

ADVANCED MODELS AND ALGORITHMS FOR SELF-SIMILAR IP NETWORK TRAFFIC SIMULATION AND PERFORMANCE ANALYSIS

Dimitar Radev* — Izabella Lokshina**

The paper examines self-similar (or fractal) properties of real communication network traffic data over a wide range of time scales. These self-similar properties are very different from the properties of traditional models based on Poisson and Markov-modulated Poisson processes. Advanced fractal models of sequential generators and fixed-length sequence generators, and efficient algorithms that are used to simulate self-similar behavior of IP network traffic data are developed and applied. Numerical examples are provided; and simulation results are obtained and analyzed.

Key words: communication networks, IP network traffic, long-range dependent self-similar processes, advanced generators of self-similar teletraffic

1 INTRODUCTION

The growth of broadband networks and the Internet has been exponential in terms of users and user-end systems as well as in traffic in recent years. High-speed communications networks are able to support a wide range of multimedia applications, such as audio, video and computer data that differ significantly in their traffic characteristics and performance requirements [6].

One possible goal for telecommunications developers is to build a unified high-speed communication network platform capable of carrying diverse traffic and supporting diverse levels of quality of service (QoS). Recent studies of real communication network traffic data have shown that teletraffic (technical term, identifying all phenomena of control and transport of information within the telecommunications networks) exhibits self-similar (or fractal) properties over a wide range of time scales [3].

The properties of self-similar IP network traffic are very different from properties of traditional models based on Poisson, Markov-modulated Poisson, and related processes [16]. The use of traditional models in networks characterized by self-similar processes can lead to incorrect conclusions about the performance of analyzed networks [5]. Traditional models can lead to over-estimation of the network performance, insufficient allocation of communication and data processing resources, and consequently difficulties in ensuring the QoS [14].

Self-similarity can be classified into two types: deterministic and stochastic [8]. In the first type, deterministic self-similarity, a mathematical object is assumed to be self-similar (or fractal) if it can be decomposed into smaller copies of itself [11]. That is, self-similarity is a property, in which the structure of the whole is contained in its parts. This work is focused on stochastic self-similarity. In that case, probabilistic properties of

self-similar processes remain unchanged or invariant when the process is viewed at different time scales. This is in contrast to Poisson processes that lose their burstiness and flatten out when time scales are changed. However, the time series of self-similar processes exhibit burstiness over a wide range of time scales. Self-similarity can statistically describe teletraffic that is bursty on many time scales.

One can distinguish two types of stochastic self-similarity. A continuous-time stochastic process \mathbf{Y}_t is strictly self-similar with a self-similarity parameter H ($1/2 < H < 1$), if \mathbf{Y}_{ct} and $c^H \mathbf{Y}_t$ (the rescaled process with time scale ct) have identical finite-dimensional probability for any positive time stretching factor c [9]. This definition, in a sense of probability distribution, is quite different from that of the second-order self-similar process, observed at the mean, variance and autocorrelation levels. The process \mathbf{X} is asymptotically second-order self-similar with $0.5 < H < 1$, if for each k large enough $\rho_k^{(m)} \rightarrow \rho_k$, as $m \rightarrow \infty$, where $\rho_k = E[(X_i - \mu)(X_{i+k} - \mu)]/\sigma^2$.

In this work the exact or asymptotic self-similar processes are used in an interchangeable manner, which refers to the tail behavior of the autocorrelations [7]. Modeling and simulation of self-similar IP network traffic is performed with advanced generators of synthetic self-similar sequences, divided into two practical classes: the sequential generators and the fixed-length sequence generators. Both classes of generators are considered in this study.

2 DISTINCTIVE PROPERTIES OF LONG-RANGE DEPENDENT SELF-SIMILAR PROCESSES

In recent communication networks the most striking feature of some second-order self-similar processes is that

* University of Rouse 8, Studentska Street 7017, Rouse, Bulgaria, dradev@abv.bg;

** SUNY Oneonta 324B Netzer Administration Bldg. Oneonta, NY 13820, USA, lokshiiv@oneonta.edu

the accumulative functions of the aggregated processes do not generate as the non-overlapping batch size m increasing to infinity. Such processes are known as Long-Range Dependent (LRD) processes [1]. This is in contrast to traditional processes used in modeling IP network traffic, all of which include the property that the accumulative functions of their aggregated processes degenerate as the non-overlapping batch size m increasing to infinity, *ie*, $\rho_k^{(m)} \rightarrow 0$ or $\rho_k^{(m)} = 0$ ($|k| > 0$), for $m > 1$. The equivalent definition of long-range dependence is given as (1).

$$\sum_{k=-\infty}^{\infty} \rho_k = \infty. \tag{1}$$

Another definition of LRD is presented as (2),

$$\rho_k \sim L(t)k^{-(2-2H)}, \text{ as } k \rightarrow \infty \tag{2}$$

where $1/2 < H < 1$ and $L(\cdot)$ slowly varies at infinity, *ie* for all $x > 0$ it could be determined as (3).

$$\lim_{t \rightarrow \infty} \frac{L(xt)}{L(t)} = 1. \tag{3}$$

The Hurst parameter H characterizes the relation in (2), which specifies the form of the tail of the accumulative function. One can show that is true for $1/2 < H < 1$, as given in (4).

$$\rho_k = \frac{1}{2} [(k+1)^{2H} - k^{2H} + (k-1)^{2H}]. \tag{4}$$

For $0 < H < 1/2$ the process is Short-Range Dependent (SRD) and could be presented as (5).

$$\sum_{k=-\infty}^{\infty} \rho_k = 0. \tag{5}$$

For $H = 1$ all autocorrelation coefficients are equal to one, no matter how far apart in time the sequences are. This case has no practical importance in real telecommunications network traffic modeling. If $H > 1$, then (6) is true.

