On the Minimum Number of Users for the Generation of $H$-ss traffic in a Simulation Scenario

Candelaria Sansores Pérez
Universidad del Caribe
Complex Systems Simulation Lab
SM 78, Lote 1, Mza 1
Cancún, México
csansores@ucaribe.edu.mx

Luis Rizo Domínguez
Universidad del Caribe
Complex Systems Simulation Lab
SM 78, Lote 1 Mza 1
Cancún, México
lrizo@ucaribe.edu.mx

Julio César Ramírez Pacheco
Universidad del Caribe
Complex Systems Simulation Lab
SM 78, Lote 1 Mza 1
Cancún, México
jramirez@ucaribe.edu.mx

Abstract: Several simulation and measurement studies have reported the well-known network traffic self-similarity and long-range dependency and the consequences for network performance. In the context of simulation, however, the impact of the number of sources has not been sufficiently emphasized for the generation of synthetic self-similar traffic. In this paper we describe a simulation scenario suitable for the testing of performance issues under self-similar traffic. Our analysis was centered on the effect of traffic aggregation over the self-similarity degree, determining the necessary number of sources to approach the verified relation $H = (3 - \min \alpha)/2$. Besides, we highlighted the performance of several Hurst parameters estimators for this type of simulation scenarios, identifying the most suited ones.

Key–Words: Self-similar traffic, Heavy-tail distributions, Performance evaluation, Simulation

1 Introduction

Simulation plays an important role for the design, performance evaluation and dimensioning of computer networks. Diverse network features can be studied with the aid of simulation scenarios. The effect of traffic behaviour on the QoS metrics such as delay, delay jitter, packet loss, etc. is such an example. In this context, packet network traffic has shown to be of self-similar nature[2][3]. Thus, current simulation scenarios must take this behaviour into account. A well known and amply cited simulation scenario is given in [1][4], where self-similar traffic was generated by the transmission of files of size $Z$ by an ensemble of $i = 32$ users. A particular feature of this scenario is that the distribution of files, $Z_i$, transmitted by user $i$ is heavy-tailed with parameter $\alpha_i$, giving rise to highly variable file sizes. In the limit as the number of users $i \to \infty$, the traffic in the network node is self-similar with $H = (3 - \min \alpha_i)/2$. Unfortunately, the Hurst parameter obtained in that paper is highly variable and overestimates when $\alpha > 1.6$ and underestimates when $\alpha < 1.6$. From the above it is noted that the scenario described in [1][4] cannot be used for simulation studies where accurate tuning of the Hurst parameter is required. In addition, estimators used to test the presence of self-similar behaviour are not the most robust. In this paper, we propose some changes to the simulation scenario and determine the necessary (and finite) number of users to approach the limit $H = (3 - \min \alpha_i)/2$ with minimum variation. It is shown that our simulation scenario can be used to effectively generate a self-similar process with Hurst parameter $H$, where $H$ shows little variation. Thus, we propose a simulation scenario which can be used to study the behaviour of network algorithms under self-similar traffic, and where $H$ can be effectively and accurately tuned by the values of $\alpha$. Also, we complement the study presented in [1][4] by including in our study several estimators of Hurst-index. In this context, the paper is organized as follows, section 2 reviews fundamentals concepts related to heavy-tail distributions, self-similar processes and the methods for generating self-similar processes from heavy-tailed distributions. It also reviews the main estimators for both of them. Section 3 provides description of the simulation scenario and points out the differences with the one described in [1][4]. Section 4 shows the results of the simulation and finally section 5 presents the concluding remarks.

2 Internet Traffic Models

2.1 Heavy-Tail Distributions

Heavy-tailed distributions are distribution functions whose tails $P(X > x)$ and $P(X \leq -x)$, for positive $x$, decrease slower than exponential rate. The latter, e.g., normal and exponential distributions, are
2.2 Estimators of tail-index

Several estimators of the tail-index $\alpha$ have been proposed, next subsections review Hill-based and QQ plots used for estimating $\alpha$.

