Generation of Self-Similar Time Series for Simulation Studies of Telecommunication Networks

Hae-Duck J. Jeong   Krzysztof Pawlikowski
Department of Computer Science

Donald C. McNickle
Department of Management

University of Canterbury
Christchurch, New Zealand

Abstract

It is generally accepted that self-similar processes may provide better models for teletraffic in modern telecommunication networks than Poisson processes. If stochastic self-similarity of teletraffic is not taken into account, it can lead to inaccurate conclusions about performance of networks. Thus, an important requirement for conducting simulation studies of networks is the ability to generate long synthetic self-similar sequences of incremental processes, to transform them into inter-event time intervals, and this should be done accurately and fast. A method for transformation of count processes into inter-event processes proposed by Leroux and Hassan [14] is studied. A case study is discussed to show how long sequences are needed in steady-state simulation of queueing models with self-similar input processes. This is compared with simulation run lengths of the same queueing models fed by Poisson processes.

1 Introduction

The search for accurate mathematical models of data streams in modern telecommunication networks has attracted a considerable amount of interest in the last few years. The reason is that several recent teletraffic studies of local and wide area networks, including the world wide internet, have shown that commonly used teletraffic models, based on Poisson or related processes with short-range dependencies (SRD), are not able to capture the self-similar, (or fractal) nature of teletraffic with long-range dependencies (LRD); see Leland et al. [13], Likhanov et al. [15], Paxson and Floyd [22], Ryu [24]. This is particularly important in networks that are engaged in such sophisticated services as variable-bit-rate (VBR) video transmission (Garrett and Willinger [6], Krunz and Makowski [11], Rose [23]). The properties of teletraffic in such scenarios are very different from both the properties of conventional models of telephone traffic and the traditional models of data traffic generated by computers.

The use of traditional models of teletraffic can result in overly optimistic estimates of performance of telecommunication networks, insufficient allocation of communication and data processing resources, and difficulties in ensuring the quality of service expected by network users (Beran [1], Neidhardt and Wang [19], Paxson and Floyd [22]). On the other hand, if the strongly correlated character of teletraffic is explicitly taken into account, this can also lead to more efficient traffic control mechanisms (Östring et al. [20]).

Several methods for generating pseudo-random self-similar count processes have been proposed. They include methods based on fractional Gaussian noise (Mandelbrot [16]), fractional ARIMA processes (Hosking [8]), the $M/G/\infty$ queue model (Krunz and Makowski [11], Leland et al. [13]), autoregressive processes (Cario and Nelson [3], Granger [7]), spatial renewal processes (Taralp et al. [26]), etc. Some of them generate asymptotically self-similar sequences and require large amounts of CPU time. For example, Hosking's method (Hosking [8]), based on the F-ARIMA$(0, d, 0)$ process, needs many hours to produce a self-similar sequence with 131,072 (2^{17}) numbers on a Sun SPARCstation (Leland et al. [13]). It requires $O(n^2)$ computations to generate n numbers. Even though exact methods of generation of self-similar sequences exist (for example: Mandelbrot [16]), they are only fast enough for short sequences. They are usually inappropriate for generating long sequences because they require multiple passes along generated sequences. To overcome this, approximate methods for generation of self-similar sequences in simulation studies of telecommunication networks have also been proposed (Lau et al. [12], Paxson [21]).

An important requirement for conducting simulation studies of telecommunication networks is the ability to generate long synthetic self-similar sequences of incremental processes, long enough to ensure arbitrary statistical precision of the final simulation results. Very little is known on appropriateness of selection of specific inter-arrival processes and transformation of inter-event time intervals for arrival counts generated by an FGN process still remains an open research issue (Leroux and Hassan [14], Paxson [21]).

In this paper we study a transformation of count process based on the *Fractional Gaussian Noise* and *Daubechies Wavelets* (*FGN-DW*) into inter-arrival processes. A generator of such count processes, based on *FGN-DW*, has been proposed in Jeong et al. [10]. Then, the inter-arrival processes are used in steady-state simulation of queueing models with self-similar input processes.

