Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

FAST SIMULATION OF BROADBAND TELECOMMUNICATIONS NETWORKS CARRYING LONG-RANGE DEPENDENT BURSTY TRAFFIC

José R. Gallardo
Dimitrios Makrakis

Advanced Communications Engineering Centre
Department of Electrical and Computer Engineering
The University of Western Ontario
London, Ontario N6A 5B9, CANADA

Luis Orozco-Barbosa

School of Information Technology and Engineering
Department of Electrical Engineering
University of Ottawa
Ottawa, Ontario K1N 6N5, CANADA

ABSTRACT

A technique for the fast simulation of broadband communications systems is proposed, which is based on regenerative Importance Sampling techniques. Our algorithm is applicable to estimate the probability of rare events when modeling the offered traffic using Fractional Stable Noise (FSN) processes (including Fractional Brownian Noise as a particular case), which have been recently proved to be able to capture both the long-range dependence and the burstiness of today's aggregate network traffic. An exact description of FSN processes is given, as well as an approximation that allows for the application of Importance Sampling techniques. The results obtained for a simple example are also included.

1 INTRODUCTION

Simulation is used quite extensively these days in the planning process of telecommunications networks. Simulation allows the network designer to draw important conclusions and make the right decisions before major capital investments are made. Theoretical and mathematical analysis serves the same purpose as simulation, but when the object of study is too complex, analysis tends to be unmanageable.

The validity of the conclusions obtained, either from simulation or theoretical analysis, depends greatly on how accurately the model captures the actual operation of the system under study. For this reason, especially during the last few years, a great amount of research has been focused on obtaining realistic models for the traffic generated by the users of telecommunications networks. Self-similarity and long-range dependence have been proved to be important features of aggregated traffic, and several models of this type have been proposed with the objective of reflecting the real statistical behaviour of the traffic inside networks. One of the most relevant models presented in the past is described in (Norros 1995) and in (Willinger et al. 1997). They propose a model based on the use of Fractional Brownian Noise.

More recently, the models proposed independently in (Gallardo, Makrakis, and Orozco-Barbosa 1998) and (Karasaridis and Hatzinakos 1998) are a generalization to the one presented in (Norros 1995) and (Willinger et al. 1997), in the sense that, rather than limiting the marginal distribution of the process to be Gaussian, an alpha-stable distribution is now used, which allows us to achieve a better agreement between the burstiness of the artificial process and that of the real traffic by selecting the proper stability coefficient α. The model is stationary and long-range dependent and corresponds to the aggregation of a relatively large number of traffic streams mixed together into a single flow.

Due to their representation by means of a stochastic integral, it is not time-efficient to generate long traces of artificial alpha-stable long-range dependent stochastic processes in a direct manner. In (Gallardo, Makrakis, and Orozco-Barbosa 1999) an algorithm for the fast generation of artificial traces of these processes is presented. Such approach uses an auto regressive (AR) model as an approximation to the actual process, based on the minimum dispersion (MD) principle. Because of the AR expression used to represent the process, in addition to being highly efficient for the generation of artificial traces, this algorithm allows for the application of Importance Sampling techniques to speed up simulations of systems involving this type of traffic.

Fast simulation is desirable when trying to estimate the probability of occurrence of rare events in communications systems, such as buffer overflows, excessive delays or transmission errors. The authors in (Huang, Devetsikiotis, and Lambadaris 1995) and (Li, Wolisz, and Popescu-Zeletin 1998) have dealt with the application of Importance Sampling techniques to fast simulation of systems involving Gaussian processes in general. Unfortunately, those results are not directly applicable to alpha-stable processes.

2 FRACTIONAL STABLE NOISE STOCHASTIC PROCESSES

Fractional Stable Motion (FSM) processes are self-similar stochastic processes with stationary increments (Samorodnitsky and Taqqu 1994). Their marginal distributions are the so-called alpha-stable distributions, which are referred to in the *Generalized Central Limit Theorem* (Feller 1966) as describing the limit behaviour of normalized sums of a relatively large number of independent identically distributed (*iid*) random variables; hence their appropriateness for modeling aggregate traffic. Fractional Brownian Motion (FBM) is a particular case of FSM, since the Gaussian distribution belongs to the alpha-stable family. Fractional Stable Noise (FSN) processes are the stationary, long-range dependent increments of FSM processes during a time interval of unit length. In this work, we are considering three members of the FSN family: *i)* Balanced Linear FSN; *ii)* Anti-balanced Linear FSN; and *iii)* Log-FSN. The exact expression of a FSN process Y_j is the following (Samorodnitsky and Taqqu 1994):

$$Y_j = \int_{-\infty}^{\infty} g(j,x) \cdot M(dx) \qquad (1)$$

where $M(dx)$ is an alpha-stable random measure, and:

$$g(j,x) = \begin{cases} ln|j+1-x| - ln|j-x| & ;\text{Log-FSN} \\ |j+1-x|^{H-1/\alpha} - |j-x|^{H-1/\alpha} & ;\text{Bal LFSN} \\ [j+1-x]^{\langle H-1/\alpha \rangle} - [j-x]^{\langle H-1/\alpha \rangle} & ;\text{Anti-bal LFSN} \end{cases} \qquad (2)$$