2.2.1 Standard Hill Estimator

Let $X_1, X_2, ..., X_n$ be a discrete time series with distribution $F_X(x)$. Now let $X_1 > X_2 > ... > X_n$ denote the ordered statistics of time series $X_1, ..., X_n$. The Hill estimator of $\gamma = \alpha^{-1}$ based on $k + 1$-upper ordered statistics, $1 < k \leq n$, is defined according to the following formula:

$$H_{k,n} = k^{-1} \sum_{i=1}^{k} \log \frac{X(i)}{X(k+1)}.$$  \hspace{2cm} (2)

The parameter $\alpha$ is estimated by plotting $k$ versus $H_{k,n}$ for $1 < k \leq n$ and looking for a stable region in the plot. The stable region must sit at height $\alpha$. Usually the Hill estimator works better when the underlying heavy-tailed distribution is Pareto. When the distribution is not of Pareto-type the Hill estimator shows volatility, i.e., irregular erratic behavior.

2.2.2 Smooth Hill Estimator

The smooth Hill estimator, $\text{smooHill}$, is obtained by applying a smoothing technique to the standard Hill estimator in order to reduce the volatility in the standard Hill plot. Let again $X_1 > X_2 > ... > X_n$ be the ordered statistics, the $\text{smooHill}$ estimator is defined as

$$\text{smoo } \hat{\alpha}_{k,n,u} = \frac{1}{(u-1)k} \sum_{j=k+1}^{u} H_{j,n},$$ \hspace{2cm} (3)

where $u \in \{2, 3\}$. Again a plot of $k$ versus $\text{smoo } \hat{\alpha}_{k,n,u}$ should stabilize at a region $\hat{\alpha}$.

2.2.3 Alternative Hill Estimator

Another variant of the standard Hill estimator is the alternative Hill estimator, $\text{altHill}$, which changes the scale of the Hill estimator. The $\text{altHill}$ estimator can be applied to the standard Hill estimator and the $\text{smooHill}$ estimator. When applied to the $\text{smooHill}$ estimator, it results in the $\text{altsmooHill}$ estimator of $\hat{\alpha}$. The $\text{altHill}$ estimator is defined as

$$H_{[n^\theta],n} = [n^\theta]^{-1} \sum_{i=1}^{[n^\theta]} \log \frac{X(i)}{X([n^\theta]+1)},$$ \hspace{2cm} (4)

where $[y]$ is the smallest integer greater of equal to $y \geq 0$. For the estimation of $\hat{\alpha}$ we plot $\theta$ versus $H_{[n^\theta],n}$ for $0 \leq \theta \leq 1$. The stable region in the plot should be the estimated value of $\hat{\alpha}$.

2.2.4 QQ-Plot

Let $X = (X_1, X_2, ..., X_n)$ be i.i.d observations with common distribution $F$. Now let $X_1 > X_2 > ... > X_n$ be the upper order statistics of $X$, i.e., $X(i) > X(j)$ iff $i < j$. Pick $k$ upper order statistics and neglect rest $k + 1$. The distribution of the $k$ exceedances, $X_1, ..., X(k)$ should be Pareto if $F$ is heavy-tailed. Taking the logarithm of the $k$ exceedances makes its distribution approximately exponential, thus the plot of the empirical quantiles of the exceedances against the theoretical quantiles of the exponential distribution should yield a straight line with slope $\alpha^{-1}$. More formally the plot of

$$\{( \log(1 - \frac{j}{k+1}), \log X(k+j+1), 1 \leq j \leq k)\},$$ \hspace{2cm} (5)

should yield approximately a $\alpha^{-1}$ slope straight line if the distribution of $X_1, X_2, ..., X_n$ satisfies the asymptotic behavior of (1). The slope of the line is computed.
by least squares regression through the points in (5) and is called the QQ estimator, i.e.,

$$\bar{\alpha}_{1,k,n}^{-1} = \frac{\sum_{i=1}^{k}(\nu_{i,k})\xi_{i,k} - \sum_{i=1}^{k}(\nu_{i,k})H_{k,n}}{k\left(\sum_{i=1}^{k}(\nu_{i,k})^2 - (\frac{1}{k}\sum_{i=1}^{k}\nu_{i,k})^2\right)},$$