$$\rho + k = \begin{cases} 1 & \text{for } k = 0, \\ \frac{1}{2}k^{2H}g(k^{-1}) & \text{for } k > 0 \end{cases} \tag{6}$$

where

$$g(x) = (1+x)^{2H} - 2 + (1-x)^{2H}. \tag{7}$$

One can see that $g(x) \rightarrow \infty$ as $H > 1$. If $0 < H < 1$ and $H \neq 1/2$, then the first non-zero term in the Taylor expansion of $g(x)$ is equal to $2H(2H-1)x^2$. Therefore, (8) is true.

$$\frac{\rho_k}{H(2H-1)k^{2H-2}} \rightarrow 1 \text{ as } k \rightarrow \infty. \tag{8}$$

In the frequency domain, an essentially equivalent definition of LRD for a process X with given spectral density (9),

$$f(\lambda) = \frac{\sigma^2}{2\pi} \sum_{k=-\infty}^{\infty} \rho_k e^{ik\lambda} \tag{9}$$

is that in the case of LRD processes, this function is required to satisfy the following property (10),

$$f(\lambda) \sim c_{f_1} \lambda^{-\gamma}, \text{ as } \lambda \rightarrow \infty \tag{10}$$

where c_{f_1} is a positive constant and $0 < \gamma < 1$, $\gamma = 2H - 1 < 1$.

As a result, LRD manifests itself in the spectral density that obeys a power-law in the vicinity of the origin. This implies that $f(0) = \sum_k \rho_k = \infty$. Consequently, it requires a spectral density, which tends to $+\infty$ as the frequency λ approaches 0.

For a Fractional Gaussian Noise (FGN) process, the spectral density $f(\lambda, H)$ is given by (11),

$$f(\lambda, H) = 2c_f(1 - \cos \lambda)B(\lambda, H) \tag{11}$$

with $0 < H < 1$ and $-\pi \leq \lambda \leq \pi$, where (12) is true,

$$c_f = \sigma^2(2\pi)^{-1} \sin(\pi H)\Gamma(2H + 1),$$

$$B(\lambda, H) = \sum_{k=-\infty}^{\infty} |2\pi i k + \lambda|^{-2H-1} \tag{12}$$

and $\sigma^2 = \text{Var}[X_k]$ and $\Gamma(\cdot)$ is the gamma function.

The spectral density $f(\lambda, H)$ in (11) complies with a power-law at the origin, as shown in (13),

$$f(\lambda, H) \rightarrow c_f \lambda^{1-2H}, \text{ as } \lambda \rightarrow 0 \tag{13}$$

where $1/2 < H < 1$.

3 ADVANCED GENERATORS OF SELF-SIMILAR IP NETWORK TRAFFIC

3.1 Sequential generators of self-similar IP network traffic

We could make a sequential Markovian model that imitates a self-similar sequence. However, this approach would have a disadvantage because of connection between the model parameters and its self-similar properties that is difficult to understand.

Markovian models for self-similar traffic require including several control parameters with a wide range of input values. As a result, controlling these values in sequential generators is much more complex than in generators of fixed-length sequences of self-similar processes with a given Hurst parameter [15].

We considered the following efficient candidate sequential generators, based on:

- Fractal-Binomial-Noise-Driven Poisson processes;
- Superposition of Fractal Renewal Processes (SFRP);
- $M/G/\infty$ Processes (MGIP);
- Pareto-Modulated Poisson Processes (PMPP);
- Spatial Renewal Processes & Fractional Gaussian Noise;
- Superposition of Autoregressive Processes (SAP).

3.1.1 Model based on fractal-binomial-noise-driven Poisson processes

For the standard Fractal Renewal Process (FRP), inter- event times are independent random variables [1]. The marginal Probability Density Function (PDF) of such a fractal renewal process can be defined as (1),

$$f(t) = \begin{cases} 0, & t \leq A, \\ \delta A^\delta t^{-(\delta+1)}, & t > A \end{cases} \quad (14)$$

where $0 < \delta < 2$. Selecting δ in this interval proves far superior to $0 < \delta < 1$ for the same required values that the inter-event time PDF can be further improved [2].

The improved PDF of the FRP decays as a power law, as shown in (15),

$$f(t) = \begin{cases} \delta A^{-1} e^{-\delta t/A}, & 0 < t \leq A, \\ \delta e^{-\delta} A t^{-(\delta+1)}, & t > A, \end{cases} \quad (15)$$

which is continuous for all t , producing smoother spectral density function [12].

A method based on the Fractal-Binomial-Noise-Driven Poisson Processes (FBNDP) adds M independent and identically distributed (iid) alternating FRPs to generate a fractal binomial noise process that serves as the rate function for a Poisson process. The FBNDP requires five input parameters to generate self-similar sequences: A , δ , R , Δt , and M . The resulting Hurst parameter H assumes the value $(\alpha + 1)/2$.

The suggested algorithm is advancing with the intervals Δt . If S is a simulation clock, which advances in time, and $S^{(j)}$ is the elapsed time of the j -th FRP sequence, then $S^{(j)} = \tau_0^{(j)} + \tau_1^{(j)} + \dots + \tau_k^{(j)}$ for some k and $j = 1, 2, \dots, M$, where $\tau_k^{(j)}$ is the inter-arrival time. The sequence of self-similar pseudo- random numbers X_0, X_1, \dots is generated through the following process.