In the previous equation, we use the notation:

$$z^{\langle a \rangle} \stackrel{\Delta}{=} |z|^a \cdot sign(z) \qquad (3)$$

for any real number z and for any $a \geq 0$. As described in (Gallardo, Makrakis, and Orozco-Barbosa 1999), a FSN process can be very accurately approximated by an auto regressive (AR) process given by:

$$Y_j \approx \sum_{i=1}^{N} a_i \cdot Y_{j-i} + (\gamma_\varepsilon)^{1/\alpha} \cdot u_j \qquad (4)$$

In equation (4), N denotes the order of the AR process and is a positive integer. The u_j's, called the innovations, are *iid* $S_\alpha(1,0,0)$ random variables, according to the notation used in (Samorodnitsky and Taqqu 1994). Finally, the coefficients a_i's and the innovation dispersion parameter γ_ε are calculated using the minimum dispersion criterion. Equation (4) is equivalent to saying that for a FSN process **Y**, the conditional mean and dispersion parameter of Y_j given the past values $\{y_{j-1}, y_{j-2}, \ldots, y_{j-N}\}$ are given by $\hat{Y}_j \stackrel{\Delta}{=} \sum_{i=1}^{N} a_i \cdot y_{j-i}$ and γ_ε, respectively. In other words, the conditional distribution of Y_j is given by:

$$Y_j \stackrel{d}{=} S_\alpha\left((\gamma_\varepsilon)^{1/\alpha}, 0, \hat{Y}_j\right) \qquad (5)$$

3 REALISTIC TRAFFIC MODEL FOR AGGREGATE TRAFFIC

The traffic model proposed and verified independently in (Gallardo, Makrakis, and Orozco-Barbosa 1998) and (Karasaridis and Hatzinakos 1998) for aggregate streams is defined as follows. Let W_j represent the number of arrivals or offered workload during the j-th time interval of unit length, then:

$$W_j = m + Y_j \qquad (6)$$

where m is the mean value of the number of arrivals per unit time and Y_j is a zero-mean FSN process.

4 ASYMPTOTIC BEHAVIOUR OF A QUEUE WITH FSN INPUT TRAFFIC

It was shown in (Gallardo, Makrakis, and Orozco-Barbosa 1998) and in (Norros 1995) that $V(t)$, the buffer occupancy in a stationary storage system, is given by:

$$V(t) = \sup_{0 \leq s \leq t} [Y(t-s) - (C-m) \cdot (t-s)]$$

$$= \sup_{0 \leq \tau \leq t} [Y(\tau) - (C-m) \cdot \tau] \qquad (7)$$

where m is the mean input rate, C is the service (or leak) rate, with $m < C$, and $Y(\tau)$ is a zero-mean stationary alpha-stable random process representing the new arrivals during a period of length τ. Assuming that the buffer size is $x >> 1$, and using the *principle of the largest term* or *Laplace's method* (which is a heuristic rule that basically translates to saying that *rare events occur in the most likely way*, as described in Duffield and O'Connell (1995)), the probability of buffer overflow can be approximated by:

$$\Pr[V(t) > x] \approx \Pr[Y(\tau_0) > (C-m) \cdot \tau_0 + x] \qquad (8)$$

where $\tau_0 = \max\left\{\frac{H \cdot x}{(C-m)(1-H)},\ t\right\}$

Therefore, analyzing the asymptotic behaviour of the queue is equivalent to analyzing the behaviour of the alpha-stable random *variable* $Y(\tau_0)$ and its probability to exceed the threshold described in equation (8).

5 IMPORTANCE SAMPLING

Importance sampling is one of the classical techniques for increasing the efficiency of Monte Carlo simulations (Bucklew, 1990; Glynn and Iglehart 1989). The basic idea is to modify the system under study by replacing one of the stochastic processes involved with a new one in order to reduce the variance of the estimator. That is usually achieved by increasing in an intelligent way the probability of occurrence of the events of interest. The estimated statistics that result from the simulation are then transformed (unbiased) to make them correspond to the original system.