(6)

where $\nu_{i,k} = -\log(i/k)$ and $\xi_{i,k} = \log\left(\frac{X(i)}{X(i+k)}\right)$. There are two different versions of the QQ-plot, namely the dynamic and static QQ-plot. The dynamic QQ-plot is similar to the Hill plot and is obtained by plotting $\{(k,1/\bar{\alpha}_{1,k,n}), 1 \leq k \leq n\}$ and finding a stable region in the plot. The static plot is obtained by choosing an appropriate value of $k$, plotting the points in (5) and finding a region where the plots looks linear, then in the linear region apply (6) which should yield the value of $\bar{\alpha}_{1}^{-1}$.

2.3 Self-similarity and long-memory

Processes with some form of scaling behaviour can be defined as stochastic signals possessing invariance properties on all or a set of scales(i.e., no characteristic scale can be identified). Examples of such processes include self-similar[26], long-memory, fractal and multifractal processes[27][24][25][28]. The paper deals with self-similar and long-memory processes, the most known of them. Strict self-similar signals(H-ss), $X = \{X_t, t \in \mathbb{R}\}$, are defined as those for which appropriate changes of scale of time and space do not vary its statistical properties, i.e., processes for which $X_{ot} = a^H X_t$, for any $t \in \mathbb{R}$, $a, H > 0$, where the equality is in terms of finite-dimensional distributions. Weak self-similarity, a more often used version, is defined as processes for which $\mathbb{E}X_{ot}X_{as} = a^{2H}\mathbb{E}X_tX_s$, for any $t, s \in \mathbb{R}$, $a, H > 0$. Note that strict self-similarity implies nonstationarity, long-memory processes on the other hand is often defined for stationary processes. Long-memory property of finite-variance stationary signals $Y = \{Y_t, t \in \mathbb{R}\}$ is possessed if $\mathbb{E}Y_{t}Y_{t+\tau} \sim c_\gamma |\tau|^\beta-1$ (equivalently as its PSD $f(\nu) \sim c_f |\nu|^{-\beta}$) as $\tau \to \infty$ (as $\nu \to 0$). Indeed, a strong relationship between these two processes exists and a given self-similar process(H-ss) with stationary increments(H-ssi) possess long-memory in its first increment process, i.e., $\mathbb{E}Y_{1}Y_{1+\tau} \sim c_\gamma t^{\beta-1}$ provided $Y = \Delta^1X(t; 1) = X(t + 1) - X(t)$ and $X$ belongs to the space of finite variance H-ssi processes. The above for example holds true for the unique Gaussian H-ssi process, namely, fractional Brownian motion(fBm) with $0 < H < 1$. Many estimators of Hurst-index have been proposed[17][33][34], R/S statistic, variance based(aggregated, differenced, detrended), periodogram-based(GPH, cumulated, whittle), wavelet based estimators(aby, delbeke)[20][30], etc.

2.4 Estimators of the self-similarity parameter

2.4.1 R/S Statistic:

The R/S Statistic[6][8][9][10][34] developed by E. Hurst when studying Nile river is defined for a process $Y(t)$ in the interval $(\tau, \tau + n)$ as

$$\frac{R}{S}(\tau, n) := \max(W(\tau, n)) - \min(W(\tau, n)).$$

(7)

where $W(\tau, n) = Y(\tau + u) - Y(\tau) - uE(\tau, n)$ and $E(\tau, n)$ and $S(\tau, n)$ denote the mean and standard deviation in the interval $(\tau, \tau + n)$. Hurst found that for long-memory records, (7) behaves as $E\left(\frac{R}{S}(\tau, n)\right) \sim n^H, H > 0.5$, in constrast, short-memory processes follow $E\left(\frac{R}{S}(\tau, n)\right) \sim n^{0.5}$. A log-log plot of the mean values of the R/S statistic values versus $n$ is an estimator of $H$.