Algorithm 1:

Step 1: For each $j = 1, 2, \dots, M$, generate $\tau_0^{(j)}$ from (16),

$$\tau_0^{(j)} = \begin{cases} -\delta^{-1} A \log(U\delta^{V-1}\delta - (V - U)), & V \geq 1, \\ AV^{1/(1-\delta)}, & V < 1, \end{cases} \quad (16)$$

where

$$V \equiv \frac{1 + (\delta - 1)e^\delta U}{\delta} \quad (17)$$

and U is an iid uniformly distributed random variable over the unit interval $[0, 1]$; set $S^{(j)} = \tau_0^{(j)}$.

Step 2: Find j^* and $S^{(j^*)}$ such that $j^* = \operatorname{argmin}_j \{S^{(j)}\}$.

Step 3: Calculate (18).

$$x = \begin{cases} 0, & \text{if } S^{(j^*)} < A, \\ 1, & \text{if } S^{(j^*)} \geq A. \end{cases} \quad (18)$$

Step 4: If $x = 1$, then X_0 should be drawn from a Poisson probability distribution with $\lambda = 1$. If $x = 0$, then $X_0 = 0$.

Step 5: Set $i = 1$, and $y = 0$. Advance the simulation clock, ie, $S \leftarrow S^{(j^*)}$.

Step 6: Construct a new inter-event time $\tau_i^{(j)}$ from (19),

$$\tau_i^{(j)} = \begin{cases} -\delta^{-1} A \log U, & U \geq e^{-\delta}, \\ e^{-1} A U^{-1/\delta}, & U < e^{-\delta}, \end{cases} \quad (19)$$

and set $S^{(j^*)} \leftarrow S^{(j^*)} + \tau_i^{(j)}$.

Step 7: Find a new j^* such that $j^* = \operatorname{argmin}_j \{S^{(j)}\}$, and compute $S^{(j)} - S$.

Step 8: Repeat Step 6 through Step 8 to obtain x as in (18).

Step 9: Advance the simulation clock, ie, $S \leftarrow S^{(j^*)}$, and set $y = y + x$.

Step 10: Repeat Step 6 through Step 10 within time slot of length Δt .

Step 11: Compute $X_i = \text{POISS}(y)$, set $y = 0$, and $i = i + 1$.

Step 12: Repeat Step 6 through Step 11 until $i = n$, where n is the number of sample points.

An approximate self-similar sequence $\{X_0, X_1, X_2, \dots\}$ was obtained. It took 1 min 38 sec on the Centrino-based HP PC (1.76 GHz) to generate a traffic sample sequence of 1 048 576 numbers (eg, about 524×10^6 inter-event times). Table 1 shows the mean values of estimated Hurst parameters H and 95% confidence intervals for the mean in parentheses, obtained using the wavelet-based H estimator for the FBNDP method with four input values: $A = 9.92$, $R = 200$, $M = 4$ to 14, and $H = 0.6$ to 0.9. The results confirm that the most appropriate aggregate level is $M = 10$.

3.1.2 Model based on superposition of fractal renewal processes

A method based on the Superposition of the Fractal Renewal Processes (SFRP) uses a group of independent and identical fractal renewal processes (FRP). This method requires three parameters, ie, α and A from the individual FRPs, and M , which is the number of FRPs superposed.

Table 1. Accuracy of sequences generated with FBNDP

M	Mean values of estimated H			
	0.6	0.7	0.8	0.9
4	0.6023 (0.575, 0.630)	0.6864 (0.659, 0.714)	0.7817 (0.754, 0.809)	0.8469 (0.819, 0.874)
6	0.6085 (0.581, 0.636)	0.6871 (0.660, 0.715)	0.7804 (0.753, 0.808)	0.8496 (0.822, 0.877)
8	0.6080 (0.581, 0.636)	0.6870 (0.660, 0.715)	0.7803 (0.753, 0.808)	0.8489 (0.821, 0.876)
10	0.6089 (0.581, 0.636)	0.6875 (0.660, 0.715)	0.7827 (0.755, 0.810)	0.8502 (0.823, 0.878)
12	0.6053 (0.578, 0.633)	0.6903 (0.663, 0.718)	0.7832 (0.756, 0.811)	0.8501 (0.823, 0.878)
14	0.6049 (0.577, 0.632)	0.6895 (0.662, 0.717)	0.7842 (0.757, 0.812)	0.8497 (0.822, 0.877)

Table 2. Accuracy of sequences generated with SFRP

M	Mean values of estimated H			
	0.6	0.7	0.8	0.9
4	0.6076 (0.580, 0.635)	0.6986 (0.671, 0.726)	0.7929 (0.765, 0.820)	0.8603 (0.833, 0.888)
6	0.6076 (0.580, 0.635)	0.6986 (0.671, 0.726)	0.7929 (0.765, 0.820)	0.8655 (0.833, 0.888)
8	0.6192 (0.592, 0.647)	0.7084 (0.681, 0.736)	0.7997 (0.772, 0.827)	0.8655 (0.838, 0.893)
10	0.6166 (0.589, 0.644)	0.7091 (0.682, 0.737)	0.7986 (0.771, 0.826)	0.8686 (0.841, 0.896)
12	0.6134 (0.586, 0.641)	0.7085 (0.681, 0.736)	0.7978 (0.770, 0.825)	0.8718 (0.844, 0.899)
14	0.6143 (0.587, 0.642)	0.7063 (0.679, 0.734)	0.7983 (0.771, 0.826)	0.8708 (0.843, 0.898)

The resulting Hurst parameter H , and the mean μ and the variance σ^2 of the marginal output distribution in related counting process during the unit time interval, can be defined according to (20)

$$\begin{aligned}
 H &= \frac{\alpha + 1}{2}, \quad \mu = E[X_n] = \lambda, \\
 \sigma^2 &= \text{Var}[X_n] = (1 + (1/t_0))^\alpha \lambda
 \end{aligned}
 \tag{20}$$

where

$$\lambda = M\delta [1 + (\delta - 1)^{-1} e^{-\delta}]^{-1} A^{1-\delta}
 \tag{21}$$

is an aggregate arrival rate of events in the unit time interval. The sequence of self-similar pseudo-random numbers X_0, X_1, \dots is generated with the following procedure.