To be more specific, assume that we have a system whose behaviour depends on the stochastic process $\boldsymbol{W}$ and we want to estimate the expected value of a certain random variable $X(\boldsymbol{W})$. The process $\boldsymbol{W}$ can represent the random input traffic to an ATM switch and $X(\boldsymbol{W})$ can be the cell loss ratio or the proportion of cells with excessive delay. Then:

$$E_{\boldsymbol{W}}\left[X(\boldsymbol{W})\right] = \int_{\boldsymbol{U_W}} X(\boldsymbol{w}) \cdot f_{\boldsymbol{W}}(\boldsymbol{w}) \cdot d\boldsymbol{w} \qquad (9)$$

In the previous equation, $\boldsymbol{U_W}$ is the sample space of $\boldsymbol{W}$ and the notation $E_{\boldsymbol{W}}[\cdot]$ denotes sampling using the process $\boldsymbol{W}$ as the random input to the system. Suppose now that $\boldsymbol{W'}$ is a modified stochastic process such that $f_{\boldsymbol{W}}(\boldsymbol{w}) = 0$ whenever $f_{\boldsymbol{W'}}(\boldsymbol{w}) = 0$ (*absolute continuity condition*). The new probability density function (*pdf*) $f_{\boldsymbol{W'}}(\boldsymbol{w})$ is usually referred to as the *twisted density*. Then we can see that:

$$E_{\boldsymbol{W}}\left[X(\boldsymbol{W})\right] = \int_{\boldsymbol{U_W}} X(\boldsymbol{w}) \cdot \frac{f_{\boldsymbol{W}}(\boldsymbol{w})}{f_{\boldsymbol{W'}}(\boldsymbol{w})} \cdot f_{\boldsymbol{W'}}(\boldsymbol{w}) \cdot d\boldsymbol{w}$$

$$= E_{\boldsymbol{W'}}\left[X(\boldsymbol{W}) \cdot \frac{f_{\boldsymbol{W}}(\boldsymbol{W})}{f_{\boldsymbol{W'}}(\boldsymbol{W})}\right] \stackrel{\Delta}{=} E_{\boldsymbol{W'}}\left[X(\boldsymbol{W}) \cdot L(\boldsymbol{W})\right] \qquad (10)$$

where the quotient $L(\boldsymbol{W}) \stackrel{\Delta}{=} f_{\boldsymbol{W}}(\boldsymbol{W}) / f_{\boldsymbol{W'}}(\boldsymbol{W})$ is known as the *likelihood ratio* or *weight function of the transformation*. Equation (10) suggests that estimating the expected value of $X(\boldsymbol{W})$ via Monte Carlo simulations using $\boldsymbol{W}$ as the random process is equivalent to estimating it using $\boldsymbol{W'}$ and unbiasing each sample by applying the likelihood ratio.

5.1 Estimator Variance

The expected value of $X(\boldsymbol{W})$ and the sample mean obtained using the standard Monte Carlo method are given respectively by $\bar{X}$ and $\hat{X}$ in equation (11) below:

$$\bar{X} = E_{\boldsymbol{W}}\left[X(\boldsymbol{W})\right]; \quad \hat{X} = \frac{1}{N}\sum_{i=1}^{N} X(\boldsymbol{w_i}) \qquad (11)$$

where N is the number of samples taken and $\{\boldsymbol{w}_i \mid i = 1,2,\dots,N\}$ is a set of independent and identically distributed (*iid*) sample paths of the process $\boldsymbol{W}$. The estimator is said to be unbiased because $E_{\boldsymbol{W}}\left[\hat{X}\right] = \bar{X}$. The variance of the estimator is given by:

$$Var_{\boldsymbol{W}}\left[\hat{X}\right] = \frac{1}{N}\left\{E_{\boldsymbol{W}}\left[X^2(\boldsymbol{W})\right] - \bar{X}^2\right\} \qquad (12)$$

When Importance Sampling (IS) is used, the unbiased sample mean is now given by:

$$\hat{X}' = \frac{1}{N}\sum_{i=1}^{N} X(\boldsymbol{w_i}) \cdot L(\boldsymbol{w_i}) \qquad (13)$$

where $\{\boldsymbol{w}_i \mid i = 1,2,\dots,N\}$ is now a set of *iid* sample paths of the modified process $\boldsymbol{W'}$. The variance of the IS estimator is now given by:

$$Var_{\boldsymbol{W'}}\left[\hat{X}'\right] = \frac{1}{N}\left\{E_{\boldsymbol{W'}}\left[X^2(\boldsymbol{W}) \cdot L^2(\boldsymbol{W})\right] - \bar{X}^2\right\} \qquad (14)$$

The major difficulty in applying the IS technique is to find a twisted density that minimizes (or at least reduces considerably, as compared to the standard Monte Carlo method) the variance of the IS estimator for a given number of samples N.