2.4.2 Block averaged methods: Variance and Absolute Moment

Consider the aggregated series $\Gamma_{m}(\{x_i\}) = X_i^{(m)}$ of a length $N$ time series[11][13]. The sample variance of the block averaged process $\text{Var}\{\Gamma_{m}(\{x_i\})\}$ for long-memory series behaves asymptotically as $\text{Var}\{\Gamma_{m}(\{x_i\})\} \sim cm^{-\beta}$, where $c$ is a constant and $\beta = 2 - 2H$. From this result, a log-log plot of $\text{Var}\{\Gamma_{m}(\{x_i\})\}$ versus $m$, for different values of $m$, and such that $m_{i+1}/m_i = C \in \mathbb{R}^+$ is an estimator of $H$. The absolute moment of $X_i^{(m)}$, $\text{AM}(m)$, behaves asymptotically as $\text{AM}(m) \sim m^{-\beta/2}$, thus a log-log plot of $\text{AM}(m)$ versus $m$ results in a line with slope $-\beta/2 = H - 1$ from which $H$ is inferred. First method is called the variance method and the latter the absolute moment one.

2.4.3 Periodogram based methods: Periodogram and Whittle

The periodogram, $I(\nu) = 1/(2\pi N)|\sum_{j=1}^{N}X_j e^{ij\nu}|^2$ for the series $\{X_j\}$ is also an estimator of $H$. The periodogram for a long-memory time series behaves as $I(\nu) \sim |\nu|^{1-2H}$ for $\nu \to 0$, therefore a log-log plot of $I(\nu)$ versus $\nu$ is used to obtain
H. The Whittle method [33][29][18][34][6] is a non-graphical MLE estimator strongly related to the periodogram defined by the following relation \( Q(\eta) := \int_{-\pi}^{\pi} \left( I(v) f(v; \eta) \right) dv + \int_{-\pi}^{\pi} \log(f(v; \eta)) dv \), where \( \eta \) is a vector of unknown parameters and \( f(v; \eta) \) is the spectral density at frequency \( v \) of the studied function, the value of vector \( \eta \) that minimizes the function \( Q \) is considered the Whittle Estimator. A discretized version of \( Q(\eta) \) is obtained as \( Q^*(\eta) = \sum_{j=1}^{(N-1)/2} \frac{I(v_j) dv_j}{f^*(v_j; \eta)} \) where \( N \) is the series length. The Whittle MLE specifies the functional form of the spectral density at all frequencies and the Local Whittle[33][29][6] assumes only the functional form when \( v \) approaches zero, namely \( f(v) \sim G(H) |v|^{-2H} \) as \( v \to 0 \) and from \( Q^*(\eta) \) the task is reduced to minimize the function

\[
R(H) = \log \left( \frac{1}{M} \sum_{j=1}^{M} I(v_j) \right) - (2H-1) \frac{1}{M} \sum_{j=1}^{M} \log v_j
\]

Its computation involves the introduction of the parameter \( M \) which is an integer less than \( \frac{N}{2} \), and satisfying \( \frac{1}{M} + \frac{M}{N} \to 0 \) as \( N \to \infty \).