A summary of the basic properties of self-similar processes is given in the next section. Generation and transformations of the *FGN-DW* count processes are described in Section 3. In Section 4, a case study is discussed to show how long sequences are needed in steady-state simulation of queueing models with self-similar input processes (called $SSM/M/1/\infty$). This is compared with simulation run lengths of the same queueing models fed by SRD processes such as $M/M/1/\infty$ queueing system. The influence of the Hurst parameter on simulation run-length is also analysed, before the final conclusions are formulated.

2 Self-Similar Processes and Their Properties

Basic definitions of self-similar processes are as follows:
A continuous-time stochastic process $\{X_t\}$ is strongly self-similar with a self-similarity parameter H $(0 < H < 1)$, known as the Hurst parameter, if for any positive stretching factor c, the rescaled process with time scale $ct, c^{-H}X_{ct}$, is equal in distribution to the original process $\{X_t\}$ (Beran [2]). This means that, for any sequence of time points $t_1, t_2, \ldots, t_n$, and for all $c > 0$, $\{c^{-H}X_{ct_1}, c^{-H}X_{ct_2}, \ldots, c^{-H}X_{ct_n}\}$ has the same distribution as $\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\}$.

In discrete-time case, let $\{X_k\}, k = 0, 1, 2, \ldots,$ be a (discrete-time) stationary process with mean μ, variance σ^2, and autocorrelation function (ACF) $\{\rho_k\}$, for $k = 0, 1, 2, \ldots$, and let $\{X_k^{(m)}\}_{k=1}^{\infty} = \{X_1^{(m)}, X_2^{(m)}, \ldots\}$, $m = 1, 2, 3, \ldots$, be a sequence of batch means, i.e., $X_k^{(m)} = (X_{km-m+1} + \ldots + X_{km})/m, k \geq 1$.

The process $\{X_k\}$ with $\rho_k \to k^{-\beta}$, as $k \to \infty, 0 < \beta < 1$, is called *exactly self-similar* with $H = 1-(\beta/2)$, if $\rho_k^{(m)} = \rho_k$, for any $m = 1, 2, 3, \ldots$. In other words, the process $\{X_k\}$ and the averaged processes $\{X_k^{(m)}\}$, $m \geq 1$, have identical correlation structure. The process $\{X_k\}$ is *asymptotically self-similar* with $H = 1 - (\beta/2)$, if $\rho_k^{(m)} \to \rho_k$, as $m \to \infty$.

The most frequently studied models of self-similar traffic belong either to the class of fractional autoregressive integrated moving-average (F-ARIMA) processes or to the class of fractional Gaussian noise processes; see Hosking [8], Leland et al. [13], Paxson [21]. F-ARIMA(p, d, q) processes were introduced by Hosking (Hosking [8]), where p is is order of auto-regression in ARIMA process, d is degree of differencing in ARIMA process, and q is order of moving average in ARIMA process. Hosking showed that these processes are asymptotically self-similar with Hurst parameter $H = d + \frac{1}{2}$, as long as $0 < d < \frac{1}{2}$. On the other hand, the incremental process $\{Y_k\} = \{X_k - X_{k-1}\}, k \geq 0$, is called the *fractional Gaussian noise* (FGN) process, if $\{X_k\}$ represents a fractional Brownian motion (FBM) random process. This process is a (discrete-time) stationary Gaussian process with mean μ, variance σ^2 and $\{\rho_k\} = \{\frac{1}{2}(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H})\}$, $k > 0$. An FBM process is characterised by three properties (Mandelbrot and Wallis [17]): (i) it is a continuous zero-mean Gaussian process $\{X_t\}$ defined for $t \geq 0$ and $0 < H < 1$, with ACF given by $\rho_{s,t} = \frac{1}{2}(s^{2H} + t^{2H} - |s-t|^{2H})$, where s is time lag and t is time; (ii) its increments $\{X_t - X_{t-1}\}$ form a stationary random process; (iii) it is self-similar with Hurst parameter H, that is, for all $c > 0$, $\{X_{ct}\} \approx \{c^H X_t\}$, in the sense of probability distribution. That is if time is changed by the ratio c, then $\{X_t\}$ is changed by c^H.