Algorithm 2:

Step 1: For each $j = 1, 2, \dots, M$, and $i = 0$, generate $\tau_i^{(j)}$ from equations (3) and (4); set $S^{(j)} = \tau_0^{(j)}$.

Step 2: Find j^* such that $j^* = \text{argmin}_j S^{(j)}$ and set $X_0 = S^{(j^*)}$.

Step 3: Advance the simulation clock, ie $S \leftarrow S^{(j^*)}$.

Step 4: Set $i = i + 1$. Construct a new inter-event time $\tau_i^{(j)}$ from equation (6) and set $S^{(j^*)} \leftarrow S^{(j^*)} + \tau_i^{(j)}$.

Step 5: Find a new j^* such that $j^* = \text{argmin}_j S^{(j)}$, and compute $S^{(j)} - S$.

Step 6: Advance the simulation clock, ie $S \leftarrow S^{(j^*)}$.

Step 7: Repeat Step 4 through Step 6 until $i = n$ is reached, where n is the number of sample points.

The obtained results from the simulation with $M = 10$, $A = 3.8$ and other conditions that similar to the FBNDP method, and with the mean values of estimated H and 95% confidence intervals for the mean in parentheses, are presented in Table 2.

The mean values for 0.6 and 0.7 of estimated H with 95% confidence intervals for the means in parentheses for FBNDP algorithm are 0.6086 (0.581, 0.636) and 0.6875 (0.66, 0.715), respectively. Their $\Delta H(\%)$ is +1.44 and -1.789, respectively.

For $H = 0.6, 0.7$ and 0.8 the overall relative error is less than 5%, but for $H = 0.9$ it is greater than 5%. The same characteristics, obtained with SFRP algorithm are 0.6166 (0.589, 0.644) and 0.7091 (0.682, 0.737), and their $\Delta H(\%)$ are +2.759 and -1.296, respectively.

The SFRP approach demonstrates more accurate results and the relative error for $H = 0.6, 0.7, 0.8$ and 0.9 is +2.76%, +1.29%, -0.17% and -3.49%, respectively. Similar to FBNDP algorithm, H values, estimated with SFRP algorithm, ranged from positively biased to negatively biased, as the H value increased.

3.1.3 Model based on superposition of autoregressive processes

A method based on the Superposition of Autoregressive Processes (SAP) generates asymptotically self-similar sequences when aggregating several independent autoregressive processes. In the simplest case this can be presented as the sum of two autoregressive processes of the first order, as shown in (22)

$$Z_{1i} = A_{1i}z_{1,i-1} + y_{1,i}, \quad Z_{2i} = A_{2i}z_{2,i-1} + y_{2i}, \quad i = 1, 2, \dots
 \tag{22}$$

where A_{1i} and A_{2i} are randomly chosen from a beta-distribution $B(\alpha_1, \alpha_2)$ on $[0, 1]$ with the shape parameters α_1 and α_2 , where $\alpha_1 > 0, \alpha_2 > 0$; y_{1i} and y_{2i} are a pair of independent and identically distributed sequences of random variables with mean of zero and variance $\sigma^2 = 1$.

Using the least-square fitting we can find that $\alpha_2 = 7.7929 * \log(H) + 4.9513$. Then, the Hurst parameter H is linearly dependent on the shape parameter α_2 of a beta-distribution, while α_1 can be selected arbitrary; for example, $\alpha_1 = 1$ in all cases that are investigated.

Table 3. Accuracy of generated sequences ($H = 0.6; 0.7$)

Methods	Mean Values of Estimated H and ΔH			
	0.6		0.7	
	\hat{H}	ΔH	\hat{H}	ΔH
FBNDP	0.6086 (0.581, 0.636)	+1.440	0.6875 (0.660, 0.715)	-1.789
SFRP	0.6166 (0.589, 0.644)	+2.759	0.7091 (0.682, 0.737)	+1.296
SAP	0.5989 (0.595, 0.603)	-0.182	0.6852 (0.680, 0.690)	-2.112

Table 4. Accuracy of generated sequences ($H = 0.8; 0.9$)

Methods	Mean Values of Estimated H and ΔH			
	0.8		0.9	
	\hat{H}	ΔH	\hat{H}	ΔH
FBNDP	0.7827 (0.755, 0.810)	-2.157	0.8502 (0.823, 0.878)	-5.538
SFRP	0.7986 (0.771, 0.826)	-0.174	0.8686 (0.841, 0.896)	-3.491
SAP	0.7845 (0.781, 0.788)	-1.937	0.8971 (0.894, 0.900)	-0.325

The PDF $f(x)$ of the beta-distribution is defined as (23),

$$f(x) = \begin{cases} \frac{x^{\alpha_1-1}(1-x)^{\alpha_2-1}}{\beta(\alpha_1, \alpha_2)}, & 0 < x < 1, \\ 0, & \text{otherwise} \end{cases} \quad (23)$$

where $\beta(\alpha_1, \alpha_2)$ is given as (24).

$$\beta(\alpha_1, \alpha_2) = \int_0^1 X^{\alpha_1-1}(1-x)^{\alpha_2-1} dx = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}. \quad (24)$$

This method, also based on a superposition of the autoregressive processes, consists of the following steps.

Algorithm 3:

Step 1: Set $i = i + 1$. Determine z_{1i} and z_{2i} using (9).

Step 2: Calculate the sum $X_i = z_{1i} + z_{2i}$, $i = 1, 2, \dots$

Step 3: Repeat Step 1 and Step 2 until $i = n$, where n is the number of sample points.