### 2.4.4 Wavelet based methods

Let \( d_x(i, j) \) denote the wavelet coefficients of a particular finite length sequence \( \{x_i\} \), it is known that for long-memory processes the variance at level \( i \) of the coefficients is given by \( \text{Var}(d_x(i, j)) = \frac{\sigma^2}{2} V_\psi(H)(2^j)^{2H+1} \), where \( V_\psi(H) \) depends on the particular wavelet and the Hurst-index and is defined by:

\[
V_\psi(H) = - \int_{-\infty}^{\infty} \gamma_\psi(\tau) |\tau|^{2H} d\tau
\]

taking the logarithm at \( \text{Var}(d_x(i, j)) \) should result in \( \log(\text{Var}(d_x(i, j))) = (2H + 1)j + K \), where \( K \) is a constant. Abry and Veitch have suggested an Hurst-index estimator based on this behaviour using Daubechies wavelets[30][16][20]. First a time average \( \mu_i \) of \( d_x(i, j) \) is computed at a given scale, where \( \mu_i \) is defined as \( \mu_i = (n_i)^{-1} \sum_{j=1}^{n_i} d^2_x(i, j) \), where \( n_i \) is the wavelet coefficient number at scale \( i \) and \( n \) the time series points. The estimated Hurst-index is then obtained from the slope of a linear regression method for \( \log(\mu_i) = \log(\frac{1}{n_i} \sum_{j=1}^{n_i} d^2_x(i, j)) \), where \( i = 1, 2, \ldots, \left\lfloor \log_2(n) \right\rfloor \).

### 2.4.5 Sources of inaccuracies

Algorithms’ accuracy are often affected by some parameters such as cut-off selection, number of aggregation levels and the minimum number of points in block size in regression based methods. Also other parameters include number of frequencies for periodogram methods, beginning and ending octave, etc. These parameters are sources of inaccuracies and bias the estimates. They must be selected carefully.

### 2.5 Self-similarity through high-variability

Self-similar traffic can be generated using the Lamperti transformation based on a stationary stochastic process or can be generated by the superposition of an infinite number of users which are superposed in a node. In this paper we concentrate in the generation of self-similar traffic based on heavy-tailed distributions. Let \( X_i \) be a random variable with a heavy-tailed distribution. Suppose the random variable can represent the file size of traffic source \( i \) or the period of transmission between successive packets. As the number of users \( i \to \infty \), then, the traffic aggregated(or superposed at a node) is self-similar with self-similarity parameter \( H = (3 - \min_\alpha_1)/2[14][15][31]. \) We used the high variability of interdepartures times for the generation of self-similar traffic.

### 3 Simulation scenario

This section presents the proposed simulation scenario which generates self-similar traffic with parameter \( H \). This scenario turns to be an appropriate model for simulations where the degree of traffic self-similarity needs to be finely and precisely adjusted. The simulation scenario is shown by the network model of figure 1. In this network, self-similar traffic is generated by an ON/OFF model, where the ON and/or OFF are heavy-tailed [14][15]. Although the required number of independent users should be infinite along this model, in practice, this condition is not feasible giving rise to self-similar traffic generators using diverse number of sources. This diversity has an important impact over traffic self-similarity generation and measurement. For instance, in the works [1][4], oriented to study the relationship between file sizes and self-similarity phenomena, the numbers of sources was set up to \( i = 32 \) and its variation seemed not to be significant to their results. In
contrast, in the experiments we performed, a significant relationship between this parameter and the generated self-similar traffic was found and thus, the network configuration of figure 1 was proposed in order to calibrate this feature. As shown in the figure 1, the network consists of $i$ nodes(or sources) and $l$ output links in a packet switched configuration. The parameter $i$ is customizable and represents statistically independent UDP sources, i.e, $S_1, S_2,$ $\ldots, S_i$ are i.i.d. $N_1$ and $N_2$ represents routers through which packets from sources $S_i$ are processed and forwarded to the destination sources $R_i$. In our configuration $N_1$ represents the node over which traffic is superposed and thus represents our measurement point. Queue length of node $N_1$ was set up to 1000 packets with buffer size of 312.5kB and its output link bandwidth was set up to 32.768Mb and latency of 30ms. Each link from $S_i$ to $N_1$ and from $R_i$ to $N_2$ has a bandwidth of 8.2Mb and a latency of 20ms. In order to obtain self-similar traffic in $N_1$, the traffic sources $S_i$ have a Pareto random variable generator for the inter-departure time $t_i$. Recall that if $t_i$ is a Pareto random variable, its CDF is given by:

$$P(t_i \leq t) = 1 - \left(\frac{t_{\min}}{t}\right)^\alpha,$$

where the minimum value of $t_i$ is $t_{\min}$ and $\alpha$ is the tail-index. The Pareto random variable has infinite variance when $1 < \alpha < 2$. In this case the mean is finite. Then, in order to keep $E(t) < \infty$ for all sources, the simulation was performed for $\alpha \in \{1.1, \ldots, 1.9\}$. Even though the mean of the inter-departure time rely upon the value of $\alpha$, in our configuration it is constant, i.e., $E(t) = 500\mu$s for all given values of $\alpha$. Likewise to normalize the data mean rate for all sources, $t_{\min}$ was tuned to each values of $\alpha$, with an initial value of 0.041ms and fixed packet size of 320 bytes. The network configuration just reviewed was used in all the experiments of the paper. We used the well-known network simulator ns-2 in a 2x2.8GHz Quad-Core Intel Xeon Macintosh platform. All results were obtained from several hundreds of runs executed for 300 simulated seconds and varying number of sources.

4 Results

4.1 Generation of Pareto series

In order to check the correctness of our simulation scenario, we first test the appropriate generation of heavy-tailed traces in ns-2 from which users send the packets (recall that Pareto series simulate inter-departure times of packets). Figure 2 shows typical packet inter-departure time series trace in ns-2 with $\alpha \in \{1.2, 1.8\}$. Top plot correspond to Pareto series with $\alpha = 1.2$ while bottom plot to Pareto with $\alpha = 1.8$. Note that traces behave qualitatively as heavy-tailed process, i.e, extreme values occur frequently. As
above mentioned, the number of extremes values (i.e., of silent periods) in ns-2 generated Pareto series occur with a non-negligible probability. Note that the 'usual' values are below 100ms and that the lower the value of \( \alpha \) is the greater this value. A Hill-plot

![Figure 3: Typical Hill plots for ns-2 generated Pareto series](image)

and a CCDF plot will confirm the appropriate generation of Pareto series in our simulation scenario. Figure 3 shows the Hill plots associated to traces of figure 2. Top plot corresponds to Pareto with \( \alpha = 1.2 \) while the bottom to \( \alpha = 1.8 \). Note that Hill-plots stabilize in a region and this region corresponds to the true value of \( \alpha \). CCDF plots are also helpful for testing if a given model follows a particular probability distribution \( F_X(x) \). CCDF plots, therefore can be used to test if a series follows a Pareto distribution. Figure 4 shows the CCDF plots corresponding to traces presented in figure 2. Again as before top plot corresponds to ns-2 generated Pareto time series with \( \alpha = 1.2 \) while bottom plot to \( \alpha = 1.8 \). Note from the figure that both time series follow accurately the reference line corresponding to an exact Pareto time series with known \( \alpha \). From the above it is seen that the generation of Pareto time series with ns-2 is accurate since Hill and CCDF plots estimate correctly the given \( \alpha \). Similar results were obtained when analyzing other time series corresponding to a given source or sources.

### 4.2 Generation of Self-Similar series

In this subsection we experimentally verify that aggregate traffic from \( N \) sources, where each source send packets with inter-departure according to a Pareto series, follows and can be modelled by a self-similar process of parameter \( H \). Recall that as \( N \to \infty \), \( H = (3 - \alpha)/2 \). This subsection only test that aggregate traffic is indeed a self-similar process. Figure 5 shows typical traffic traces obtained in a network node in our simulation scenario. Note that the series obtained behaves in accordance with a self-similar trace. The bottom trace corresponds to a trace with \( H = 0.9 \)
and the top plot to a trace with $H = 0.95$.