Main properties of self-similar processes include:

- *Slowly decaying variance.* The variance of the sample mean decreases more slowly than the reciprocal of the sample size, that is, $Var[\{X_k^{(m)}\}] \to c_1 m^{-\beta_1}$ as $m \to \infty$, where c_1 is a positive constant and $0 < \beta_1 < 1$.

- *Long-range dependence (LRD).* A process $\{X_k\}$ is called a stationary process with *LRD* if its ACF $\{\rho_k\}$ is non-summable, that is, $\sum_{k=0}^{\infty} \rho_k = \infty$. The speed of decay of autocorrelations is more like hyperbolic than exponential.

- *Hurst effect*. Self-similarity manifests itself by a straight line of slope β_2 on a log-log plot of the R/S statistic. For a given set of numbers $\{X_1, X_2, \ldots, X_n\}$ with sample mean $\hat{\mu} = E\{X_i\}$ and sample variance $S^2(n) = E\{(X_i - \hat{\mu})^2\}$, Hurst parameter H is presented by the *rescaled adjusted range* $\frac{R(n)}{S(n)}$ (or R/S *statistic*) where $R(n) = \max\{\sum_{i=1}^{k}(X_i - \hat{\mu}), 1 \leq k \leq n\} - \min\{\sum_{i=1}^{k}(X_i - \hat{\mu}), 1 \leq k \leq n\}$ and S is estimated by $S(n) = \sqrt{E\{(X_i - \hat{\mu})^2\}}$. Hurst found empirically that for many time series observed in nature the expected value of $\frac{R(n)}{S(n)}$ asymptotically satisfies the power law relation, i.e., $E[\frac{R(n)}{S(n)}] \rightarrow c_2 n^H$ as $n \rightarrow \infty$ with $0.5 < H < 1$ and c_2 is a finite positive constant; see for example Beran [2], Cox [4], Leland et al. [13].

- *$1/f$-noise*. The spectral density $f(\lambda; H)$ obeys a power law near the origin, i.e., $f(\lambda; H) \rightarrow c_3 \lambda^{1-2H}$, as $\lambda \rightarrow 0$, where c_3 is a finite positive constant and $0.5 < H < 1$.

We will use these properties to investigate characteristics of generated self-similar sequences.

3 Teletraffic Generation and Simulation

We claim that the *FGN-DW* transformation is sufficiently fast for generation of synthetic self-similar sequences, to be used as simulation input data (Jeong et al. [10]). Let us briefly introduce the *FGN-DW* method itself.

3.1 Generation of Self-Similar Teletraffic Using FGN-DW

General strategy behind our method is the same as in Paxson [21]. The algorithm consists of the following steps (for more detailed discussions, see Jeong et al. [10]):

Step.1 Calculate a sequence of values $\{f_1, f_2, \cdots, f_n\}$ using

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

where $c_f = \sigma^2 (2\pi)^{-1} sin(\pi H) \Gamma(2H+1)$ and $O(\cdot)$ represents the residual error, $f_i = \hat{f}(\frac{\pi i}{n})$, corresponding to the spectral density of an FGN process for frequencies from $\frac{\pi}{n}$ to π.

Step.2 Adjust $\{f_i\}$ by multiplying them by realisations of an independent exponential random variable with mean equal 1.

Step.3 Generate a sequence $\{x_1, x_2, \cdots, x_n\}$ of complex numbers such that $|x_i| = \sqrt{\hat{f}_i}$ and the phase of x_i is uniformly distributed between 0 and 2π. The random phase technique, taken from Schiff [25], preserves the distribution of spectral density of $\{\hat{f}_i\}$, but ensures that different sequences generated using the method will be independent. It also makes the marginal distribution of the final sequence normal. The phase randomisation makes the different frequency components independent (Paxson [21]).

Step.4 Calculate two synthetic coefficients of orthonormal Daubechies wavelets which are used in the inverse discrete wavelet transform (IDWT) (Nartallo et al. [18]). Then, generate the approximately self-similar FGN sequence in time domain by using the IDWT from $\{x_i\}$. We use the Daubechies wavelets because it produces more accurate coefficients of wavelets than Haar wavelets (Wickerhauser [27]).

Using the above steps, the proposed FGN-DW method gives a fast generator of well approximated self-similar sequences of numbers representing FGN processes which can be interpreted as differential arrival rates.

3.2 Transformation of Self-Similar Count Processes into Inter-Arrival Processes

Simulation studies of telecommunication networks require generation of arbitrarily long sequences of inter-arrival times of transmitted data packets. Therefore, a mechanism is needed to transform self-similar processes representing the arrival counts, or differential arrival rates (see Subsection 3.1), into the suitable sequences of inter-arrival times of packets, while preserving appropriate characteristics. In this paper, we use a count process generated by *FGN-DW* to obtain exponential inter-arrival distribution in the way suggested by Leroux and Hassan [14]. The transformation is based on comparison of the count values in consecutive short time intervals, called bins. This transformation follows the following steps:

1. To obtain non-negative integers representing the bin counts, an exponential transformation is applied to the samples to eliminate the negative values which can occur in the original sequence representing differential arrival rates.

2. The resulting real-valued samples are rounded to the nearest integers, to get integer values of bin counts.

3. To generate the inter-arrival time of the next packet use an exponential generation function: $X = -\frac{1}{\lambda} * \log(r)$ (Jain [9]), where r is a random number uniformly distributed on $(0, 1)$, and λ is the bin count from step 2. The resulted distribution of inter-arrival times is controlled by a single parameter λ, which changes from bin to bin. Because of its connection with exponential distribution, we will call it SSM.

4 Numerical Results

The first issue is to show how well the self-similarity of the original arrival counts is preserved when the arrival counts are converted into suitable inter-arrival times. Assuming that preservation of self-similarity means to have coefficient H changed by less than 10 %, our results in Table 1 - 3 show that inter-event process preserves well the self-similarity of the original sequence generated by the FGN-DW (except $H \leq 0.7$ if it is estimated from the periodogram). Large relative differences in self-similarity in Table 1 (as measured by H) of count processes and inter-event processes

H	Count Processes	Inter-Event Processes	Relative Inaccuracy
0.5	.4999	.7033	40.7 %
0.6	.5960	.7416	24.4 %
0.7	.6913	.7992	15.6 %
0.8	.7883	.8427	6.9 %
0.9	.8840	.8703	1.5 %

Table 1: $\hat{H}$ estimated from periodogram plots.

H	Count Processes	Inter-Event Processes	Relative Inaccuracy
0.5	.5398	.5932	9.9 %
0.6	.5974	.6551	9.7 %
0.7	.7067	.7223	2.2 %
0.8	.7876	.7993	1.5 %
0.9	.8565	.8708	1.7 %

Table 2: $\hat{H}$ estimated from RS-statistic plots.

H	Count Processes	Inter-Event Processes	Relative Inaccuracy
0.5	.4991	.5633	12.9 %
0.6	.5974	.6303	5.5 %
0.7	.6895	.7121	3.3 %
0.8	.7876	.7924	0.6 %
0.9	.8532	.8620	1.0 %

Table 3: $\hat{H}$ estimated from variance-time plots.

may be caused by the analytical method used for obtaining the periodogram plot in the frequency domain. The inter-arrival times of the packets generated according to the exponential distribution are used in steady-state simulation of a $SSM/M/1/\infty$ queueing system, i.e., a queueing system with a self-similar inter-arrival process based on a time dependent exponential distribution.