The asymptotically self-similar traffic sequence $\{X_1, X_2, \dots\}$ of 1048576 numbers was generated in 3.5 seconds on the Centrino-based HP PC (1.76 GHz). Unlike other sequential generators, the SAP generator does not require an aggregation level to be assumed as an input parameter. But it requires the shape parameter α_1 as input parameter.

Mean values of estimated H obtained with algorithms for sequential generators are shown in Table 3 and Table 4, respectively, for Hurst parameter, equal to 0.6, 0.7 and 0.8, 0.9.

The cumulative relative error for the FBNDP method was less than 5%, except for $H = 0.9$. For the SEFR method the cumulative relative error was +6.21%, +1.78%, -1.43% and -5.37%, for $H = 0.6, 0.7, 0.8$ and 0.9, respectively.

Similar to the FBNDP method, estimated H values ranged from positively biased to negatively biased, as the H value increased. For $\alpha_2 = 2.9$ and 8.1 all values of the Hurst parameter from the traffic sample sequence of the SAP method were higher than the required values.

The relative error was +24.14% and +10.45%, respectively; in that case, the results were overestimated. For $\alpha_2 = 21.3$ and 71.5, the relative error was -0.01% and -8.45%.

3.2 Fixed-length sequence generators of self-similar IP network traffic

The most frequently studied discrete-time models of self-similar IP network traffic belong to fractional autoregressive integrated moving-average (F-ARIMA) and fractional Gaussian noise (FGN) processes, and possible fixed-length sequence generators are based on:

- Fractional autoregressive integrated moving-average processes (F-ARIMA);
- Fast Fourier transform (FFT);
- Fractional Gaussian noise and Daubechies wavelets;
- Random midpoint displacement (RMD);
- Successive random additions.

3.2.1 Model based on fractional autoregressive integrated moving-average processes

Let us consider F-ARIMA(0, d , 0) method for generating self-similar sequences, where d is the fractional differencing parameter, $0 < d < 1/2$, and let generate the process $\mathbf{X} = \{X_i: i = 0, 1, 2, \dots, n\}$ with a normal marginal distribution, the mean of zero and the variance σ_0^2 , and the autocorrelation function (ACF), $\{\rho_k\}$ ($k = 0, \pm 1, \dots$) that defined as (25),

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)}, \quad (25)$$

where

$$\gamma_k = \sigma_0^2 \frac{(-1)^k \Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)}.$$

The fixed-length sequence of self-similar pseudo-random numbers is generated through the following process.

Algorithm 4:

Step 0: Set $N_0 = 0$ and $D_0 = 1$. X_0 , the first pseudo random element in the output self-similar sequence, is generated from the normal distribution $N(0, \sigma_0^2)$, where σ_0^2 is the required variance of the X_i .

Step i: ($i = 1, \dots, n - 1$). Compute mean_i and var_i of X_i recursively, using the following equations

$$\begin{aligned}
 N_i &= \rho_i - \sum_{j=1}^{i-1} \Phi_{i-1,j} \rho_{i-j}, \\
 D_i &= D_{i-1} - N_{i-1}^2 / D_{i-1}, \\
 \Phi_{ii} &= N_i / D_i, \\
 \Phi_{ij} &= \Phi_{i-1,j} - \Phi_{ii} \Phi_{i-1,i-j}, j = 1, \dots, i - 1,
 \end{aligned}$$

where Φ_{ij} , $i = 0, j = 0, \dots, n - 1, j = 0, \dots, n - 1$ is given by

$$\begin{aligned}
 \Phi_{ii} &= - \binom{i}{j} \frac{(j-d-1)!(i-d-j)!}{(-d-1)!(i-d)!}, \\
 \text{mean}_i &= \sum_{j=1}^i \Phi_{ij} X_{i-j}, \\
 \text{var}_i &= (1 - \Phi_{ii}^2) \text{var}_{i-j}.
 \end{aligned}$$

Generate X_i from $N(\text{mean}_i, \text{var}_i)$. Increase i by 1.

Step N: If $i = n$, then stop.

A self-similar traffic sequence $\{X_1, X_2, \dots, X_n\}$ was obtained. It took 4 hours, 1 min and 24 sec on the Centrino-based HP PC (1.76 GHz) to generate the F-ARIMA traffic sample sequence with 1 048 576 numbers. The algorithm, based on the F-ARIMA method, is too rigorous to be used to generate long sample sequences [13].

3.2.2 Model based on fractional Gaussian noise and Daubechies wavelets

In this work we propose a new advanced generator of pseudo-random self-similar sequence based on Fractional Gaussian Noise (FGN) and Daubechies Wavelet (DW), called the FGN-DW method.

The generators based on the DW produce more accurate self-similar sequences than those based on Haar Wavelets (HW) [4]. In that case not only estimates of H obtained from the DW are closer to the true values than those from the HW, but also variances obtained from the DW are much lower.

The reason behind is that the DW produce smoother wavelet coefficients that are used in the Discrete Wavelet Transform (DWT), than the HW. The HW are discontinuous, and they do not have good time-frequency localisation properties, since their Fourier transforms decay as $|\lambda|^{-1}$, $\lambda \rightarrow \infty$, meaning that resulting decomposition has a poor scale. Therefore, the DW produce more accurate coefficients than the HW.

The suggested approach for generating synthetic self-similar FGN sequences in a time domain is based on a DWT. Wavelets can provide compact representations for a class of FGN processes, because the structure of wavelets naturally matches the self-similar structure of the LRD processes.

