The graphical methods for estimating Hurst parameter of self-similar network traffic

Mitko Gospodinov, Evgeniya Gospodinova

***Abstract**: The modern high-speed network traffic exhibits the self-similarity. The degree of self-similarity is measured by the Hurst parameter. In this paper are used two graphical techniques for estimating Hurst parameter of pseudo-random self-similar sequences, based on the fractional Gaussian noise (FGN) method. The analyses show that the FGN method always produces self-similar sequences, with relative inaccuracy of the estimating Hurst parameter below 8%. This generator should be recommended for practical simulation studies of high-speed telecommunication networks, since it is very accurate.*

***Key words**: Hurst parameter, fractal Gaussian noise, self-similar traffic, R/S plot, and variance-time plot.*

1. Introduction

Data traffic is main component of computer communication systems and traffic models are of importance for assessing their performance. Several recent teletraffic studies of communication networks, including wide area network (WAN), local area network (LAN) and World Wide Web have shown that commonly used teletraffic models, based on Poisson or related processes are not able capture the self-similar nature of teletraffic [2], [4]. These models of teletraffic result in overly optimistic estimates of performance of computer networks, insufficient allocation of communication and data processing resources, and difficulties in ensuring the quality of service expected by network users [3], [1].

The self-similarity means that the statistical properties (all moments) of a stochastic process do not change for all aggregation levels [5]. The main properties of self-similar processes include:

- Slowly decaying variance – the variance of the sample is decreased more slowly than the reciprocal of the sample size.
- Long-range dependence - the process is called a stationary process with long-range dependence if its autocorrelation function is non-summable. The speed of decay of autocorrelations is more hyperbolic than exponential.
- Hurst effect – it expresses the degree of self-similarity.

In this paper are used two graphical techniques for estimating Hurst parameter of pseudo-random self-similar sequences, based on the Fast Fourier Transform algorithm of synthesizing fractional Gaussian noise.

2. The graphical methods for estimating Hurst parameter

2.1 Rescaled adjusted range statistics (R/S method)

Statistical self-similar means that the statistical properties for the entire data set are same for sub-sections of the data set (the two halves of the data set have the same statistical properties as the entire data set). This is applied to estimating the Hurst exponent, where the rescaled range is estimated over sections of different size. As shown in Fig.1, the rescaled range is calculated for the entire data set ($RSave_0=RS_0$). Then the rescaled range is calculated for the two halves of the data set, resulting in RS_0 and RS_1. The process continues by dividing each of the previous sections in half and calculating the rescaled range for each new section. The rescaled range values for each section are averaged. The subdivision stops when the region gets too small (at least 8 data points).

The Hurst parameter is estimated by calculating the average rescaled range over multiple regions of the data. For a given set of observations, $\{X_1, X_2, \ldots, X_n\}$ with sample mean $\mu=E\{X_i\}$ is defined a sequence of adjusted partial sums:

$$W_j=(X_1+ X_2+\ldots+X_j) - j\overline{X}(n), \quad j=1, 2, 3,\ldots, n \qquad (1)$$

Where $\overline{X}(n)$ is the arithmetic mean of the first n observations.
The range $R(n)$ is defined by:

$$R(n) = \max(0, W_1, W_2, \ldots,W_n) - \min(0, W_1, W_2, \ldots,W_n) \qquad (2)$$

The sample standard deviation $S(n)$ of the observations $X_1, X_2, \ldots, X_n$ is defined by:

$$S(n) = \sqrt{E(X_i-\mu)^2} \qquad (3)$$

The Hurst parameter is presented by the rescaled adjusted range:

$$R/S \text{ statistics} = R(n)/S(n) \qquad (4)$$

The expected value of $R(n)/S(n)$ asymptotically satisfies the power law relation:

$$E\left[\frac{R(n)}{S(n)}\right] \rightarrow cn^H, \quad \text{as } n\rightarrow\infty \qquad (5)$$

Where H > 0.5 is Hurst parameter and c>0 is a finite constant.