### 4.3 Self-similarity and $\alpha$ relation

Figure 6 shows the simulation results when considering 30 traffic sources. The same number of traffic sources was used in [1][4]. As can be noted from the figure, the same kind of behaviour is obtained as those of [1][4]. Note that no estimator can follow the reference line $H = (3 − \min \alpha_i)/2$. In fact it is noted that the estimates $\hat{H} \sim H_{ref} + k$, where $k$ is a constant. From this it can be said that no simulation scenario, neither [1][4] nor the proposed by us is capable of finely and accurately generating self-similar traffic for performance purposes when $N = 32$. Precisely generating self-similar traffic is important for testing the behaviour of algorithms or novel protocols and checking its behaviour under varying degrees of correlation or persistence. In fact, this degree can be accurately varied based on the tail-index of the traffic source. Figure 7 shows the variation of the Hurst-index when estimated with five different estimators. Note that variance-type method presents high variability. R/S statistic presents low variability but unfortunately its bias is high. From the two figures is concluded that when using 30 traffic sources in the simulation scenario shown above, self-similar traffic is effectively generated but the Hurst-index of this generated traffic presents high bias and variability. Figure 8 shows the simulation results when using 50 traffic sources. Note that significant improvements in the bias are obtained. In fact, Whittle, wavelet and R/S statistic behave reasonable well. Periodogram and variance method present high bias and variability. Recall that the most robust estimators are those based on wavelets and the MLE estimation ones(Whittle). Our paper takes these estimators into account and our conclusions area based on the results obtained from these.
As above, figure 9 shows the standard deviation of the estimations of the Hurst-index but when using 50 traffic sources. Note that although R/S statistic shows low bias, the variance is high and thus is not suggested for deciding which the number of required traffic sources is. Whittle and wavelet are the most robust among all the estimators[33][34][12][20] studied and thus can be used for the task of deciding the required number of traffic sources for obtaining self-similar traffic with low-bias and variance. We also performed the same kind of analysis to 70 traffic sources obtaining similar results as those for 50, this behaviour can be observed in Figure 10. From the above figures, we can conclude that the required number of users neccessary to generate accurate self-similar traffic is at least 50 traffic sources. Also, variance and R/S statistic methods can not be used for such a task. In our work, Whittle and wavelet based methods were used to decide the required number of users for the accurate generation of self-similar signals. References [1][4] showed the results for 30 traffic sources and the methods used to test the presence were variance and R/S statistic. We suggest that the relationship between self-similarity parameter and QoS parameter presented in that paper must be re-evaluated.

5 Concluding Remarks

This paper described background information on self-similar processes and heavy-tailed distributions. It also detailed the simulation scenario for generating self-similar traffic from heavy-tailed sources. This scenario diverges from previous reported results in two aspects: i) the number of sources and ii) the increment in the number of Hurst parameter estimators evaluated. Results found show that the number of independent users impact the accuracy of the Hurst-index and conclude that the required number of independent traffic sources must be at least 50. Also, according to variation of Hurst-index estimation, the Whittle and wavelet methods were the most suited for this type of simulation scenarios. As further work we propose the analysis of the relationship of self-similarity parameter and QoS performance under the scenario proposed. Also, it would be interesting to include several self-similar traffic flows in the topology to study its relation on adjacent nodes and also to establish the mathematical relationship among Hurst-indexes.

Acknowledgements: This work has been supported by Project QROO-2008-01-92231 funded by Consejo Nacional de Ciencia y Tecnologia (CONACYT) and Project PROMEP/103.5/08/2968 funded by PROMEP program of Mexico. Authors would like to thank José Antonio Careaga for his help in this work.

References:


