Generation and Testing of Self-Similar Traffic in ATM Networks

Anand R. Prasad, Borut Stavrov, Frits C. Schoute

Telecommunications and Traffic Control Systems Group
Delft University of Technology
P.O. Box 5031
2600 GA Delft
The Netherlands
Tel./Fax: +31 15 278 1797/1774
Email: AnandR@Octopus.Et.TuDelft.NL

Abstract — Recent findings from detailed studies of traffic measurements from variety of different packet networks have brought up a surprising discrepancy between the traditional traffic modelling techniques and the actual network traffic. The studies have shown the actual network traffic to be statistically *self-similar* with significant implications for the design of future multi-service integrated networks. This new traffic feature can be effectively captured within fractal models like: *fractional Brownian motion* (fBm) and *fractional ARIMA* processes. Although these formal mathematical models provide an elegant solution to the modelling of the *self-similar* phenomena, an comprehensive queuing analysis of these models is still lacking. Therefore simulations with synthetic *self-similar* input traffic are essential for gaining better understanding of the queuing problems and some initial experience with the performance of the future networks. Consequently fast generation of long traces of *self-similar* processes becomes an important task. In this paper we will use an fBm generation method called *Successive Random Addition* (SRA) algorithm and carry out a rigorous statistical analysis on the generated traces. Our results will show that the traces are indeed *self-similar*, although the parameters obtained may slightly differ from their target values. Our conclusion is that for qualitative studies the SRA algorithm provides a very good traffic source, whereas for quantitative analysis some caution is recommended. We will also mention some possible applications of the algorithm in performance-related network implementations.

I. Introduction

In the last decade networking technology has experienced a tremendous growth and development. Much of this attention is thanks to the enormous popularity of the Broadband Integrated Services Digital Network (B-ISDN) based on the Asynchronous Transfer Mode (ATM) technology, or better known under the well publicised name of the "Information Superhighway". B-ISDN promises to be a single, high speed, service independent, flexible and efficient network, that will be able to transport voice, data video and multimedia over the same infrastructure. ATM provides the technology needed for carrying out this task; it is supposed to efficiently handle aggregate traffic streams with very different QoS requirements and provide essentially infinitely scaleable connectivity.

From traffic modelling point of view, the emergence of modern high speed networks and the prevailing trend towards B-ISDN, combines drastically new and different transmission and switching technologies with a very heterogeneous mixture of services and applications. The traditional approach based almost exclusively on *Poisson* (or more generally, Markovian) assumptions, that has worked so well in the past, has undergone an extensive criticism in the new environment, especially because of the complete absence of any practical validation of the increasingly more complex models (e.g., Markovian Arrival Process [3] or Batch-Markovian Arrival Process [4]) and their apparent differences with the traffic behaviour seen in the actual networks.

In the past few years detailed studies of measurements from different working networks (e.g., CCSN/SS7 at 56 kbs, ISDN at 1.5 Mbs, Ethernet LAN at 10 Mbs, VBR video for ATM with a mean rate of 5.34 Mbs), have been performed (see [13],[1],[11],[14]). The studies have produced several remarkable results: (i) the actual network traffic is clearly distinguishable from the traffic generated by the traditional theoretical models, (ii) the network traffic is statistically *self-similar*, i.e. the traffic bursts are spanning several time scales from milliseconds to hours, (iii) the traffic is much more bursty than expected and furthermore aggregating the traffic streams typically intensifies the burstiness (*self-similarity*), which is in sharp contrast to the conventional teletraffic wisdom based on the folklore theorem[1].

This fractal behaviour of the traffic is very different from any currently used traffic model. Mathematically, the *self-similarity* manifests itself through the presence of long-range dependence in the traffic trace. Long-range dependent processes on the other hand are characterised by an autocorrelation function that decays hyperbolically in the lag. Given that long-range dependence is ubiquitous in the measured traffic and cannot be captured by any of the currently employed theoretical models, there has been an increasing concern about the practical relevance of the traditional traffic models and the validity of the traffic engineering guidelines based on them.

This is even further emphasised in the light of the recent findings, that the performance of the queuing models with long-range dependent input processes can be

[1] The folklore theorem basically says: "multiplexing a large number of independent traffic streams results in a *Poisson* like (smoother) traffic.