5.2 Uniformly Bounded Likelihood Ratios

Let $X(\boldsymbol{w})$ be a function that assigns nonzero values to those sample paths within a rare event $\boldsymbol{B} \subset \boldsymbol{U_W}$, and assigns a zero value to the sample paths that do not belong to $\boldsymbol{B}$. The indicator function $I_B(\boldsymbol{w})$, which is 1 for all $\boldsymbol{w} \in \boldsymbol{B}$ and 0 for all $\boldsymbol{w} \notin \boldsymbol{B}$, is an example of that kind of functions. Another example could be a function that assigns the proportion of cells with excessive delay within an ATM switch to the corresponding cell arrival sequence $\boldsymbol{w}$. Let us note that if:

$$L(\boldsymbol{w}) \leq k < 1 \;\; ; \;\; \forall \boldsymbol{w} \in \boldsymbol{B}, \;\; k = \text{constant} \qquad (15)$$

then:

$$\frac{Var_{\boldsymbol{W}'}\left[\hat{X}'\right]}{Var_{\boldsymbol{W}}\left[\hat{X}\right]} = \frac{E_{\boldsymbol{W}'}\left[X^2(\boldsymbol{W}) \cdot L^2(\boldsymbol{W})\right] - \overline{X}^2}{E_{\boldsymbol{W}}\left[X^2(\boldsymbol{W})\right] - \overline{X}^2}$$

$$= \frac{E_{\boldsymbol{W}}\left[X^2(\boldsymbol{W}) \cdot L(\boldsymbol{W})\right] - \overline{X}^2}{E_{\boldsymbol{W}}\left[X^2(\boldsymbol{W})\right] - \overline{X}^2} < k \qquad (16)$$

This equation clearly shows that when inequality (15) is satisfied, the IS variance is reduced by a factor k, as compared to traditional Monte Carlo simulations. The condition that $X(\boldsymbol{w})$ be zero outside a proper subset $\boldsymbol{B}$ of $\boldsymbol{U_W}$ is necessary, because the last inequality in equation (16) would not be satisfied otherwise, since $L(\boldsymbol{w})$ cannot be less than 1 for all sample paths in $\boldsymbol{U_W}$.

When inequality (15) is satisfied, it is said that the likelihood ratio is *uniformly bounded* within $\boldsymbol{B}$ (Juneja 1994). This result offers an alternative to trying to minimize the IS variance itself, which tends to be rather complicated most of the time. We can try instead to minimize the maximum value that $L(\boldsymbol{w})$ can take within $\boldsymbol{B}$ or, at least, guarantee that this maximum value is less that 1. In other words, if we are using a parametric approach in the sense that the twisted density function depends on a certain parameter ξ (which could be one of the parameters α, σ, or μ of an alpha-stable random variable, for instance), then our best choice when applying this technique is to use ξ_0 that satisfies:

$$\underset{\boldsymbol{w} \in \boldsymbol{B}}{Max}\; L_{\xi_0}(\boldsymbol{w}) \leq \underset{\boldsymbol{w} \in \boldsymbol{B}}{Max}\; L_{\xi}(\boldsymbol{w}) \;\; ; \;\; \forall \xi \qquad (17)$$

and it will provide variance reduction as long as:

$$\underset{\boldsymbol{w} \in \boldsymbol{B}}{Max}\; L_{\xi_0}(\boldsymbol{w}) < 1 \qquad (18)$$

Because of the lack of closed-form expressions for the *pdf* of alpha-stable random variables, this approach is more attractive in our case than trying to minimize the IS sample variance given in equation (14).

6 REGENERATIVE SIMULATIONS

The regenerative approach to simulations is motivated by the fact that many stochastic systems have the property of starting afresh probabilistically from time to time; that is, whenever the regenerative condition is reached, the evolution of the system is independent of its past and governed by the same probability law. This enables us to separate the course of the simulation into *iid* blocks, called regenerative cycles (Crane and Lemoine 1977). When regenerative methods are not used, Importance Sampling "breaks down" for long simulations (Glynn and Iglehart 1989), in the sense that the typical likelihood ratio goes to zero (due to the fact that the sample space of the random process $\boldsymbol{W}$ increases exponentially with the simulation length), making it necessary to collect an increasingly larger number of samples in order to obtain a significant enough estimate. The effect of this *breakdown* is that, even though IS estimators are unbiased, the estimate can be several orders of magnitude smaller that the actual value when the number of samples is small (Devetsikiotis and Townsend 1993). When using the regenerative approach, on the other hand, since the likelihood ratio is applied within each cycle, it is maintained within reasonable bounds regardless of the overall simulation time.