The second issue is to look at how many observations are needed in sequential steady-state simulation of a queueing system with self-similar input processes. An $M/M/1/\infty$ and $SSM/M/1/\infty$ queueing systems have been simulated in AKAROA-2 (Ewing et al. [5]), a fully automated tool of sequential simulation designed for stopping simulation with a pre-specified statistical precision of the final results.

Figure 1 shows the number of theoretically required observations, and the mean number of empirical observations need for stopping simulation with the required statistical precision, for different values of traffic intensity ρ, when analysing steady-state mean response time in an $M/M/1/\infty$ queueing system. Each mean experimental run length of simulation is obtained from 30 replications. Confidence intervals of relative half-widths of 10% or less, at 0.95 confidence level, are also shown in Figure 1. While for $\rho \leq 0.6$ the mean number of empirical observations is slightly

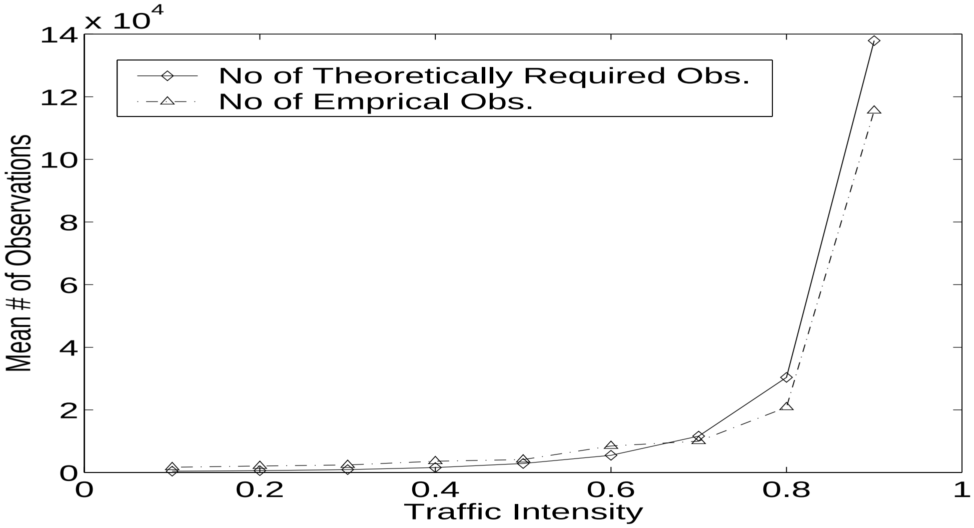

Figure 1: Mean numbers of observations needed in analysis of steady-state mean response time: in $M/M/1/\infty$ queueing system with relative precision $\leq 10\%$, 0.95 confidence level.