0-7803-3177-X/96/$5.00©1996 IEEE

drastically different from the one obtained using traditional short-range dependent models, which typically results in too optimistic QoS guarantees (e.g., see [6], [15]). This gives a rise to the need for a new traffic models, capable of capturing the fractal nature of the real traffic, and their performance analysis.

During the past decade Personal Wireless Communication (PWC) systems have developed very rapidly. Recently studies of mobile broadband system (MBS) and ATM oriented MBS have drawn the researchers' attention (see, e.g. [7]). The goal is to make B-ISDN services available to the mobile user, e.g. a broadband wireless multimedia service. In other words the system has to be able to transmit most of the broadband services and the conventional narrowband services over the air interface without a great loss of quality. The MBS in fact provides the interface between the fixed (ATM) part of the network and the mobile user. Keeping in mind the fractal nature of the ATM traffic, self-similar traffic modelling is also relevant for the design of the future PWC systems.

From a modelling viewpoint, accounting for the self-similarity does not require very complex models. On the contrary, as discussed in [1], two models that yield an elegant and parsimonious representation of the self similar phenomena are *fractional Gaussian noise (FGN)* and *fractional autoregressive integrated moving-average (ARIMA)* processes. LAN traffic can be successfully modelled using FGN process with three parameters: mean, variance and Hurst parameter; fractional ARIMA processes with four or five parameters seem to describe (variable bit rate) VBR video traffic reasonably accurate [14].

The trouble is however that the queuing theory behind these models has just started to develop. There are very few analytical results available at this moment (see [12]). Most of the performance results are obtained via trace-driven simulations (see [1],[8],[14],[15]). Therefore simulation studies, using a synthetic self-similar process as input traffic, are essential for gaining better understanding of the queuing behaviour under such conditions and a more realistic performance analysis of the future networks. This is why the generation of long traces of self-similar processes has become increasingly important to researchers (e.g., see [10]).

The accuracy and quality of different generation methods has been studied in [10],[5] . In this paper we use a fBm generation method, called *Successive Random Addition* algorithm, derived from the *Random Midpoint Displacement* algorithm used in [10], for generation of FGN processes. The algorithm has been used for some time now in the field of fractal geometry, but its appropriateness as a traffic source has not been studied yet.

The generated traces are then tested for the extent of their self-similarity by means of two statistical tests: *Rescaled-Adjusted Range or, R/S Statistic* [5] and *Variance-Time Analysis* [1].

Section 2 discusses the definition and the properties of the self-similar processes. Section 3 describes the traffic generation: the definition of fBm and description of the SRA algorithm. The R/S statistic and the Variance-Time analysis are explained in section 4. Finally our conclusions and future research activities are summarised in section 5.

II. Self-similar processes: definition and properties

Although it sounds very intuitive, self-similarity is not only a popular name, but a rigorously defined mathematical concept. We follow here the approach used in [1],[11].

Let $X = (X_t: k = 1,2, \ldots)$ be a wide-sense stationary stochastic process, i.e. a process with constant mean μ, finite variance $\sigma^2 = E[(X_t-\mu)^2]$ and an autocorrelation function $r(k) = E[(X_t-\mu)(X_{t+k}-\mu)] / E[(X_t-\mu)^2]$, $k\geq 0$, that depends only on k. In particular, we will assume that X has an autocorrelation function of the form:

$$r(k) \sim a_1 \cdot k^{-\beta}, \text{ as } k \to \infty \qquad (1)$$

where $0 < \beta < 1$ and a_1, a_2, ... are all finite positive constants.

Let's construct a new time series by averaging the original series X over non-overlapping blocks of size m. In other words, for each m = 1, 2, 3, ..., let $X^{(m)} = (X_k^{(m)}: k=1,2,\ldots)$ denote the new covariant stationary time series where $X^{(m)}$ is defined as $X_k^{(m)} = 1/m(X_{km-m+1} + \ldots + X_{km})$, $(k\geq 1)$. The process X is called *(exactly) second-order self-similar* with self-similarity parameter $H = 1-\beta/2$, if for all $m = 1,2, \ldots$, $\text{var}(X^{(m)}) = \sigma^2 m^{-\beta}$ and

$$r^{(m)}(k) = r(k), \text{ for } k \geq 0. \qquad (2)$$

In other words, X is *exactly self-similar* if the aggregated processes $X^{(m)}$ are indistinguishable from X, at least with respect to their second order statistical properties.