Let $\{\beta_1, \beta_2, \cdots, \beta_M\}$ be the regeneration epochs, such that $1 \stackrel{\Delta}{=} \beta_0 < \beta_1 < \cdots < \beta_M$. Consider the input process instance $\boldsymbol{w} = (w_1, w_2, \cdots, w_K) \stackrel{\Delta}{=} (\tilde{\boldsymbol{w}}_1, \tilde{\boldsymbol{w}}_2, \cdots, \tilde{\boldsymbol{w}}_M)$, where $\tilde{\boldsymbol{w}}_i \stackrel{\Delta}{=} \left(w_{\beta_{i-1}}, w_{\beta_{i-1}+1}, \cdots, w_{\beta_i - 1}\right)$, for $i \in \{1, 2, \ldots, M\}$. The portion of the simulation having $\tilde{\boldsymbol{w}}_i$ as input is an *iid* replica of the portion having $\tilde{\boldsymbol{w}}_j$ as input, for $i, j \in \{1, 2, \ldots, M\}$, $i \neq j$. Following with our telecommunications emphasis, if our goal is to estimate the proportion of cells with a certain property (lost due to buffer overflow or having excessive delay), then:

$$X(\boldsymbol{w}) = \frac{N(\boldsymbol{w})}{D(\boldsymbol{w})} \qquad (19)$$

where $N(\boldsymbol{w})$ is the number of cells with the specified property and $D(\boldsymbol{w})$ is the total number of observed cells. From here:

$$X(\boldsymbol{w}) = \frac{\frac{1}{M} \cdot \sum_{i=1}^{M} N(\tilde{\boldsymbol{w}}_i)}{\frac{1}{M} \cdot \sum_{i=1}^{M} D(\tilde{\boldsymbol{w}}_i)} \xrightarrow{M >> 1} \frac{\overline{N(\tilde{\boldsymbol{w}})}}{\overline{D(\tilde{\boldsymbol{w}})}} \qquad (20)$$

For the queueing system that we want to analyze, the regenerative condition would be met when the queue is found empty.

7 TWISTED DENSITY FUNCTION

We will consider three different ways of transforming the input process: *i)* by modifying the mean arrival rate m described in equation (6); *ii)* by modifying the stability coefficient α of the FSN process Y_j, mentioned in equation (4); and *iii)* by modifying the innovation dispersion parameter γ_ε, also mentioned in equation (4). Since we want to observe buffer overflows and/or excessive delays, option *(i)* above is intended to increase the average traffic load, while options *(ii)* and *(iii)* intend to intensify the burstiness of the source by increasing the probability that Y_j has bigger values. In what follows, we will evaluate the potential performance of each one of these options using the *uniformly bounded likelihood ratios* criterion, described in section 5.2. According to the discussion in section 4, regarding the asymptotic behavior of a queueing system with FSN input traffic, we will select a twisted density as if we were dealing with an individual alpha-stable random *variable*.

7.2 Modifying the Mean Arrival Rate

Suppose we have two alpha-stable random variables Y_1 and Y_2, such that $Y_1 = Y_2 + \mu$, where μ is a constant. The likelihood ratio relating these two variables for a value y beyond the threshold is given by:

$$L(y) = \frac{f_{Y_1}(y)}{f_{Y_2}(y)} \xrightarrow[y>>1]{} \left(\frac{y}{y-\mu}\right)^{-\alpha-1} \approx 1 \qquad (21)$$

This approach is asymptotically inefficient in the sense that $L(y)$ is uniformly bounded, but the bound is very close to 1 when the threshold is very large. An additional disadvantage of this approach is that, if a mean arrival rate is chosen that is very close to the service rate or greater, the regeneration period mentioned in section 6 will increase, reducing the effectiveness of the regenerative approach.

7.3 Modifying the Stability Coefficient

Now, suppose that the two random variables Y_1 and Y_2 have a different stability coefficient. The likelihood ratio is now given by:

$$L(y) = \frac{f_{Y_1}(y)}{f_{Y_2}(y)} \xrightarrow[y>>1]{} \frac{\alpha_1 \cdot K_{\alpha_1}}{\alpha_2 \cdot K_{\alpha_2}} \cdot \left(\frac{y}{\sigma_Y}\right)^{-(\alpha_1-\alpha_2)} \qquad (22)$$

where $K_\alpha \overset{\Delta}{=} \dfrac{1-\alpha}{2\cdot\Gamma(2-\alpha)\cdot \mathrm{Cos}(\pi\,\alpha/2)}$

In the previous equation, σ_Y is the scale parameter of both Y_1 and Y_2. It can be seen from equation (22) that a great variance reduction can be achieved when using this approach, as long as α_2 is smaller than α_1.