<table>
<tr><td colspan="16">RS₀</td><td>RSave₀</td></tr>
<tr><td colspan="8">RS₀</td><td colspan="8">RS₁</td><td>RSave₁</td></tr>
<tr><td colspan="4">RS₀</td><td colspan="4">RS₁</td><td colspan="4">RS₂</td><td colspan="4">RS₃</td><td>RSave₂</td></tr>
<tr><td colspan="2">RS₀</td><td colspan="2">RS₁</td><td colspan="2">RS₂</td><td colspan="2">RS₃</td><td colspan="2">RS₄</td><td colspan="2">RS₅</td><td colspan="2">RS₆</td><td colspan="2">RS₇</td><td>RSave₃</td></tr>
<tr><td>RS₀</td><td>RS₁</td><td>RS₂</td><td>RS₃</td><td>RS₄</td><td>RS₅</td><td>RS₆</td><td>RS₇</td><td>RS₈</td><td>RS₉</td><td>RS₁₀</td><td>RS₁₁</td><td>RS₁₂</td><td>RS₁₃</td><td>RS₁₄</td><td>RS₁₅</td><td>RSave₄</td></tr>
</table>

Fig. 1 Estimating the Hurst parameter

For a short range-model, the expected value will be described by a power function with an exponent of 0.5.

$$E\left[\frac{R(n)}{S(n)}\right] \rightarrow dn^{0.5}, \quad \text{as } n\rightarrow\infty \qquad (6)$$

Where d is a finite constant. The difference between (5) and (6) is called the Hurst effect.

The gradient of a plot of log(R/S) against log (n) is the Hurst parameter.

2.2 Variance-time plot

The variance-time plot analysis is based on property of slowly decaying variance of self-similar processes undergoing aggregation. The m-averaged process $X^{(m)}=(X_1^{(m)}, X_2^{(m)},\ldots)$ of a discrete-time stationary parent process $X_1, X_2,\ldots$ as:

$$X_j^{(m)} = \frac{1}{m}\sum_{i=(j-1)m+1}^{jm} X_i \qquad j=1, 2,\ldots, N/m \qquad (7)$$

Where m and j are positive integers.
The variance is defined as:

$$Var[X^{(m)}]=1/(N/m)\sum_{j=1}^{N/m}(X_j^{(m)}-\overline{X})^2 \qquad (8)$$

The variances of the aggregated processes $X^{(m)}$(m=1, 2, 3,…) decrease linearly (for large m):

$$Var[X^{(m)}]=Var[X]/m^{ß} \qquad (9)$$

Where H=1-ß/2.

The variance-time plot is obtained by plotting log(Var($X^{(m)}$)) against log(m) and by fitting a sample least squares line through the resulting points in the plane, ignoring the small values for m. Values of the estimate *ß* of the asymptotic slope between –1 and 0 suggest self-similarity and an estimate for the degree of self-similarity is given by H=1- ß/2.

3. Simulation results

For estimating Hurst parameter is created generator of pseudo-random self-similar sequences, based on the fractional Gaussian noise. The generator is implemented in C++ language. It is used to generate over 100 sample sequences of 32 768 numbers starting from different random seeds for H = 0.50, 0.55, 0.60, 0.65, 0.70, 0.80, 0.85, 0.90 end 0.95.