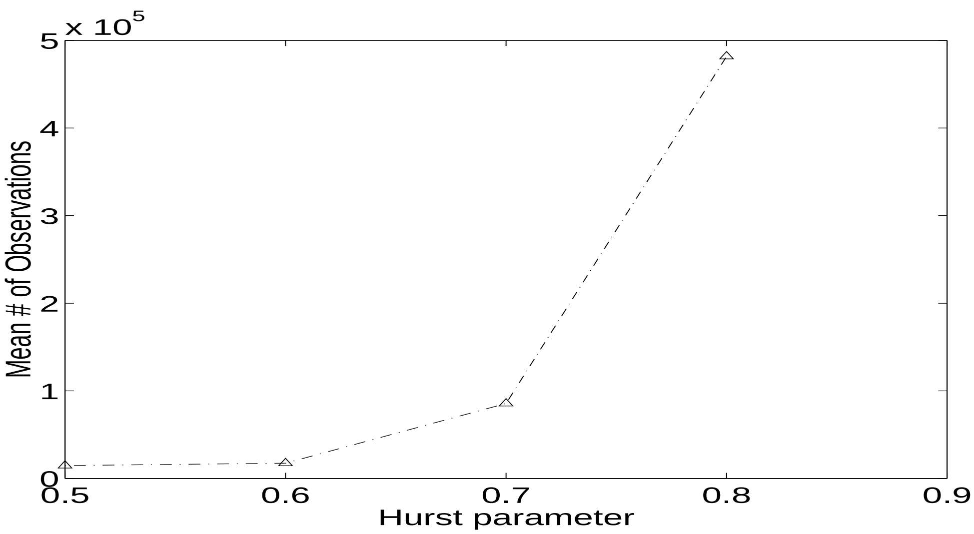

Figure 2: Mean numbers of observations needed in sequential analysis of steady-state mean response time: in $SSM/M/1/\infty$ queueing system with precision $\leq 10\%$, traffic intensity $\rho = 0.22$.

higher than the number of theoretically required observations, for $0.6 < \rho \leq 0.9$ the mean number of empirical observations is lower than the number of observations required theoretically. Both numbers quickly grow as $\rho \to 1$. The range of the reported mean numbers of empirical observations needed is between 1,686 and 115,544.

The results in Figure 2 show that, analysis of an $SSM/M/1/\infty$ queueing system with self-similar input processes requires much more observations than an $M/M/1/\infty$ queueing system. Mean number of observations, or simulation run length, significantly increases with H of the input stream. For $\rho = 0.22$, the range of mean numbers of empirical observations for the $SSM/M/1/\infty$ queueing system is between 14,635 and 481,754, while the $M/M/1/\infty$ queueing system needs between 2,169 and 2,195 observations.

5 Conclusions

We have examined two main issues of stochastic simulation of telecommunication networks with self-similar teletraffic.

First, we looked at how well the self-similarity of the original arrival count processes are preserved when the arrival counts are converted into suitable inter-arrival times. Using exponential distribution as the resulted inter-arrival distribution we have shown that the self-similarity of original counts processes can be preserved, although more robust transformers are still needed.

Next, we looked at how many observations are needed in sequential steady-state simulation of a queueing system with self-similar input processes. As we have shown, assuming self-similar inter-arrival processes, one needs much more observations to obtained the final simulation results with a required precision than when assuming Poisson inter-arrival processes, exhibiting SRD. To secure a predefined statistical precision of simulation final results, one would need to generate arbitrary long sequences of inter-arrival times, and this would need to be done sequentially. At this stage, an algorithm for sequential generation of self-similar sequences has not been proposed yet.

References

[1] Beran, J. (1992), Statistical Methods for Data with Long Range Dependence, *Statistical Science*, **7**, 404-427.

[2] Beran, J. (1994), *Statistics for Long-Memory Processes*, Chapman and Hall, New York.

[3] Cario, M.C. and Nelson, B.L. (1998), Numerical Methods for Fitting and Simulating Autoregressive-to-Anything Processes, *INFORMS Journal on Computing*, **10**, 72-81.

[4] Cox, D.R. (1984), Long-Range Dependence: a Review, *In Statistics: An Appraisal*, Iowa State Statistical Library, The Iowa State University Press, 55-74.

[5] Ewing, G., Pawlikowski, K., and McNickle, D. (1998), Akaroa 2.5: User's Manual, *Technical Report TR-COSC 07/98*, Department of Computer Science, University of Canterbury, Christchurch, New Zealand.

[6] Garrett, M.W. and Willinger, W. (1994), Analysis, Modeling and Generation of Self-Similar VBR Video Traffic, In *Computer Communication Review, Proceedings of ACM SIGCOMM'94*, London, UK, **24**, 269-280.

[7] Granger, C.W.J. (1980), Long Memory Relationships and the Aggregation of Dynamic Models, *Journal of Econometrics*, **14**, 227-238.