7.4 Modifying the Innovation Dispersion Parameter

The effect of changing the innovation dispersion parameter of the process is equivalent to multiplying each sample Y_j by a constant factor $(\gamma'_\varepsilon/\gamma_\varepsilon)^{1/\alpha}$, where γ'_ε is the new dispersion parameter. Assume that the two random variables Y_1 and Y_2 are now related by $Y_2 = (\sigma_2/\sigma_1)\cdot Y_1$. This time, the likelihood ratio is given by:

$$L(y) = \frac{f_{Y_1}(y)}{f_{Y_2}(y)} \xrightarrow[y>>1]{} \left(\frac{\sigma_1}{\sigma_2}\right)^{\alpha} \qquad (23)$$

This method can give some variance reduction, but it is not as efficient as the one described in section 7.2.

8 LIKELIHOOD RATIO OF THE TRANSFORMATION

According to the discussion in the previous section and based on the asymptotic behavior of the queueing system, we selected the method of changing the stability parameter of our traffic in order to intensify the burstiness of the input stream. We will not try to maximize the variance reduction achieved because, as mentioned before, we do not have closed form expressions for the *pdf* of alpha-stable random variables. Thus, the modified process that we propose to use is:

$$W'_j = m + Y'_j \qquad (24)$$

where, Y'_j has now the conditional distribution $S_{\alpha'}\left((\gamma_\varepsilon)^{1/\alpha}, 0, \hat{Y}_j\right)$, where $\hat{Y}_j$, γ_ε, and α are as given in section 2 and α' is the new stability coefficient. Now, assume that a sample path $\boldsymbol{w}$ of observed traffic consists of K samples $\{w_1, w_2, w_3, \ldots, w_K\}$, then:

$$f_{\boldsymbol{W}}(\boldsymbol{w}) = f_{W_1}(w_1)\cdot f_{W_2}(w_2 \mid w_1)\cdot\ldots\cdot f_{W_K}(w_K \mid w_1, w_2, \ldots, w_{K-1}) \qquad (25)$$

A similar expression applies to $f_{W'}(\boldsymbol{w})$. From here we can conclude that:

$$L(\boldsymbol{w}) = \prod_{j=1}^{K} L_j(\boldsymbol{w}) \qquad (26)$$

where, using equations (4) and (25) we have:

$$L_j(\boldsymbol{w}) = \begin{cases} \dfrac{f_{W_1}(w_1)}{f_{W_1}'(w_1)} & ; \text{ for } j = 1 \\ \dfrac{f_{W_j}(w_j \mid w_{j-1}, w_{j-2}, \ldots, w_1)}{f_{W_j}'(w_j \mid w_{j-1}, w_{j-2}, \ldots, w_1)} & ; \text{ for } 2 \le j \le N \\ \dfrac{f_{W_j}(w_j \mid w_{j-1}, w_{j-2}, \ldots, w_{j-N})}{f_{W_j}'(w_j \mid w_{j-1}, w_{j-2}, \ldots, w_{j-N})} & ; \text{ for } j > N \end{cases} \qquad (27)$$

From equation (6) we obtain:

$$\begin{aligned} & f_{W_j}(w_j \mid w_{j-1}, w_{j-2}, \ldots, w_{j-N}) \\ & = f_{Y_j}\left(w_j - m \mid y_{j-1}, y_{j-2}, \ldots, y_{j-N}\right) \end{aligned} \qquad (28)$$

Now, from equation (5) and the properties of alpha-stable random variables:

$$L_j(\boldsymbol{w}) = \frac{f_\alpha\left(\dfrac{w_j - m - \hat{Y}_j}{\sigma_\varepsilon}\right)}{f_{\alpha'}\left(\dfrac{w_j - m - \hat{Y}_j}{\sigma_\varepsilon}\right)} \qquad (29)$$

In equation (29), $\sigma_\varepsilon \stackrel{\Delta}{=} (\gamma_\varepsilon)^{1/\alpha}$ represents the innovation scale parameter of both the original and the modified processes, and $f_\alpha(\cdot)$ is the *pdf* corresponding to a normalized $S_\alpha(1,0,0)$ random variable.

9 RESULTS

As a specific example, we are including in this section the results obtained using both direct and fast simulation for a simple system consisting of a traffic source and a server with a constant service rate, as shown in Figure 1. The goal was to estimate the blocking probability (or probability of packet loss) of the queue.

