in S-Plus is included).

In fact, the studies with F-ARIMA models often make use of Hosking’s method [Garrett and Willinger 1994; Huang et al. 1995a, 1995b] and sometimes the Haslett–Raftery method [Krunz and Makowski 1998; Adas and Mukherjee 1996].

In any case, if there is no need to model the SRD structure, the best choice is undoubtedly an FGN-based model, due to the existence of an excellent approximate method proposed by Paxson [1997], that is four times more efficient than the Davies–Harte method for a given sample size.

In order to avoid that disadvantage, and to increase the efficiency in synthesizing F-ARIMA traces, we propose a new generation method for Gaussian F-ARIMA(p, d, q) that is significantly more efficient than the Davies–Harte method, and even slightly more than Paxson’s method for FGN.

We must mention the method proposed in Kaplan and Kuo [1994] for synthesizing “asymptotic FGN” by means of Haar wavelet transforms. The main idea is to find a model whose correlation decays asymptotically like FGN but where the short-term correlations can be modeled independently like an AR(1) process. Thus the characteristics of this model are very similar to those of F-ARIMA(1, d, 0). In general, wavelet methods hold great promise for modeling and synthesizing self-similar traffic, due to their computational efficiency and the natural match between the notion of “scaling” in a wavelet transform and the notion of “invariance across different scales” in a self-similar process [Paxson 1997]. However, at the moment, this method and others are somewhat limited due to difficulties in parameter estimation.

6. A NEW METHOD FOR SYNTHESIZING GAUSSIAN F-ARIMA(p, d, q)

The proposed method is an extension of Paxson’s [1997] method to F-ARIMA(p, d, q) processes. The basic idea consists of using the same result utilized by Paxson, belonging to Graf [1983], which states that the power spectrum estimated (or periodogram) for a given frequency is distributed asymptotically as an independent exponential random variable with mean equal to the actual power. Therefore, the method obtains a hypothetical periodogram of the designated process—F-ARIMA(p, d, q) in our case—by randomizing its theoretical power spectrum, given by expression (14).

The steps of the algorithm for generating one trace X_t with length N (even) are basically the same as those of Paxson’s method.

- (1) Construct a sequence of values $\{f_1, \dots, f_{N/2}\}$, where $f_k = f(2\pi k/N; \sigma_\epsilon^2, d, \phi_p, \theta_q)$, and f is the spectral density given by (14) for frequencies from $2\pi/N$ to π . Without loss of generality, we assume $\sigma_\epsilon^2/2\pi = 1$, because later we propose a normalization for obtaining variance 1.
- (2) Randomize the sequence $\{\hat{f}_k\}$ by multiplying each f_k by an independent exponential random variable with mean 1.
- (3) Construct a sequence of complex values $\{z_1, \dots, z_{N/2}\}$, such that $|z_k| = \sqrt{\hat{f}_k}$ and the phase of z_k is uniformly distributed between 0 and 2π , except for the phase of $z_{N/2}$ that must be zero. This randomization of the phase preserves the spectral density (and thus autocorrelation) corresponding to $\{\hat{f}_k\}$, but ensures that different traces generated with the method will be independent. It also makes the final marginal distribution normal [Paxson 1997].
- (4) Construct the sequence $\{z'_0, \dots, z'_{N-1}\}$ as

$$z'_k = \begin{cases} 0 & \text{if } k = 0 \\ z_k & \text{if } 0 < k \leq N/2 \\ \bar{z}_{N-k} & \text{if } N/2 < k < N \end{cases}$$

in such a way that now the sequence z'_k is symmetric about $z'_{N/2}$, corresponding to the Fourier transform of a real-valued sequence.

- (5) Calculate the inverse DFT of the sequence z'_k to obtain X_t , the approximate synthetic trace corresponding to F-ARIMA(p, d, q). As in the Davies–Harte method, the main advantage lies in the use of an FFT algorithm for calculating the inverse DFT.

The resulting trace has zero mean, but its standard deviation is unknown, although it can be estimated, and so we can obtain a standard normal marginal distribution by dividing each sample of the trace by this estimation.

However, as the usual estimator of the variance, the sample variance, has a strong bias for series exhibiting LRD, higher as H increases [Beran 1994], we propose using the following approximation of the variance in order to obtain higher quality.

Let $\{X_t, t = 0, \dots, N-1\}$ be the sequence to be generated, let $\chi(k)$ be its DFT, and let $I_x(k)$ be its periodogram for frequencies $\lambda_k = (2\pi/N) \cdot k$. From

$$I_x(k) = \frac{1}{2\pi N} \cdot |\chi(k)|^2$$

and since we take $|\chi(k)|^2 = |z_k|^2 = \hat{f}(\lambda_k)$ in the algorithm (Step 3), then

$$I_x(k) = \frac{1}{2\pi N} \cdot \hat{f}(\lambda_k).$$

Thus the spectral density corresponding to X_t is really $(f(\lambda))/(2\pi N)$, so that the theoretic variance will be

$$\sigma_x^2 = \int_{-\pi}^{\pi} \frac{f(\lambda)}{2\pi N} d\lambda.$$

We can obtain an approximation of the variance $\tilde{\sigma}_x^2$ by approximating the integral at the frequencies $\lambda_k = (2\pi/N) \cdot k$:

$$\tilde{\sigma}_x^2 = \frac{2\pi}{N} \sum_{k=1}^{N-1} \frac{f(\lambda_k)}{2\pi N} = \frac{1}{N^2} \sum_{k=1}^{N-1} f(\lambda_k).$$

It is easy to see that $\tilde{\sigma}_x^2 \rightarrow \sigma_x^2$ as $N \rightarrow \infty$. Note that $\tilde{\sigma}_x^2$ is not an estimator of the variance, but an approximation, obtained from the theoretical power spectrum, that is not random. Nevertheless, when the sample size is large enough (as usual in the context of LRD), the approximation is good enough and more suitable than the sample variance.