This paper confirms that the FGN-DW method is sufficiently fast being used for practical generation of synthetic self-similar sequences as simulation input data. The wavelet analysis transforms a sequence onto a time-scale grid, where the term scale is used instead of frequency, because the mapping is not directly related to frequency as in the Fourier transform.

The wavelet transform generates the wavelet coefficients $d_x(i, j)$ from a sequence of given numbers. For an LRD process, the variance of the wavelet coefficients at each level i is defined by (26),

$$E[d_x(j, \cdot)]^2 = CP_f 2^{i(2H-1)} \tag{26}$$

where $C > 0$ and (18) is true,

$$P_f = 2(2\pi)^{1-2H} c_\gamma E(2H - 1) \sin((1 - H)\pi) \tag{27}$$

where $E(\cdot)$ is the Euler function and c_γ is a positive constant. The power parameter P_f that plays a major role in fixing the absolute size of the LRD process and generates effects in the applications is an independent quantitative parameter with the dimensions of variance.

A spectral estimator can be obtained though calculating time average Q_i of the $d_x(i, \cdot)$ at a given scale that could be described as (28),

$$Q_i = \frac{1}{n_i} \sum_{j=1}^{n_i} d_x^2(i, j), \tag{28}$$

where n_i is the number of wavelet coefficients at scale i , ie $n_i = 2^i n$, where n is the number of data points. The estimator uses a weighted linear regression as the variance of $\log_2(Q_i)$ vary with $\log_2(2^i)$. An estimated Hurst parameter \hat{H} could be obtained with the linear regression (29),

$$y_i = \log_2(Q_i) = (2\hat{H} - 1)i + c + 1/(n_i \ln 2) \tag{29}$$

where c is a constant.

The wavelet transform delivers good resolution in both time and scale, as compared to the Fourier transform, which provides only good frequency resolution.

The fixed-length sequence of self-similar pseudo-random numbers can be generated through the recently developed process that consists of the following steps.

Table 5. Accuracy of generated sequences ($H = 0.6; 0.7$)

Wavelet coeff.	Mean Values of Estimated H and ΔH			
	0.6		0.7	
	\hat{H}	ΔH	\hat{H}	ΔH
2	0.6019 (0.574, 0.630)	+0.323	0.6984 (0.671, 0.726)	-0.229
4	0.6026 (0.575, 0.630)	+0.433	0.7039 (0.676, 0.731)	+0.554
8	0.6026 (0.575, 0.630)	+0.430	0.7031 (0.676, 0.731)	+0.445
16	0.6013 (0.574, 0.629)	+0.214	0.6987 (0.671, 0.726)	-0.185

Table 6. Accuracy of generated sequences ($H = 0.8; 0.9$)

Wavelet coeff.	Mean Values of Estimated H and ΔH			
	0.8		0.9	
	\hat{H}	ΔH	\hat{H}	ΔH
2	0.7943 (0.767, 0.822)	-0.709	0.8898 (0.862, 0.917)	-1.137
4	0.8055 (0.778, 0.833)	+0.684	0.9074 (0.880, 0.935)	+0.821
8	0.8039 (0.776, 0.831)	+0.486	0.9049 (0.877, 0.932)	+0.545
16	0.7962 (0.769, 0.824)	-0.474	0.8938 (0.866, 0.921)	-0.694

Table 7. Variances of Hurst parameter

Wavelet coeff.	Mean Values of Estimated H and ΔH			
	0.6	0.7	0.8	0.9
2	1.949e-04	2.154e-04	2.393e-04	2.670e-04
4	2.508e-04	2.540e-04	2.586e-04	2.645e-04
8	2.440e-04	2.432e-04	2.438e-04	2.456e-04
16	2.116e-04	2.055e-04	2.007e-04	1.973e-04

Algorithm 5:

Step 1: Given: H , Start for $i = 1$ and continue until $i = n$. calculate a sequence of values $\{f_1, f_2, \dots, f_n\}$ using (30),

$$f(\lambda, H) = c_f |\lambda|^{1-2H} + O(|\lambda|^{\min(3-2H, 2)}), \quad (30)$$

where $c_f = \sigma^2(2\pi)^{-1} \sin(\pi H)\Gamma(2H + 1)$, $O(\cdot)$ represents the residual error and $f_i = \hat{f}(\pi/n; H)$, the value of the frequencies f_i corresponds to the spectral density of an FGN process for f_i ranging between $\pi/n \div \pi$.

Step 2: Multiply $\{f_i\}$ by realizations of the independent exponential random variable with the mean of one to obtain $\{\hat{f}_i\}$, because the spectral density estimated for a given frequency is distributed asymptotically as the independent exponential random variable with the mean $f(\lambda, H)$.

Step 3: Generate a sequence $\{Y_1, Y_2, \dots, Y_n\}$ of complex numbers such that $|Y_i| = \sqrt{\hat{f}_i}$ and the phase of Y_i is uniformly distributed between 0 and 2π . This random phase technique preserves the spectral density corresponding to $\{\hat{f}_i\}$. It also makes the marginal distribution of the final sequence normal and produces the requirements for FGN.

Step 4: Calculate two synthetic coefficients of orthonormal Daubechies wavelets. The output sequence $\{X_1, X_2, \dots, X_n\}$ representing approximately self-similar FGN process is obtained by applying the inverse Daubechies wavelets transformation of the sequence $\{Y_1, Y_2, \dots, Y_n\}$.

It took only 2 seconds on Centrino-based HP PC (1.76 GHz), to generate a sequence of 1 048 576 numbers using this algorithm.

4 PERFORMANCE ANALYSIS

The Hurst parameter estimates for the wavelet-based H estimator are shown in Table 5 ($H = 0.6$ and 0.7) and Table 6 ($H = 0.8$ and 0.9) with different wavelet coefficients. The obtained results are averaged over 30 sequences, and they show that for all input H values the F-ARIMA and the FGN-DW methods produce sequences with less biased H values than other methods.