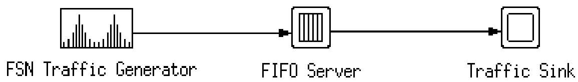

Figure 1: Configuration Used in our Simulations

The traffic source is modeled as a modified balanced Linear FSN process with α = 1.95 and H = 0.903. The average arrival rate is assumed to be 1965 cells/s, or 833 Kbps. These parameters are compatible with those found in (Gallardo, Makrakis, and Orozco-Barbosa 1998) for a traffic source generating an aggregate VBR video stream, which could correspond to a video-on-demand service provider. The service rate of the FIFO server in Figure 1 is 2358.5 cells per second or 1 Mbps, approximately 20% greater than the mean arrival rate. The modified stability coefficient α' used for the twisted density function was 1.6. A set of 100 *iid* simulation was run for both the direct and the fast algorithms. The simulations were run for 2000 seconds.

Figure 2 compares the results obtained from direct and fast simulation regarding the blocking probability vs. buffer size in the server. It can be observed from that figure that the results are satisfactorily similar for buffer sizes of 333, 1000, and 3000 cells. The direct simulation gives a relatively deviated output for a buffer size of 9000 cells. Direct simulation proved incapable of estimating the blocking probability when the buffer size is greater than 9000 cells (it gave zero probability), since the event becomes too rare to be observed even after 100 simulation runs.

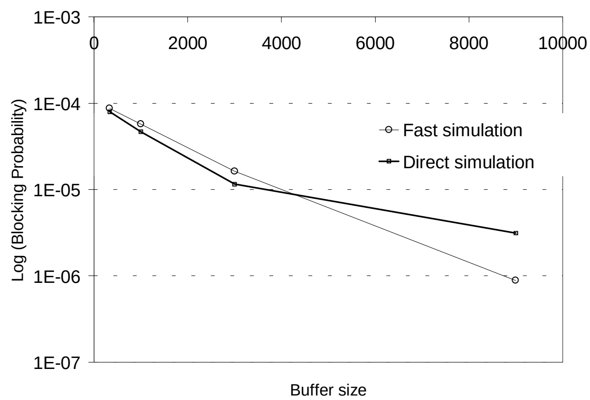

Figure 2: Comparison of Results Obtained from Direct and Fast Simulation

In addition to the results shown in Figure 2, Figure 3 shows the results obtained using the fast simulation algorithm for buffer sizes of 27000, 54000 and 108000 cells.

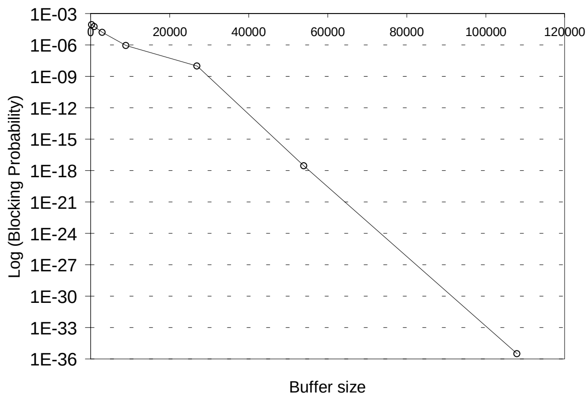

Figure 3: Results Obtained from Fast Simulation

Table 1 shows the estimated as well as the observed variance reduction for each case. The estimated variance reduction (referred to as k in equation (16)) is calculated according to the asymptotic behaviour and equation (22).

Table 1: Estimated and observed variance reduction

Buffer size (cells)	**Estimated variance reduction**	**Observed variance reduction**
333	0.0693	0.2400
1000	0.0668	1.2000
3000	0.0644	1.4000
9000	0.0620	0.0395
> **9000**	$\lesssim$ 0.06	Undefined

10 CONCLUSIONS

A technique for the fast simulation of broadband communications systems has been proposed. This technique is applicable when modeling the offered traffic using Fractional Stable Noise processes, which have been recently proved to be very accurate in capturing the long-range dependence and burstiness of today's aggregate network traffic. The fast simulation algorithm proposed in this paper is based on regenerative Importance Sampling techniques. It has been shown that there is a satisfactory agreement between the results obtained with both fast and direct simulations when the event analyzed is not too rare and that the fast algorithm provides a noticeable variance reduction when the relevant event is rare.

REFERENCES

Bucklew, J. A. 1990. "*Large Deviation Techniques in Decision, Simulation, and Estimation*", New York; Toronto: Wiley.

Crane, M. A., Lemoine, A. J. 1977. "*An Introduction to the Regenerative Method for Simulation Analysis*", Berlin; New York: Springer-Verlag.

Devetsikiotis, M., Townsend, J. K. 1993. "*An Algorithmic Approach to the Optimization of Importance Sampling Parameters in Digital Communication System Simulation*", IEEE Transactions on Communications, Vol. 41, No. 10 (Oct), pp 1464-1473.