[8] Hosking, J.R.M. (1984), Modeling Persistence in Hydrological Time Series Using Fractional Differencing, *Water Resources Research*, **20**, 1898-1908.

[9] Jain, R. (1991), *The Art of Computer Systems Performance Analysis*, John Wiley & Sons, Inc., New York.

[10] Jeong, H.-D.J., McNickle, D., and Pawlikowski, K. (1999), Fast Self-Similar Teletraffic Generation Based on FGN and Wavelets, In *Proceedings of IEEE International Conference on Networks (ICON'99)*, Brisbane, Australia, (In press).

[11] Krunz, M. and Makowski, A. (1998), A Source Model for VBR Video Traffic Based on $M/G/\infty$ Input Processes, In *Proceedings of IEEE INFOCOM'98*, San Francisco, CA, USA, 1441-1448.

[12] Lau, W-C., Erramilli, 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, WA, 466-472.

[13] 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 Transactions on Networking*, **2**, 1-15.

[14] Leroux, H. and Hassan, M. (1998), On the Inter-Arrival Distribution of FGN Arrival Processes, *Technical Report*, School of Computer Science and Software Engineering, Monash University, Caulfield East, Melbourne, Victoria 3145, Australia.

[15] Likhanov, N., Tsybakov, B., and Georganas, N.D. (1995), Analysis of an ATM Buffer with Self-Similar ("Fractal") Input Traffic, In *Proceedings of IEEE INFOCOM'95*, 985-992.

[16] Mandelbrot, B.B. (1971), A Fast Fractional Gaussian Noise Generator, *Water Resources Research*, **7**, 543-553.

[17] Mandelbrot, B.B. and Wallis, J.R. (1969), Computer Experiments with Fractional Gaussian Noises, *Water Resources Research*, **5**, 228-267.

[18] Nartallo, C.M., Prelcic, N.G., Galan, S.J.G., and Cabanelas, C.S. (1996), *Wavelet Toolbox for Matlab*, Available through `http: //www.tsc.uvigo.es/wavelets/uvi_wave.html`.

[19] Neidhardt, A.L. and Wang, J.L. (1998), The Concept of Relevant Time Scales and Its Application to Queueing Analysis of Self-Similar Traffic (or Is Hurst Naughty or Nice?), In *Proceedings of ACM SIGMETRICS'98*, Madison, Wisconsin, USA, 222-232.

[20] Östring, S., Sirisena, H. and Husson, I. (1999), Dual dimensional ABR control scheme using predictive filtering of self-similar traffic, In *Proceedings of ICC'99*, Canada, Submitted.

[21] Paxson, V. (1997), Fast, Approximate Synthesis of Fractional Gaussian Noise for Generating Self-Similar Network Traffic, *ACM SIGCOMM, Computer Communication Review*, **27**, 5-18.

[22] Paxson, V. and Floyd, S. (1995), Wide-Area Traffic: the Failure of Poisson Modeling, *IEEE/ACM Transactions on Networking*, **3**, 226-244.

[23] Rose, O. (1997), *Traffic Modeling of Variable Bit Rate MPEG Video and Its Impacts on ATM Networks*, PhD thesis, Bayerische Julius-Maximilians-Universität Würzburg.

[24] Ryu, B.K. (1996), *Fractal Network Traffic: from Understanding to Implications*, PhD thesis, Graduate School of Arts and Science, Columbia University.

[25] Schiff, S. J. (1992), Resolving time-series structure with a controlled wavelet transform, *Optical Engineering*, **31**, 2492-2495.

[26] Taralp, T., Devetsikiotis, M., and Lambadaris, I. (1998), Efficient Fractional Gaussian Noise Generation Using the Spatial Renewal Process, In *Proceedings IEEE International Communications Conference (ICC'98)*, Atlanta, GA, USA, S41.3.1-S41.3.5.

[27] Wickerhauser, M.V. (1994), *Adapted Wavelet Analysis : from Theory to Software*, A K Peters, Publ., Wellesley, Massachusetts.