The means of estimated H values, obtained with the suggested wavelet-based algorithm, are shown in Table 7. Analysis of the mean times required to generate sequence of a given length demonstrates that the sequential generators are more attractive for the practical simulation studies of computer networks than the F-ARIMA-based generator, since they are much faster.

However, these generators require more input parameters, and selecting appropriate values is a problem that remains. In addition, the problem of how to define the relationship between the Hurst parameter and two shape parameters of a beta-distribution in the case of SAP also remains.

Let us consider generation of self-similar time series with the help of Algorithm 5 with one million values Hurst parameter $H = 0.5$; time scaling factor equal to 10 and number of wavelet coefficients equal to 10. The obtained results are shown in Fig. 1.

For the generated time series the values $Q(j_1)$ were calculated as function of j_1 , as is shown on Fig. 2. When linear regression (29) was implemented, better representation was obtained if to compare to the analytical values, as shown in Fig. 3. The obtained results demonstrate self-similar and long range dependent properties of the generated time series sequence, as is shown in Fig. 4. The Hurst parameter H was determined with a very high precision that obtained the exact value of 0.504 for the original value of $H = 0.5$. The obtained results were estimated based on correspondence criteria with parameter

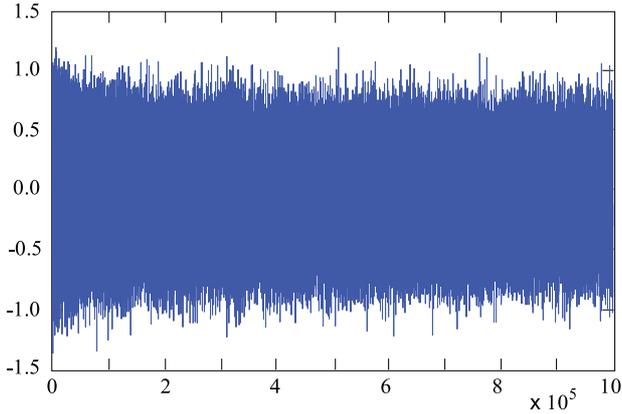


Fig. 1. Self-similar time series with 1 000 000 values

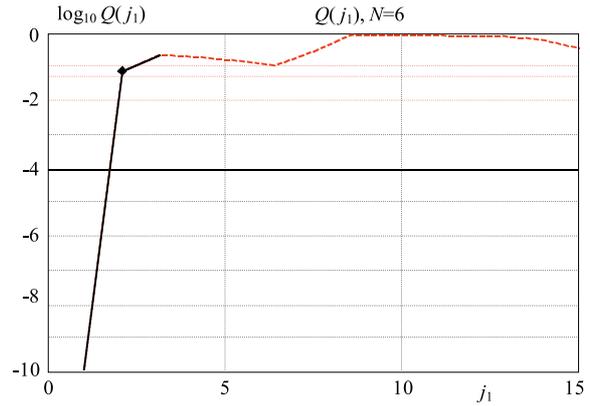


Fig. 2. Evaluation of self-similar time series for different scaling parameters

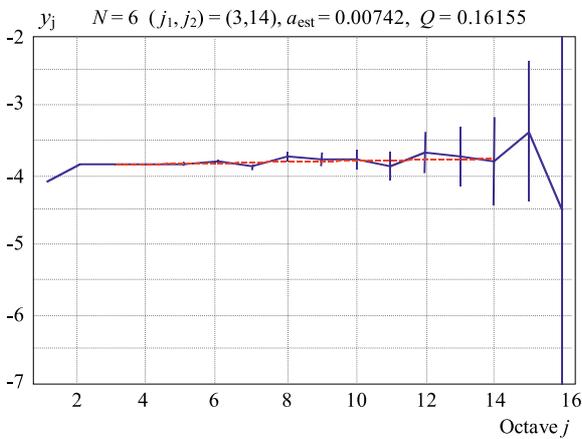


Fig. 3. Comparison between generated time series and analytical values

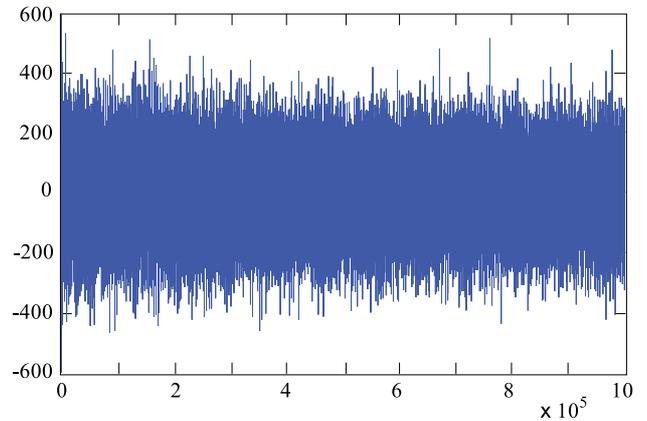


Fig. 4. Product of self-similar and Pareto time series

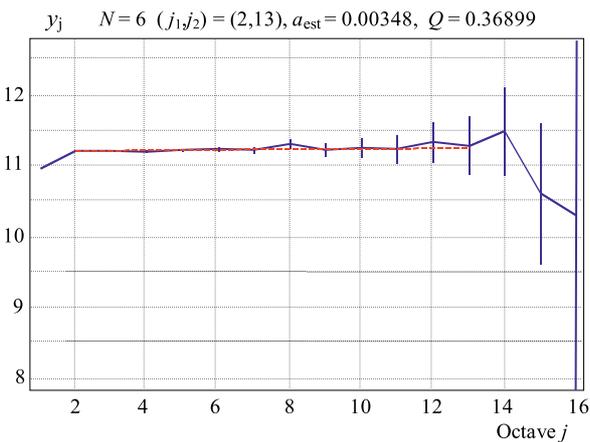


Fig. 5. Comparison between generated time series and obtained values with weighted linear regression

$Q = 0.16155$, and time series with equivalent length of 1 000 000 values were generated with Pareto distribution and $\alpha = 1.8$. Then, they were multiplied with self-similar time series, generated with Algorithm 5.