Duffield, N.G., O'Connell, N. 1995. "*Large deviations and overflow probabilities for the general single-server queue, with applications*", Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 118, pp 363-374. Also Dublin Institute for Advanced Studies - Applied Probability Group Technical Report DIAS-STP-93-30.

Feller, W. 1966. "*An Introduction to Probability Theory and its Applications*", Vol. II, New York: Wiley.

Gallardo, J.R., Makrakis, D., Orozco-Barbosa, L. 1998. "*Use of Alpha-Stable Self-Similar Stochastic Processes for Modeling Traffic in Broadband Networks*". SPIE's Conference on Performance and Control of Network Systems II. Boston, Massachusetts, November 1-6, 1998. Extended version to appear in Performance Evaluation Journal (1999).

Gallardo, J.R., Makrakis, D., Orozco-Barbosa, L. 1999. "*Fast Generation of Artificial Traces of Alpha-Stable Long-Range Dependent Stochastic Processes*". Submitted to IEEE Transactions on Signal Processing (Sep. 1998).

Glynn, P. W., Iglehart, D. L. 1989. "*Importance Sampling for Stochastic Simulations*", Management Science, Vol. 35, No. 11 (Nov), pp 1367-1392.

Huang, C., Devetsikiotis, M., Lambadaris, I. 1995. Kaye, A. R., "*Fast Simulation for Self-Similar Traffic in ATM Networks*", Proceedings of IEEE International Conference on Communications (ICC'95), Seattle, Washington, June 1995.

Juneja, S. K. 1994. "*Efficient Rare Event Simulation of Stochastic Systems (Polling Systems)*", Ph.D. Thesis, Stanford University, Department of Operations Research.

Karasaridis, A., Hatzinakos, D. 1998. "*Broadband Heavy-Traffic Modeling using Stable Self-Similar Processes*", Proceedings of 2nd Canadian Conference on Broadband Research, Ottawa, Canada, June 1998.

Li, J-S, Wolisz, A., Popescu-Zeletin, R. 1998. "*Fast Simulation of Self-Similar Traffic*", Proceedings of IEEE International Conference on Communications (ICC'98), Atlanta, Georgia, June 1998.

Norros, I. 1995. "*On the Use of Fractional Brownian Motion in the Theory of Connectionless Networks*", IEEE JSAC, Vol. 13, No. 6 (Aug.), pp 953-962.

Samorodnitsky, G., Taqqu, M. S. 1994. "*Stable Non-Gaussian Random Processes*", New York: Chapman & Hall.

Willinger, W. Taqqu, M. S., Sherman, R., Wilson, D. V. 1997. "*Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the*

Source Level", IEEE/ACM Transactions on Networking, Vol. 5, No. 1 (Feb.), pp 71-86.

AUTHOR BIOGRAPHIES

JOSE R. GALLARDO received the B.S. degree in Physics and Mathematics from the National Polytechnic Institute in México City, México, in 1988, and the M.Sc. degree in Electrical Engineering from CICESE Research and Higher Education Center in Ensenada, México, in 1991. From 1991 to 1993 he worked as part of the academic staff at CICESE, from where he is currently on leave of absence. He is a candidate for the D.Sc. degree in Electrical Engineering at the George Washington University, Ashburn, VA, USA, and is working as a Research Associate at the Advanced Communications Engineering Centre of the University of Western Ontario, London, ON, Canada. His main areas of interest are traffic modeling, traffic control, and simulation of broadband communications networks.

DIMITRIOS MAKRAKIS received the Diploma from the University of Patras, Greece, in 1982, the M.A.Sc. degree from the University of Toronto, Canada, in 1984, and the Ph.D. degree from the University of Ottawa, Canada, in 1992, all in Electrical Engineering. In 1996, he joined the Department of Electrical and Computer Engineering of the University of Western Ontario as Assistant Professor and Executive Director of the Advanced Communication Engineering Centre. His research interests are in the area of networks and communication system technology, high speed networks, personal communications networks, teletraffic studies, intelligent networks and performance evaluation.

LUIS OROZCO-BARBOSA received the B.S. degree in Electrical Engineering and Computer Science from the Metropolitan Autonomous University (UAM), Mexico, in 1979, the Diplome D'Etudes Approfondies from ENSIMAG, France, 1984, and the Doctorat de l'Université from Université Pierre et Marie Curie, Paris, France, in 1987, both in Computer Science. In 1987, he joined the University of Ottawa, where he is currently an associate professor at the Electrical Engineering Department. His current research interests include broadband switching architectures, multimedia applications, traffic modeling and management, intelligent networks, and performance evaluation of high-speed networks.