On the other hand, FFT algorithms imply division by N in the case of inverse DFT.⁷ So, in order to obtain variance equal to 1, each sample of the resulting sequence must be divided by N and by $\tilde{\sigma}_x$, that is, must be divided by $\sqrt{\sum_{k=1}^{N-1} f(\lambda_k)}$. It is clear that this normalization also can be used in Paxson's method for FGN.

6.1 Issues Related to Efficiency

If we compare our method with Davies–Harte, at first glance we can see why the new method is significantly more efficient. The Davies–Harte method uses two FFTs of $2N$ points, whereas the new method uses only one FFT of N points, decreasing the computational cost about four times. For example, synthesizing a F-ARIMA(0, d , 0) trace of a length of one million on a Pentium-166 processor takes 76 seconds with the Davies–Harte method, but only 17.5 seconds with the proposed method, that is, more than four times faster.

Paxson's method is also less efficient, requiring 31 seconds. This difference with Paxson's method is basically due to the computational difference in the calculation of the sequence $\{f_1, \dots, f_{N/2}\}$ (corresponding to the sampled spectral density obtained in Step 1 of the algorithm) between FGN and F-ARIMA(0, d , 0). In the case of FGN, such calculation involves an infinite summation—expression (7)—for which no closed form is known, and so an approximation must be obtained. Paxson proposes an approximation with six terms and a correcting factor that is good enough. However, even making use of this approximation, this calculation takes more time than evaluating (14) for F-ARIMA(0, d , 0), that has a closed form, and this represents an additional advantage of the proposed method.

An interesting improvement for both methods could be storing the sequence $\{f_1, \dots, f_{N/2}\}$ in disk, so that this sequence can be used in subsequent generations instead of being calculated, thus saving more time. This improvement allows us to save much time in cases where the sampled power spectrum remains invariable (i.e., the target correlation structure (H , p and q) and the sample size) and can be shared. For example, symmetrical configurations of the network are often used in simulation studies, in which traffic patterns are identical for each source and it is necessary to synthesize several independent

⁷The reason lies in the fact that both algorithms, direct and inverse, are the same except for the sign of the exponentials. However, in the case of inverse DFT, it is necessary to divide by the sample size, and this operation is often made subsequently.

traces with the same correlation structure. Furthermore, this sampled spectrum can be used to run many simulations due to the use of the independent replication technique. In addition, all simulation experiments where other parameters (marginal distribution, mean, variance, network parameters, etc.) are modified can share the same sampled spectrum.

It is easy to infer that the efficiency gain in our method is higher as the order p or q increases. But, even in the simplest cases, F-ARIMA(0, d , 0) and F-ARIMA(1, d , 0), the gain is about 10 and 15%, respectively. In short, it is advisable to store the sampled spectrum in disk if we must synthesize many traces with the same correlation structure.

It is also obvious that the efficiency gain for Paxson's method is significantly higher than that for our method, even though we make use of an approximation with few terms for the former. In fact, both methods have the same computational cost if the sampled spectrum is read from disk. On the other hand, storing the spectrum in disk could allow us to make use of an approximation with more terms that possibly involves a higher quality of the traces.

6.2 Issues Related to the Quality of the Traces

Next, we present a summary of a comprehensive study that evaluates the quality of the traces generated by our method. The goal of this study is not to prove that the method really synthesizes sample paths corresponding to a F-ARIMA(p , d , q) process, because this is not true, as Graf's result is only asymptotically true and the method uses a finite power spectrum [Paxson 1997].

Instead, this study intends to show that the method produces sample paths that are indistinguishable for practical purposes from those corresponding to a true F-ARIMA process. By this we mean that its marginal distribution is normal with zero mean and variance 1, that its autocorrelation function adequately fits the true function and the estimation of H is close to the true value, and that its spectral density is consistent with F-ARIMA.

Moreover, our method is also compared with the best method proposed in the literature, that is, Davies–Harte.

In all cases, the mean values shown in the tables were obtained by averaging the estimations over 100 independent traces of a length of one million.

Normality. The traces synthesized by both methods correspond to a Gaussian process for practical purposes. In order to analyze the marginal distribution we have used the χ^2 , Kolmogorov–Smirnov, and Anderson–Darling goodness-of-fit tests with satisfactory results, except for high values of H , where the tests usually reject many true hypotheses due to the presence of a strong LRD [Beran 1994]. However, visual tests such as the QQ- and PP-plots show that the traces “look” normal even in those extreme cases. In Figures 3 and 4 we can see that the plots for $H = 0.95$ fit a straight line with slope 1 that passes through the origin almost perfectly.

Finally, in relation to the mean and variance, we can observe in Table I that the traces generated by our method have, for practical purposes, zero mean and variance 1. We can also see that skewness and kurtosis are practically zero, indicating again the proximity to the normal distribution.

Instead, the Davies–Harte method exhibits certain problems for zero mean and variance 1, mainly as H increases (see Table II), although it passes all tests of normality. This may produce undesirable effects on changing marginal

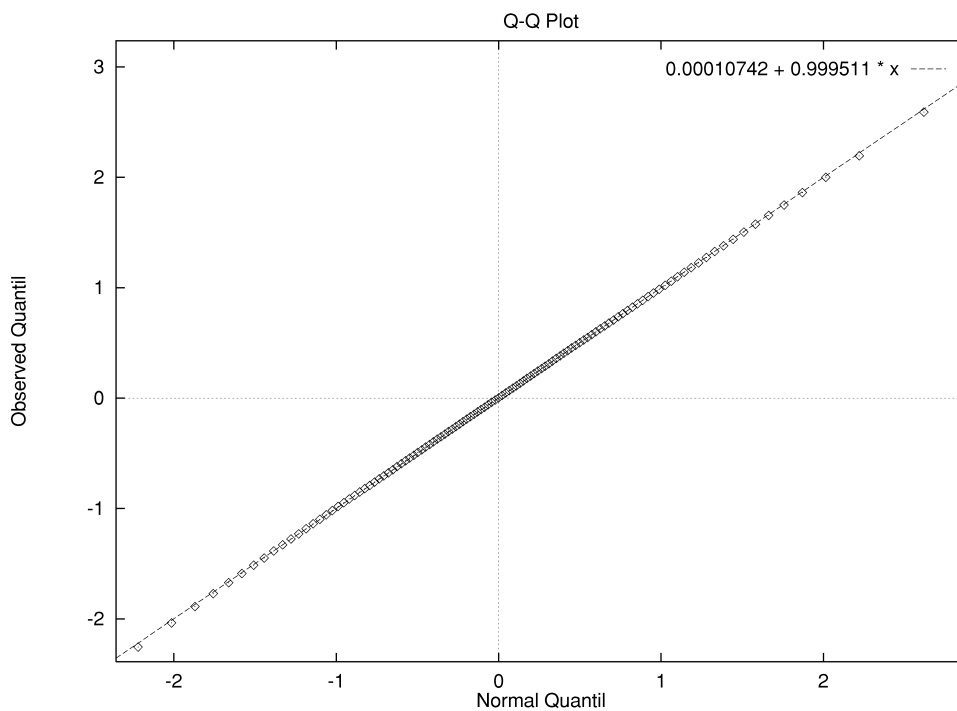


Fig. 3. Proposed method with $H = 0.95$: QQ-plot for normal distribution.

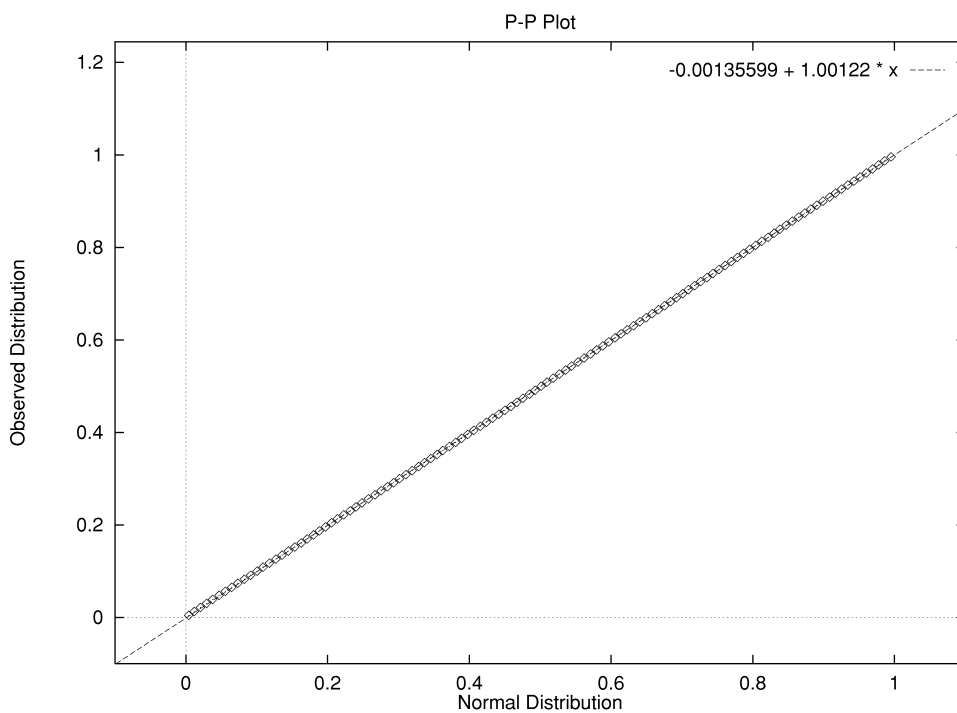


Fig. 4. Proposed method with $H = 0.95$: PP-plot for normal distribution.

Table I. Mean Values of Basic Statistics for Our Method

H	Sample Mean	Sample Deviation	Skewness	Kurtosis
0.500	-2.45072e-11	0.999833	-0.000395	0.001147
0.600	1.83910e-11	0.999952	0.000192	0.000829
0.700	1.11712e-11	0.999992	-0.000242	0.001021
0.800	-1.70470e-11	1.000080	-0.000275	-0.002498
0.900	-1.91762e-11	1.000680	-0.000316	-0.003279
0.950	1.07779e-11	1.004170	0.005729	-0.005913
0.975	8.05334e-12	1.000390	0.008572	-0.009112
0.990	9.42135e-12	0.985671	-0.009460	-0.030973

Table II. Mean Values of Basic Statistics for the Davies–Harte Method

H	Sample Mean	Sample Deviation	Skewness	Kurtosis
0.500	0.000209	0.999981	-0.000086	-0.001760
0.600	0.000516	0.999860	-0.000297	0.000147
0.700	0.002783	1.000010	-0.000717	-0.002089
0.800	0.000214	0.997575	-0.000816	0.000185
0.900	-0.019418	0.971073	0.005464	0.000464
0.950	-0.010069	0.885158	0.006671	-0.010367
0.975	0.002867	0.714338	-0.002490	-0.013063
0.990	0.043249	0.500878	-0.018778	0.001360

Table III. Estimations of H for Our Method by Whittle’s Estimator

H_{true}	Bias $\equiv \hat{H} - H_{\text{true}}$	$\hat{H}_i^{\text{MIN}} - \hat{H}_i^{\text{MAX}}$	$\hat{H}_i < H_{\text{true}}$ (%)	Out of CI (%)
0.5100	+0.0010	0.5068 – 0.5146	42	0
0.5250	+0.0007	0.5205 – 0.5322	54	5
0.5500	-0.0000	0.5459 – 0.5557	59	1
0.6000	+0.0000	0.5947 – 0.6064	46	4
0.6500	+0.0003	0.6455 – 0.6553	53	1
0.7000	-0.0003	0.6943 – 0.7080	46	3
0.7500	+0.0003	0.7471 – 0.7568	43	3
0.8000	-0.0001	0.7939 – 0.8076	61	4
0.8500	-0.0003	0.8447 – 0.8545	50	1
0.9000	+0.0001	0.8936 – 0.9053	42	4
0.9500	+0.0004	0.9443 – 0.9541	39	5
0.9750	+0.0003	0.9717 – 0.9814	44	0
0.9900	+0.0002	0.9854 – 0.9932	47	1

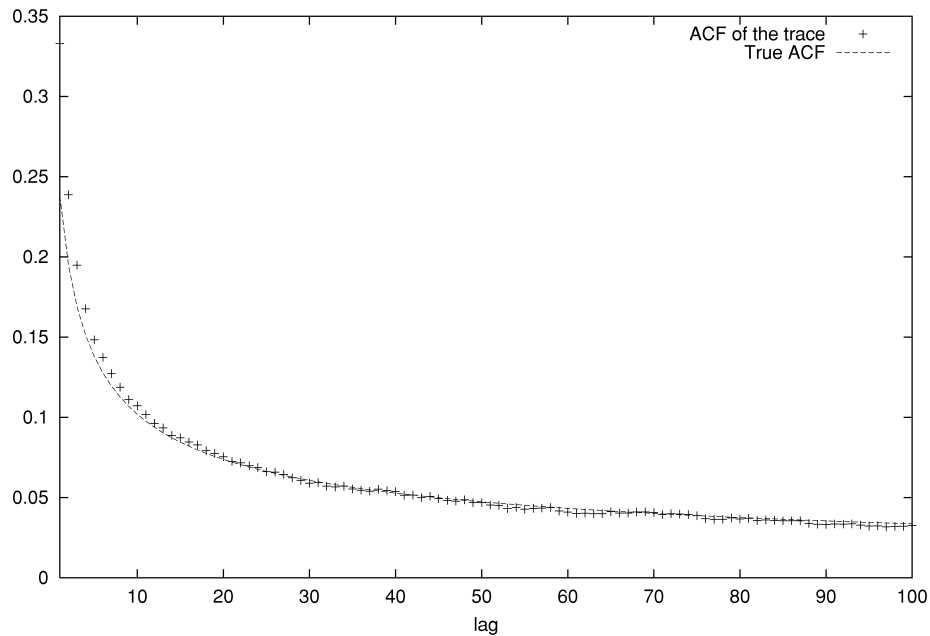
distribution, since it is necessary to know this distribution (parameters included) as accurately as possible.

Correlation Structure. Our method synthesizes high-quality traces, even higher than those generated by the Davies–Harte method. In Tables III and IV we can see the mean estimations of H, by averaging the estimations obtained after applying Whittle’s estimator to 100 traces. Our method gives rise to deviations from the true value that are slightly smaller than those for the Davies–Harte method. Besides, the new method does not apparently exhibit any trend (estimations are randomly spread on both sides of the true value) whereas the Davies–Harte method exhibits a slight trend. Finally, the number of estimations out of the 95% confidence interval was less with our method.

The reason why our method synthesizes traces with higher quality might be that this method obtains the periodogram from the exact expression of the power spectrum, whereas the Davies–Harte method obtains it by means of a DFT of the autocovariances.

Table IV. Estimations of H for the Davies–Harte Method by Whittle’s Estimator

H_{true}	Bias $\equiv \bar{\hat{H}} - H_{\text{true}}$	$\hat{H}_i^{\text{MIN}} - \hat{H}_i^{\text{MAX}}$	$\hat{H}_i < H_{\text{true}}$ (%)	Out of CI (%)
0.5100	+0.0004	0.5029 – 0.5146	31	5
0.5250	+0.0005	0.5225 – 0.5303	66	2
0.5500	–0.0004	0.5459 – 0.5537	32	1
0.6000	–0.0003	0.5947 – 0.6064	49	5
0.6500	+0.0007	0.6474 – 0.6553	51	1
0.7000	+0.0003	0.6963 – 0.7061	45	3
0.7500	–0.0004	0.7451 – 0.7549	61	3
0.8000	–0.0004	0.7939 – 0.8037	70	5
0.8500	+0.0009	0.8486 – 0.8525	21	2
0.9000	–0.0010	0.8955 – 0.9053	75	2
0.9500	+0.0009	0.9463 – 0.9561	30	8
0.9750	+0.0008	0.9697 – 0.9814	21	5
0.9900	–0.0002	0.9853 – 0.9932	48	3

Fig. 5. Simulated and true autocorrelation functions for F-ARIMA(0, d, 0) with $H = 0.75$.

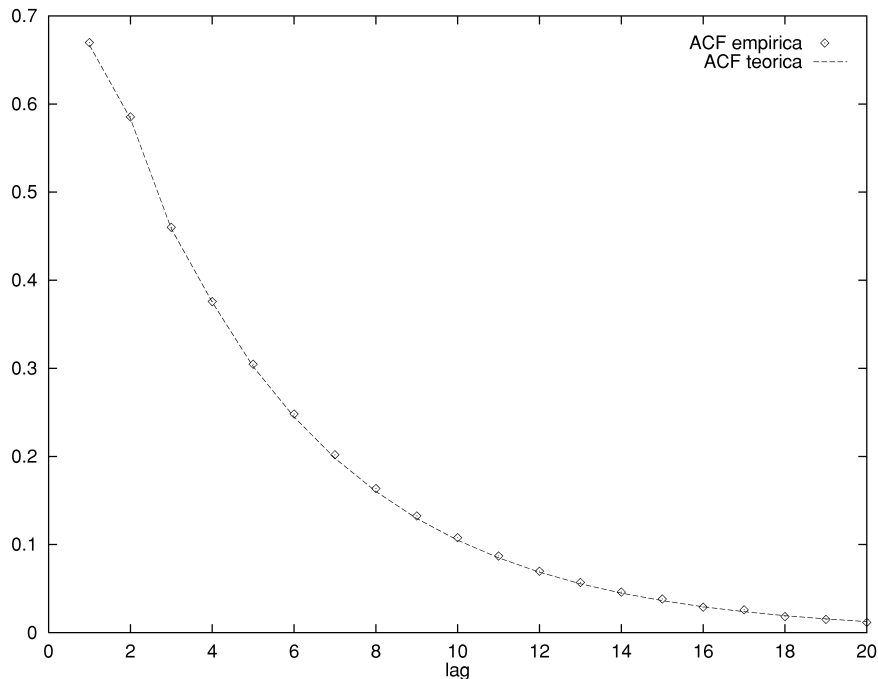
To see how the autocorrelation function of the traces fits the true correlations, we show an example of both functions for F-ARIMA(0, d, 0) in Figure 5. On the other hand, our new method allows us to synthesize any SRD process belonging to the ARMA(p, q) family. For instance, in Figure 6 we can also see both autocorrelation functions, simulated and true, for an AR(2) process.

Spectral Density. Both methods easily pass the strict Beran goodness-of-fit test for the spectral density [Beran 1992]. As can be seen in Table V, the percentage of rejections is practically always lower than or equal to the level of significance.

Visual Test. Finally, it would also be interesting to observe if the synthesized traces pass the visual test, namely, if the aggregation preserves the distribution, comparing this behavior with traces corresponding to non-LRD processes. Thus, in Figure 7 we represent three traces on different time scales

Table V. Results of Beran's Test for Both Methods ($\alpha = 0.05$)

H	Proposed Method			Davies–Harte Method		
	\bar{t}	P-Value	Rejections (%)	\bar{t}	P-Value	Rejections (%)
0.500	0.318424	0.399942	3	0.318245	0.557305	2
0.525	0.318362	0.453918	4	0.317952	0.786700	2
0.550	0.318286	0.521159	1	0.318512	0.326722	8
0.600	0.318100	0.679482	4	0.318277	0.529119	4
0.650	0.317994	0.758574	0	0.318369	0.447762	8
0.700	0.318458	0.371068	8	0.318330	0.482181	6
0.750	0.318534	0.460968	5	0.318590	0.266887	6
0.800	0.318514	0.325121	6	0.318252	0.551159	2
0.850	0.318171	0.621160	5	0.318004	0.751592	2
0.900	0.318171	0.621160	3	0.318430	0.394802	6
0.950	0.318380	0.438114	6	0.318299	0.509647	4
0.975	0.318122	0.661799	3	0.318236	0.565187	2

Fig. 6. Simulated and true autocorrelation functions for AR(2) with $\alpha_1 = 0.5$ and $\alpha_2 = 0.25$.

(aggregation levels): one generated with the proposed method, corresponding to a F-ARIMA(0, d, 0) process with $H = 0.9$ (on the left); another one generated with the same method corresponding to a SRD process, ARMA(2, 2) (in the middle), and a sequence of a pure random process (on the right). All sequences are composed of 10 million samples, with exponential marginal distribution, so that all values are positive.

We can see as in the case of F-ARIMA(0, d, 0), in effect, the increase of the aggregation level gives rise to plots that are intuitively very “similar” (in a distributional sense). Instead, this effect is not observed in the other two cases,

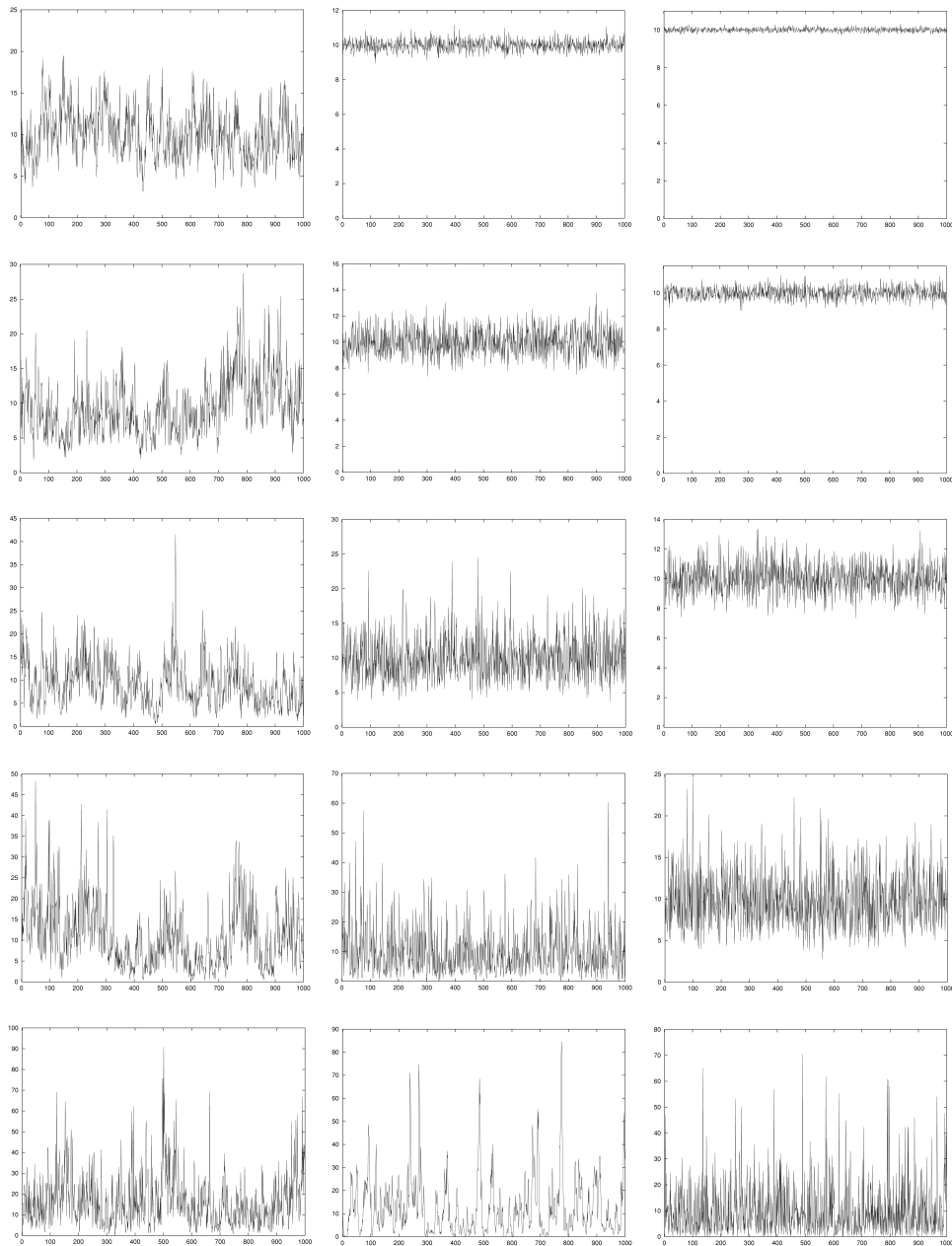


Fig. 7. Representation in different time scales (1, 10, 100, 1,000, and 10,000) of a synthesized trace of a F-ARIMA(0, d , 0) process (left), an ARMA(2, 2) process (middle), and a pure random process (right).

where the points are concentrated closer and closer to the mean as the aggregation level increases.

To sum up, the main conclusion is that the new method to synthesize F-ARIMA, proposed in this article, is the best choice due to its excellent computational efficiency for generating high-quality traces that correspond to a F-ARIMA process for practical purposes. Moreover, the method permits us to synthesize a generic process F-ARIMA(p, d, q) without applying any additional ARMA(p, q) filter to F-ARIMA(0, $d, 0$) traces.

7. CONCLUSIONS

This article deals with new problems arising in the application of self-similar models to performance analysis. Undoubtedly, one of the most important problems is the efficient synthetic generation of high-quality traces corresponding to self-similar processes. The existing real traces are not suitable, since they do not allow us to modify their correlation structure or other statistical properties of interest and, in most cases, the sample size is not enough to achieve certain statistical requirements. In addition, simulation experiments in the context of LRD will usually require very long independent sample paths, and therefore synthetic generation is essential.

Moreover, although a great part of this article is devoted to synthetic generation of self-similar processes, other interesting issues related to network simulation and self-similar processes are also dealt with:

- the drawbacks of using models based in counting processes;
- the existence of a duality in the LRD structure between counting and inter-arrival time processes, that allows us to use the latter;
- the selection of Gaussian self-similar processes in order to be able to change their marginal distribution without altering their long-range correlation structure;
- the selection of Gaussian F-ARIMA(p, d, q) processes, instead of FGN, due to their flexibility to model SRD and LRD simultaneously.

Finally, with regard to the synthetic generation of Gaussian F-ARIMA, the main conclusions of this article are as follows.

- Due to the large number of long runs that must be used in an exhaustive simulation study in the context of LRD, the time spent in synthesizing a very large trace is critical. For this reason, exact methods must be discarded since the computational cost is of the order of the sample size squared.
- The computational cost can be reduced significantly if we make use of approximate methods that, however, require consideration of an additional criterion: the goodness of the approximation, evaluated by the quality of the synthesized traces. However, the approximate methods proposed in the literature for synthesizing F-ARIMA are less efficient than those for FGN (Paxson's method, mainly). For this reason, despite their flexibility, F-ARIMA processes have been little used in simulation studies and, especially, if the SRD structure is not modeled.
- The new approximate method proposed in this article can very efficiently generate high-quality traces corresponding, for practical purposes, to a Gaussian

F-ARIMA(p, d, q) process. A comparison with existing methods shows that the new method is significantly more efficient, even more than the best method for FGN. So, we trust that the computational efficiency is not an obstacle for using much more flexible models based in F-ARIMA(p, d, q).

ACKNOWLEDGMENTS

We would like to acknowledge the anonymous TOMACS reviewers for their constructive criticisms and suggestions for improving the presentation of this article.

REFERENCES

- ADAS, A. AND MUKHERJEE, A. 1996. Delay-jitter bound and statistical loss bound for heterogeneous correlated traffic—architecture and equivalent bandwidth. Technical Report GIT-CC-96-05 (April), Georgia Institute of Technology, Atlanta, GA.
- BERAN, J. 1992. A goodness-of-fit test for time series with long range dependence. *J. Royal Statist. Soc., Series B* 54, 3, 749–760.
- BERAN, J. 1994. *Statistics for Long-Memory Processes*, Volume 61 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.
- BERAN, J., SHERMAN, R., TAQQU, M. S., AND WILLINGER, W. 1995. Long-range dependence in variable-bit-rate video traffic. *IEEE Trans. Commun.* 43, 2/3/4, 1566–79.
- BOX, G. E. P., JENKINS, G. M., AND REINSEL, C. 1994. *Time Series Analysis: Forecasting and Control* (third ed.). Prentice Hall, Englewood Cliffs, NJ.
- COX, D. R. 1984. Long range dependence: A review. In *Statistics: An Appraisal*, H. A. David and H. T. David, Eds., 55–74. Iowa State University Press, Ames, IA.
- COX, D. R. AND ISHAM, V. 1980. *Point Processes*. Chapman and Hall, London.
- CROVELLA, M. E. AND BESTAVROS, A. 1997. Self-similarity in World Wide Web traffic: Evidence and possible causes. *IEEE/ACM Trans. Netw.* 5, 6 (Dec.), 835–846.
- DAVIES, R. B. AND HARTE, D. S. 1987. Tests for Hurst effect. *Biometrika* 74, 95–102.
- ERRAMILI, A., NARAYAN, O., AND WILLINGER, W. 1996. Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Trans. Netw.* 4, 209–223.
- ERRAMILI, A., SINGH, R. P., AND PRUTHI, P. 1995. An application of deterministic chaotic maps to model packet traffic. *Queueing Syst.* 20, 171–206.
- FELDMANN, A., GILBERT, A. C., HUANG, P., AND WILLINGER, W. 1999. Dynamics of IP traffic: A study of the role of variability and the impact of control. In *Proceedings of ACM SIGCOMM '99*, 301–313.
- FENDICK, K. W., SAKSENA, V. R., AND WHITT, W. 1989. Dependence in packet queues. *IEEE Trans. Commun.* 37, 1173–1183.
- GARRETT, M. W. AND WILLINGER, W. 1994. Analysis, modeling and generation of self-similar VBR video traffic. In *Proceedings of ACM SIGCOMM '94* (London), 269–280.
- GRAF, H. P. 1983. Long-range correlations and estimation of the self-similarity parameter. Ph.D. Thesis, ETH Zürich.
- GRANGER, C. W. J. 1980. Long memory relationships and the aggregation of dynamic models. *J. Econometrics* 14, 227–238.
- GRANGER, C. W. J. AND JOYEUX, R. 1980. An introduction to long-range time series models and fractional differencing. *J. Time Series Anal.* 1, 15–30.
- GRASSE, M., FRATER, M. R., AND ARNOLD, J. F. 1997. Origins of long-range dependence in variable bit rate video traffic. In *Proceedings of the Fifteenth International Teletraffic Congress ITC-15*, V. Ramaswami and P. E. Wirth, Eds., Elsevier Science B. V., 1379–1388.
- GROSSGLAUSER, M. AND BOLOT, J.-C. 1996. On the relevance of long-range dependence in network traffic. In *Proceedings of ACM SIGCOMM '96* (Stanford University, CA, August), 15–24.
- HASLETT, J. AND RAFTERY, A. E. 1989. Space-time modeling with long-memory dependence: Assessing Ireland's wind power resource (with discussion). *App. Stat.* 38, 1–50.
- HEYMAN, D. P. AND LAKSHMAN, T. V. 1996. What are the implications of long-range dependence for VBR-video traffic engineering? *IEEE/ACM Trans. Netw.* 4, 3 (June), 301–317.