A covariance stationary process X is called *asymptotically (second-order) self-similar* with a self-similarity parameter $H = 1-\beta/2$, if for k large enough:

$$r^{(m)}(k) \to r(k), \text{ as } m \to \infty \qquad (3)$$

with $r(k)$ given by (1).

Maybe the most striking feature of the second-order self-similar processes is that the aggregated processes $X^{(m)}$ possess a non-degenerate correlation structure as $m \to \infty$. This is also visible in the measured traces, because the aggregated traffic, i.e. number of bytes/cells/packets per time unit, looks the same regardless of the choice of the time unit over which we observe the process. This behaviour is in stark contrast to the traditional traffic models, currently in use, whose aggregated processes $X^{(m)}$, tend to second-order pure noise (i.e., for all $k\geq 1$, $r^{(m)}(k) \to 0$, as $m \to \infty$).

There is one very interesting consequence of this property, which was observed in all of the measurement studies. Namely, within the time scales of interest there is no natural length of the **burst**, i.e. at time scales ranging from milliseconds to minutes and hours, the bursts have the same qualitative appearance. This was first noticed by

Mandelbrot and Wallis, who have found the bursts "too elusive to be useful" (see [5]).

Mathematically, self-similarity manifests itself in a presence of *long-range dependence* (LRD). LRD or, also known as the *Joseph Effect*, captures the phenomena of persistent and long periods (bursts) of consecutive large or small values of the self-similar process. More formally defined, LRD processes are characterised by an autocorrelation function that decays hyperbolically in the lag k, which implies a non-summable autocorrelation function $\Sigma_k\, r(k) = \infty$. Traditional *short-range dependent* (SRD) processes on the other hand, have exponentially decaying autocorrelation function, like $r(k) \sim \rho^k$, for $0 < \rho < 1$.

Other properties of the self-similar processes are: (i) slowly decaying variances, more formally defined as $\mathrm{var}(X^{(m)}) \sim a_2 m^{-\beta}$; (ii) the spectral density near the origin possesses $1/f$ *noise* behaviour, whereas currently considered traffic models have bounded spectral densities fro low frequencies.

Statistically the self-similarity can be checked with a number of different methods, some of which will be discussed in the following sections. An interesting result of these tests is an estimate of the so-called *Hurst parameter*, H, which measures the degree of self-similarity in a given time series. A H-value of 0.5, indicates, that the underlying process that generated the time series, is a SRD process, whereas H-values strictly in the interval between 0.5 and 1 are characteristic for LRD (self-similar) processes. What makes the Hurst parameter so interesting, is that it can be used as a measure of burstiness. This is particularly important since the currently popular measures are behaving very poorly in the presence of self-similar traffic, due to its extremely variable inter-arrival times.

III. Self-similar traffic generation

Generation of fractal processes has been an attractive research area for years. There are a lot of known algorithms (exact and approximate). However their appropriateness fro generation of network traffic has rarely been investigated. We have chosen for an approximate algorithm called *Successive Random Addition* (SRA), which is closely related to the Random Midpoint Displacement method, used in [10]. First some basic definitions and properties of the fBm will be explained, followed by an detailed description of the generation algorithm.

A. Fractional Brownian Motion

Fractional Brownian motions and fractional Gaussian noises, as defined in [5], are respectively, generalisations of *Brownian motion* and *white Gaussian noise*. The term *fractional* can be best explained by considerations from spectral theory. Classically ordinary Brownian motion $V(t)$ is defined as:

$$V(t) = \int_{-\infty}^{t} W(s)ds \qquad (4)$$

where $W(s)$ denotes white Gaussian noise. Formally, Gaussian noise is defined as a random process with spectral density which is independent of the frequency f. As a result its (repeated) integral(s) and derivatives(s) all have spectral densities of the form f^{-2k}, with k as an integer. FGN are defined as processes where k can take non-integer values and it is usually written like: $0.5 - H$, for $0 < H < 1$. This explains the term fractional.

Similarly fBm is a second-order self-similar process $V_H = (V_H(t) : t \geq 0)$, with a self-similarity parameter H, $0 < H < 1$, i.e. for all $a > 0$:

$$V_H(at) = a^H \cdot V_H(t) \qquad (5)$$

where the equality should be interpreted in terms of a probability distribution function. One of its interesting properties is that fBm has stationary increments which are typically related through a simple H scaling law:

$$V_H(t_2) - V_H(t_1) \sim (t_2 - t_1)^H \qquad (6)$$

This increment process is actually an FGN with parameter H.

B. Successive Random Addition

Since the SRA algorithm is based upon the RMD method, we will begin with a description of the later.

RMD is a recursive generating technique which sacrifices mathematical purity for execution speed in its approximation of the fBm. Consider a simple approximation of fBm, $V_H(t)$, where the mean square increment, for points separated by a time $\Delta t = 1$, is σ^2. Then, for points separated by time t, using (6), we obtain:

$$E\{(V_H(t) - V_H(0))^2\} = t^{2H} \cdot \sigma^2 \qquad (7)$$

If, $V_H(0) = 0$, then the points at $t = \pm 1$ are chosen as samples of a Gaussian random variable with variance σ^2 to satisfy (6). Given these initial conditions, one defines the midpoints at:

$$V_H(\pm\frac{1}{2}) = 0.5 \cdot [V_H(0) + V_H(\pm 1)] + \delta_1 \qquad (8)$$

where δ_1 is a Gaussian random variable with zero mean and variance Δ_1^2, which has to satisfy (6),.

$$\Delta_1^2 = \frac{\sigma^2}{2^{2H}} - \frac{1}{4}\mathrm{var}[\Delta V_H(1)]$$

$$= \frac{\sigma^2}{2^{2H}} \cdot [1 - 2^{2H-2}] \qquad (9)$$

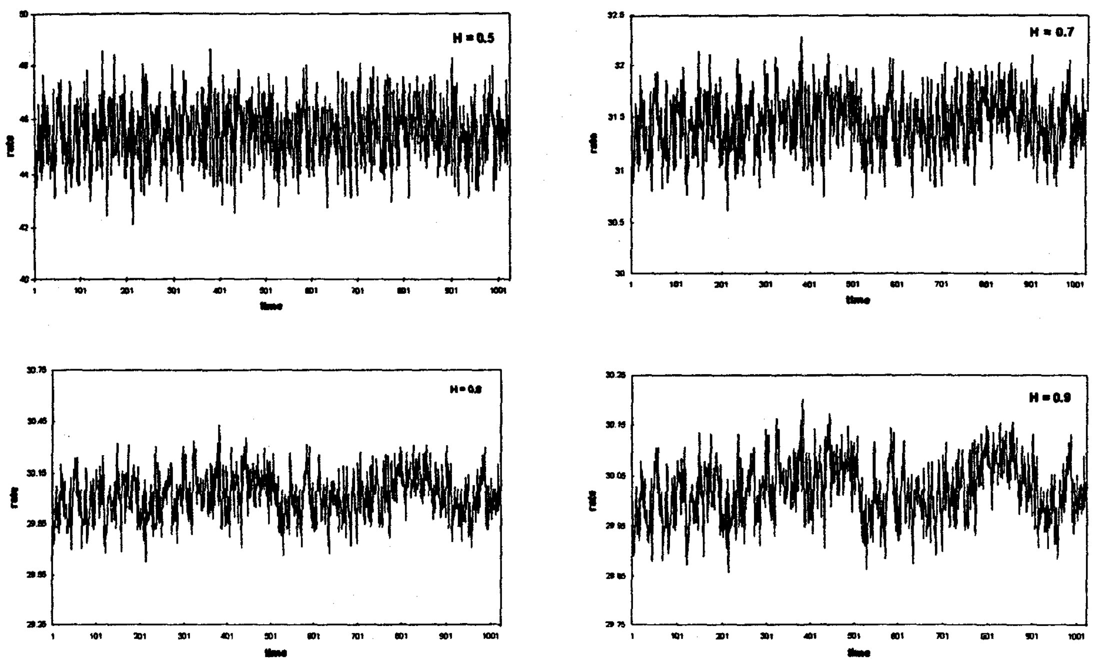

Figure 1. FGN traces with $M = 30$, $a = 2.5$, for $H = 0.5, 0.7, 0.8$, and 0.9.

As H goes to 1, Δ_1^2 goes to 0, no new fluctuations are added at smaller stages and $V_H(t)$ remains a collection of smooth line segment connecting the start points.

At n^{th} stage, a random Gaussian variable δ_n is added to the midpoint of the stage $n-1$ with a variance:

$$\Delta_n^2 = \frac{\sigma^2}{(2^n)^{2H}} \cdot [1 - 2^{2H-2}]. \qquad (10)$$

This is where the error is introduced, namely this process does not have stationary increments. Once a given point at t_i has been determined, its value remains unchanged in all later stages. All additional stages change independently and the correlation required for the fBm with $H > 1/2$ are not present.

So as to deal with the non-stationarity of the midpoint displacement technique, a *Successive Random Addition* is introduced. In the SRA algorithm the midpoints are interpolated the same way as in the RMD technique, but, a displacement of a suitable variance is added to all the points at each stage of recursive subdivision and not just the midpoints [2].

Approximate fBm processes, generated by the algorithm can be interpreted as the cumulative arrival processes [9].

$$A(t) = Mt + \sqrt{aM}V_H(t) \qquad (11)$$

where M is the mean rate and a the peakedness factor, defined as: ratio of variance to the mean number of cells in a unit time interval. The rate process, or the number of arrivals per time unit, in the interval from time t to t+1 is defined as:

$$\tilde{A}(t) = M + \sqrt{aM}(V_H(t+1) - V_H(t)) \qquad (12)$$

This formula has been used for generating the traces presented in Fig.1. The value of M is 30 and a is chosen to be 2.5. The four traces correspond to values oh $H = 0.5, 0.7, 0.8, 0.9$. The first graph ($H = 0.5$), corresponds to white noise and it is virtually featureless. As H increases above 0.5, the traces become richer in low frequency terms. This can be seen through the presence of trends and cyclic swings which are visible in the traces.

The SRA algorithm is simple to implement and it can generate long sequences (e.g., 100.000 observations) in very short instances.

IV. Statistical analysis

There are several methods to investigate the presence of the long-range dependence in a empirical time series. The two methods described bellow are *Rescaled-Adjusted Range Statistics (R/S statistic)* and *Variance-Time analysis*. They are also used for estimation of the Hurst parameter H.

A. Rescaled-Adjusted Range Statistics (R/S statistic)

Given a random process X_i at time i and the cumulative bit rate $Y_j = \Sigma^j_{i=1} X_i$ up to time j, then the expression:

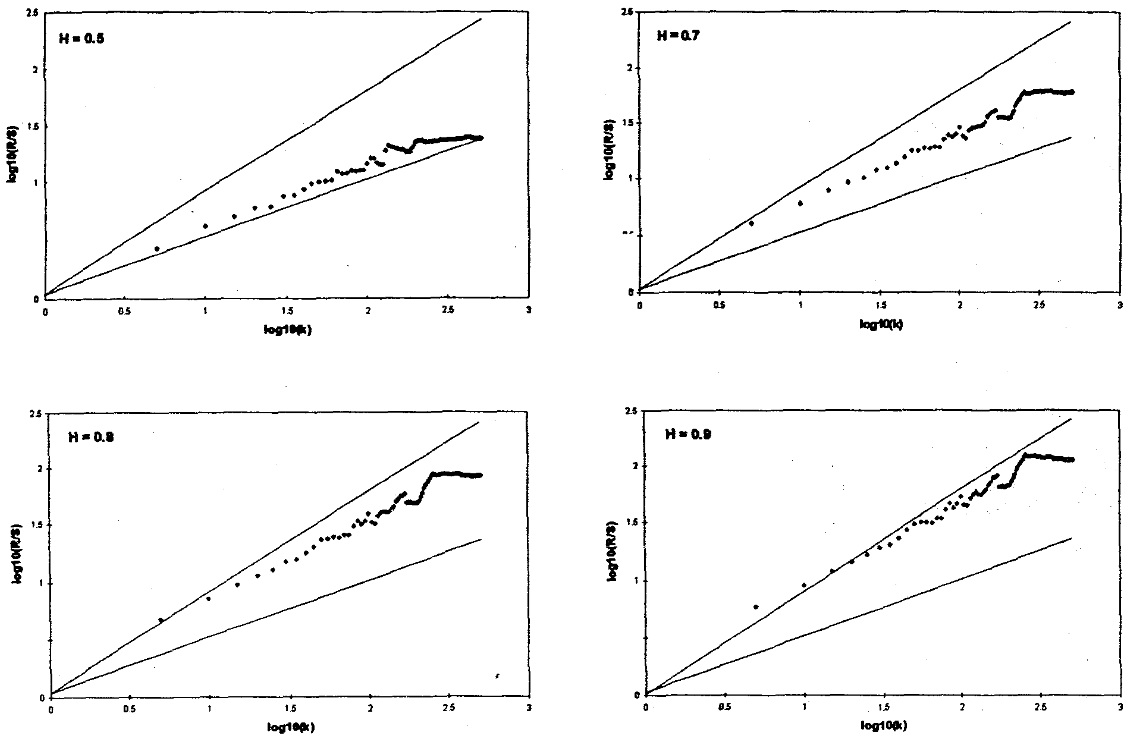

Figure 2. *R/S* statistic plots for H = 0.5, 0.7, 0.8 and 0.9

$$R(t,k) = \max_{0\le i\le k}[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)]$$

$$-\min_{0\le i\le k}[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)] \qquad (13)$$

is called the adjusted range. In order to study the properties that are independent of the scale, $R(t,k)$ is standardised by:

$$S(t,k) = \sqrt{k^{-1}\sum_{i=t+1}^{t+k}(X_i - \overline{X}_{t,k})^2} \qquad (14)$$

where $\overline{X}_{t,k} = k^{-1}\sum_{i=t+1}^{t+k} X_i$, is the sample mean and $S^2(t,k)$ is the sample variance. The ratio:

$$R/S = \frac{R(t,k)}{S(t,k)} \qquad (15)$$

is called the *rescaled adjusted range* or R/S-statistic. According to Hurst's work for large values of k, log *R/S* is scattered around a straight line with a slope that exceeds 1/2. In probabilistic terminology this means that for k large,

$$\log E[R/S] \approx a + H\cdot\log k \qquad (16)$$

for $H > 1/2$. Let $Q = Q(t,k) = R(t,k)/S(t,k)$ be the *R/S* statistic defined previously. Then the *R/S* method can be summarised as follows [3]: (1) calculate Q for all possible, or for a sufficient number of different values for t and k; (2) plot log Q against log k; (3) draw a straight line $y = a + b \log k$ that corresponds to the "ultimate" behaviour of the data. The coefficients a and b can be estimated by the least squares or any similar method.

Fig.2. presents the *R/S* plots of the traces given in Fig.1. The straight lines correspond to trends of 0.5 and 0.9 respectively. It is easy to see how different values of H produce different trends. The estimated values for H = 0.5, 0.7, 0.8, 0.9 are 0.51, 0.73, 0.79, 0.83. One can see from these values that for larger values of H the generated traces are less accurate. This does not present a problem for qualitative analysis, however for quantitative studies one is advised to test the trace's parameters before using the trace.

B. Variance-Time Analysis

One of the striking properties of long-range dependent processes is that the variance of the sample mean converges slower to zero than n^{-1} [3].This can be tested with the following method.

Let k be an integer. For different values of k in the range $2 \le k \le n/2$, and a sufficient number (say m_k) of subseries of length k, calculate the sample means of $X_1(k)$, $X_2(k)$,.....,$X_m(k)$ and the overall mean. Then, for each k, calculate the sample variance of the sample means $X_j(k)$ for $j = 1,...,m_k$:

$$\overline{X}(k) = m_k^{-1}\cdot\sum_{j=1}^{m_k}\overline{X}_j(k). \qquad (17)$$

For each k, calculate the sample variance of the sample means $X_j(k)$ ($j = 1,\dots,m_k$):

$$S^2(k) = (m_k - 1)^{-1} \sum_{k=1}^{m_k} (\overline{X_j}(k) - \overline{X}(k))^2 \quad (18)$$

Plot $\log S^2(k)$ against $\log k$. For large values of k, the points in this plot are expected to be scattered around a straight line with negative slope $2H-2$. The variance-time plots are given in Fig. 3. The estimated values of H are very similar to the one obtained with the R/S method. The same conclusions apply here too.

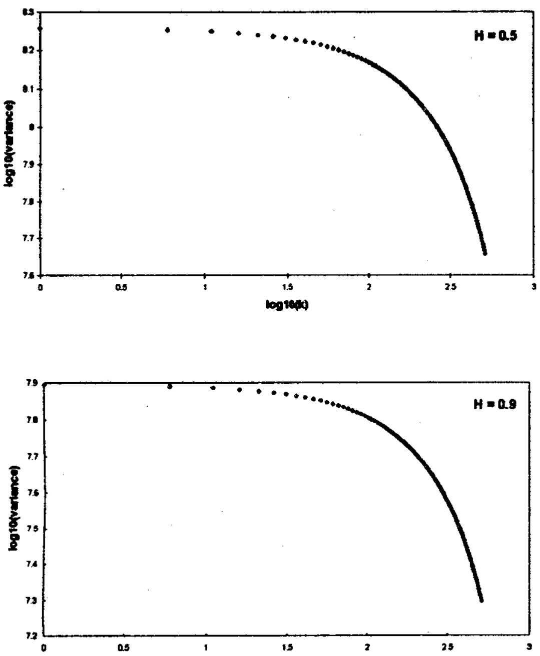

Figure 3. Variance-Time plots for $H = 0.5, 0.9$.

V. Conclusions

We have used the Successive Random Addition algorithm to generate self-similar traffic traces. Long traces of 100,000 observations or more can be generated in a few seconds. The algorithm is also simple to implement and incorporate in larger simulation modules. The statistical analysis has shown that SRA is very good for qualitative analysis. For quantitative studies the results generated for $H < 0.9$ are very satisfactory. Further work is needed to get better results for higher values of H. To get a better insight to the usability of the *SRA* algorithm, extensive tests for other values of Hurst parameter should be done.

Future research is also directed towards testing the nature of the traffic which has passed through several ATM switches, in a complex network. The SRA model presented here, is used for the generation of the self-similar input traffic.

References

[1] E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)", *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1- 15, February 1994.

[2] H.O. Peitgen and D. Saupe, *The Science of Fractal Images*, Springer-Verlag, New York, 1988.

[3] J. Beran, *Statistics of Long-Memory Processes*, Chapman and Hall, New York, 1994.

[4] H.E. Hurst, "Long-Term Storage Capacity of Reservoirs", *American Society of Civil Engineers*, vol. 116, pp. 770-799, 1951.

[5] B.B. Mandelbrot and J.R. Wallis, "Computer Experiments with Fractional Gaussian Noises Part 1-3, Rescaled Ranges and Spectras", *Water Resources Research*, vol. 5, no. 1, pp. 242-267, February 1969.

[6] I. Norros, "A Storage Model with Self-Similar Input", *Queueing Systems 16*, pp. 387-396, 1994.

[7] A.R. Prasad, "Asynchronous Transfer Mode Based Mobile Broadband System", *IEEE Third Symposium on Communication and Vehicular Technology in Benelux*, Eindhoven, The Netherlands, pp. 143-148, October 25-26 1995.

[8] R.G. Addie, M. Zukerman and T. Neame, "Performance of a Single Server Queue with Self-Similar Input", *ICC'95*, Seattle, 'pp.461-464, June 18-22 1995.

[9] I. Norros, "On the use of Fractional Brownian Motion in the Theory of Connectionless Networks", *IEEE Journal on Selected Areas in Communications*", vol. 13, no. 6, pp. 953-962, August 1995.

[10] W.C. Lau, A. Erramilli, J.L. Wang and W. Willinger, "Self-Similar Traffic GEneration: The Random Midpoint Algorithm and its Properties", *ICC'95*, Seattle, pp. 466-472, June 18-22 1995.

[11] W.E. Leland, M.S. Taqqu, W. Willinger and D.V. Wilson, "On the Self-Similar NAture of Ethernet Traffic", *Proc. ACM/SIGCOMM'93*, pp. 183-193, 1993.

[12] I. Norros, J. Virtamo, A. Simonian and D. Veitch, "A Benes Formula for a Buffer with Fractional Brownian Input", *9th ITC Specialists Seminar Teletraffic Modelling and Measurement in Broadband and Mobile Communications*, pp. 293-304, November 7-9 1995.

[13] D.E. Duffy, A.A. McIntosh, M. Rosenstein and W. Willinger, "Sattistical Analysis of CCSN/SS7 Traffic Data from Working CCS Subnetworks", *IEEE Journal of Selected Areas of Communication*, no. 12, pp. 544-551, 1994.

[14] M.W. Garrett and W. Willinger, "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic", *Proc. ACM/Sigcomm'94*, London, pp. 269-280, September 1994.

[15] H.J. Fowler and W.E. Leland, "Local Area Network Tradffic Characteristics, with Implications for Broadband Network Congestion Management", *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 7, pp. 1139-1149, September 1991.