The implementation of wavelet criteria provided good results for $Q(j_1)$ as function of j_1 , which were very similar to the results presented in Fig. 2. Considering the whole time interval $(j_1, j_2) = [1, 16]$, the optimal values were registered at $(j_1, j_2) = [2, 13]$, in addition to the fact that the correspondence criteria parameter Q increased from 0.16155 to 0.36899.

The linear regression provided an opportunity to define the Hurst parameter with better precision (0.502 versus 0.504, for the exact value of 0.5). The logarithmic diagram obtained with weighted linear regression is shown in Fig. 5. The confidential intervals are shown with use of vertical lines. They are small for $j_1 < 13$ and obtained with confidential probability of 0.95 after generation of 2^{10} samples.

5 CONCLUSIONS

Communication network traffic exhibits self-similar (or, fractal) properties over a wide range of time scales that are very different from the properties of traditional models. Self-similar models are appropriate for IP network traffic as they provide capacity to estimate the net-

work performance and quality, allocate the resources, and ensure the QoS.

One of the problems that telecommunications researchers face during simulation is how to generate long synthetic sequential self-similar sequences. To solve the problem three aspects must be considered: how accurately self-similar process can be generated; how quickly the methods generate long self-similar sequences, and how appropriately self-similar processes can be used in sequential simulation.

Advanced fractal models and efficient algorithms for self-similar IP network traffic simulation were developed and applied. Numerical examples were provided; and the simulation results were obtained and analyzed.

REFERENCES

- [1] BOBBIO, A.—HORVÁTH, A.—SCARPA, M.—TELEK, M.: Acyclic discrete phase type distributions: Properties and a parameter estimation algorithm, - *Performance Evaluation*, **54(1)** (2003), 1–32.
- [2] BORST, S. C.—MITRA, D.: Virtual Partitioning for Robust Resource Sharing: Computational Techniques for Heterogeneous Traffic, *IEEE Journal on Selected Areas in Communications* **16** No. 5 (1998), 668-678.
- [3] BOXMA, O.—COHEN, J.: Self-Similar Network Traffic and Performance Evaluation, John Wiley & Sons, New York, 2000.
- [4] DAUBECHIES, I.: Ten Lectures on Wavelets, vol. 61 of CBMS-NFS Regional Conference Series in Applied Mathematics, SIAM Press, Philadelphia, Pennsylvania, 1992.
- [5] FARAJ, R.: Modeling and Analysis of Self-Similar Traffic in ATM Networks, PhD thesis, 2000.
- [6] GIAMBENE, G.: Queueing Theory and Telecommunications: Networks and Applications, Springer, NY, 2005.
- [7] HAYES, J.: GANESH BABU, T.: Modeling and Analysis of Telecommunications Networks, John Wiley and Sons, NY, 2004.
- [8] JEONG, H.-D.: Modelling of Self-Similar Teletraffic for Simulation, PhD Thesis, University of Canterbury. USA, 2002.
- [9] KUSHNER, H.: Heavy Traffic Analysis of Controlled Queueing and Communications Networks, Springer, NY, 2001.
- [10] MEHDI, J.: Stochastic Models in Queueing Theory, Academic Press, 2003.
- [11] PARK, K.—WILLINGER, W.: Self-Similar Network Traffic and Performance Evaluation, John Wiley & Sons, New York, 2000.
- [12] RADEV, D.—LOKSHINA, I.: Modeling and Simulation of Self-Similar Teletraffic, Proceedings of the Industrial Simulation Conference (ISC) 2007, Delft, The Netherlands, 2007, pp. 355–359.
- [13] RADEV, D.—LOKSHINA, I.: Modeling and Simulation of Self-Similar Wireless IP Network Traffic, Proceedings of the IEEE Wireless Telecommunications Symposium (WTS) 2009, Prague, Czech Republic, 2009, pp. 1–6.
- [14] RUBINSTEIN, R.—MELAMED, B.: Modern Simulation and Modeling, John Wiley & Sons, New York, 1998.
- [15] STATHIS, C.—MAGLARIS, B.: Modeling the Self-Similar Behaviour of Network Traffic, *Journal of Computer Networks* **34** No. 1 (2000), 37-47.
- [16] TRIVEDI, K.: Probability and Statistics with Reliability, Queueing, and Computer Science Applications, John Wiley & Sons, New York, 2001.

Received 26 February 2010

Dimitar Radev, DSc is Professor at Department of Communication Technique and Technologies at University of Rousse, Bulgaria. His main research interests are Teletraffic Theory, Modeling and Simulation in Communication Networks, and Performance Analysis of Queueing Systems.

Izabella Lokshina, PhD is Associate Professor of MIS and chair of Management, Marketing and Information Systems Department at SUNY Oneonta, USA. Her main research interests are Intelligent Management Information Systems, Telecommunications Networks, Complex System Modeling, Simulation and Performance Analysis.



EXPORT - IMPORT
of periodicals and of non-periodically
printed matters, books and CD-ROMs

Krupinská 4 PO BOX 152, 852 99 Bratislava 5, Slovakia
tel: ++421 2 638 39 472-3, fax: ++421 2 63 839 485
info@slovart-gtg.sk; <http://www.slovart-gtg.sk>