- HEYMAN, D. P., TABATABAI, A., AND LAKSHMAN, T. V. 1991. Statistical analysis and simulation study of video teleconferencing traffic in ATM networks. In *Proceedings IEEE GLOBECOM '91*, 21–27. Also in *IEEE Trans. Circuits Syst. Video Technol.*, 2, 1, (March).
- HOSKING, J. R. M. 1981. Fractional differencing. *Biometrika* 68, 1, 165–176.
- HOSKING, J. R. M. 1984. Modeling persistence in hydrological time series using fractional differencing. *Water Resources Res.* 20, 12, 1898–1908.
- HUANG, C., DEVETSIKIOTIS, M., LAMBADARIS, I., AND KAYE, A. R. 1995a. Fast simulation for self-similar traffic in ATM networks. In *Proceedings of IEEE ICC '95* (Seattle, Sept.), 438–444.
- HUANG, C., DEVETSIKIOTIS, M., LAMBADARIS, I., AND KAYE, A. R. 1995b. Modeling and simulation of self-similar variable bit rate compressed video: A unified approach. In *Proceedings of ACM SIGCOMM '95* (Cambridge, MA), 114–125.
- HURST, H. E. 1951. Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civil Eng.* 116, 770–799.
- KAPLAN, L. M. AND KUO, C.-C. J. 1994. Extending self-similarity for fractional Brownian motion. *IEEE Trans. Signal Proc.* 42, 12 (Dec.), 3526–3530.
- KLEINROCK, L. 1975. *Queueing Systems*. Wiley, New York.
- KRUNZ, M. AND MAKOWSKI, A. 1998. Modeling video traffic using $M/G/\infty$ input processes: A compromise between Markovian and LRD models. *IEEE J. Selected Areas Commun.* 16, 5, 733–748.
- KRUNZ, M., SASS, R., AND HUGHES, H. 1995. Statistical characteristics and multiplexing of MPEG streams. In *Proceedings of IEEE INFOCOM '95* (Boston, April), 455–462.
- LAU, W.-C., ERRAMILI, A., WANG, J. L., AND WILLINGER, W. 1995. Self-similar traffic generation: The random midpoint displacement algorithm and its properties. In *Proceedings of IEEE ICC '95* (Seattle, Sept.), 466–472.
- LELAND, W. E., TAQQU, M. S., WILLINGER, W., AND WILSON, D. V. 1993. On the self-similar nature of Ethernet traffic. *Comput. Commun. Rev.* 23, 183–193.
- LELAND, W. E., TAQQU, M. S., WILLINGER, W., AND WILSON, D. V. 1994. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans. Netw.* 2, 1 (Feb.), 1–15.
- LI, S. Q. AND HWANG, C. L. 1993. Queue response to input correlation functions: Continuous spectral analysis. *IEEE/ACM Trans. Netw.* 1, 6 (Dec.), 678–692.
- LIKHANOV, N., TSYBAKOV, B., AND GEORGANAS, N. 1995. Analysis of an ATM buffer with self-similar (“fractal”) input traffic. In *Proceedings of IEEE INFOCOM '95* (Boston, April), 985–992.
- LÓPEZ-ARDAO, J., SUÁREZ, A., AND LÓPEZ-GARCÍA, C. 1998. On the effects of long-range dependence on the performance of FDDI. In *Proceedings of the Seventh IEEE Workshop on Computer-Aided Modeling, Analysis and Design of Communication Links and Networks (CAMAD'98)* (Sao Paulo, Brazil, August 13).
- LÓPEZ-ARDAO, J. C. 1999. Contribución al análisis del impacto de la correlación en las prestaciones de las redes de alta velocidad. Ph.D. Thesis, Universidad de Vigo (Spain).
- MANDELBROT, B. 1967. Some noises with $1/f$ spectrum, a bridge between direct current and white noise. *IEEE Trans. Inf. Theory* 13, 2 (April), 289–298.
- MANDELBROT, B. B. AND NESS, J. W. V. 1968. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* 10, 4 (Oct.), 422–437.
- NORROS, I. 1994. A storage model with self-similar input. *Queueing Syst.* 16, 387–396.
- PARULEKAR, M. 1997. Buffer engineering for self-similar traffic. Ph.D. Thesis, College Park, University of Maryland.
- PARULEKAR, M. AND MAKOWSKI, A. M. 1996. Tail probabilities for a multiplexer with self-similar traffic. In *Proceedings of IEEE INFOCOM '96* (San Francisco, March), 1452–1459.
- PAXSON, V. 1997. Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic. *Compu. Commun. Rev.* 27, 5 (Oct.), 5–18.
- PAXSON, V. AND FLOYD, S. 1995. Wide-area traffic: The failure of Poisson modeling. *IEEE/ACM Trans. Netw.* 3, 226–244.
- RYU, B. K. AND ELWALID, A. 1996. The importance of long-range dependence of VBR video traffic in ATM traffic engineering: Myths and realities. In *Proceedings of ACM SIGCOMM '96* (Stanford University, CA, August), 3–14.
- SAMORODNITSKY, G. AND TAQQU, M. S. 1994. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, London.
- STATISTICAL SCIENCES, INC. 1991. The S+ package, Version 3.0.

- SUÁREZ, A., LÓPEZ-ARDAO, J., LÓPEZ-GARCÍA, C., RODRÍGUEZ-RUBIO, R., AND SOUSA, M. E. 1999. Concurrent methodology on LRD simulation studies. In *proceedings of the Thirteenth European Simulation Multiconference (ESM '99)* (Warsaw, Poland, May), 121–128.
- TSYBAKOV, B. AND GEORGANAS, N. D. 1997. On self-similar traffic in ATM queues: Definitions, overflow probability bound, and cell delay distribution. *IEEE/ACM Trans. Netw.* 5, 3 (June), 397–409.
- WILLINGER, W. 1995. Traffic modeling for high-speed networks: Theory versus practice. In *Stochastic Networks*, Vol. 71 of *IMA Volumes in Mathematics and its Applications*, F. Kelly and R. Williams, Eds., 395–409, Springer-Verlag, New York.
- WILLINGER, W., TAQQU, M. S., AND ERRAMILI, A. 1996. A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks. In *Stochastic Networks: Theory and Applications*, F. P. Kelly, S. Zachary, and I. Ziedins, Eds., Oxford University Press, New York.
- WILLINGER, W., TAQQU, M. S., LELAND, W. E., AND WILSON, D. V. 1995. Self-similarity in high speed packet traffic: Analysis and modeling of Ethernet traffic measurements. *Statis. Sci.* 10, 67–85.

Received November 1999; Revised April 2000; Accepted May 2000