The estimates of the Hurst parameter obtained from the R/S statistic and the variance-time plot have been used to analyze the accuracy of the generator. The relative inaccuracy ΔH is calculating using the formula:

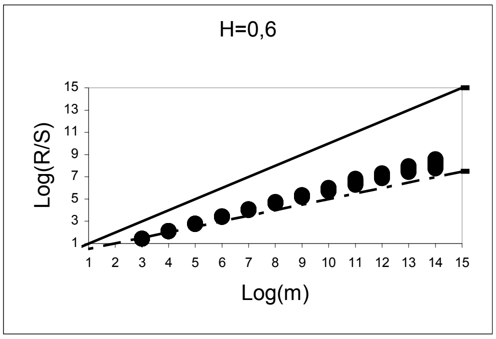
$$\Delta H=\frac{\hat{H}-H}{H}*100\% \qquad (10)$$

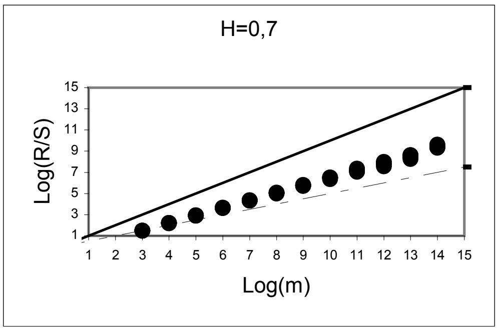
Where *H* is the required value of the Hurst parameter and $\hat{H}$ is the empirical mean value over a number of independently generated sequences. The presented numerical results are all averaged over 100 sequences.

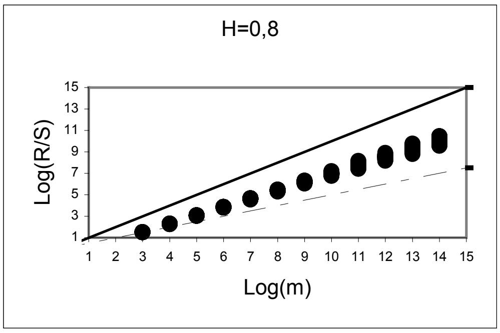
- The plots of R/S statistic indicate the self-similar nature of the generated sequences. The results for H=0.6, 0.7, 0.8 and 0.9 are shown in Figure 2. An estimate of H is given by $\hat{H}=\hat{\beta}$, where $\hat{\beta}$ is asymptotic slope of the R/S statistic plot. The relative inaccuracy, of the estimated Hurst parameter, obtained from the R/S statistic plot is given in Table 1. This method of analysis of H does not show that this generator has a persistently positive or negative bias of $\hat{H}$, as the variance-time plot did.
- The variance-time plot also indicates the self-similar nature of the generated sequences. The results are shown in Figure 3. The analysis of variances of the aggregate processes is accomplished at levels of m = 8, 16, 32, 64, 128, …, 1024. An estimate of *H* is given by $\hat{H}=1-\hat{\beta}/2$, where $\hat{\beta}$ is slope. In Table 1 are give the relative inaccuracy of the estimated Hurst parameters. The method shows high quality in the sense of the accuracy of H. The relative inaccuracy increasing with the increase in H. The results suggest that the output sequences have a negatively biased $\hat{H}$.

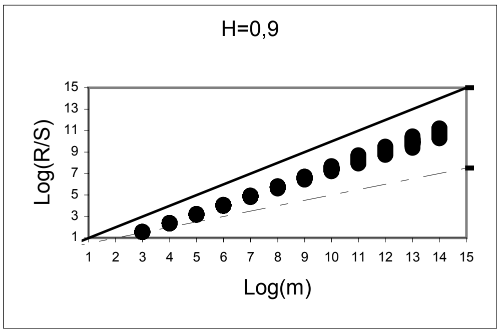
Table 1. Comparison of the two methods for estimating the Hurst parameter

Target H	R/S statistics				Variance-time plot			
	Mean	Relative inaccuracy %	Standard deviation	Mean-squared error	Mean	Relative inaccuracy %	Standard deviation	Mean-squared error
0.50	0.511	+2.20	0.042	0.205	0.496	-0.80	0.041	0.202
0.55	0.558	+1.45	0.050	0.224	0.545	-0.91	0.052	0.228
0.60	0.605	+0.83	0.059	0.243	0.593	-1.17	0.058	0.240
0.65	0.650	+0.00	0.069	0.263	0.640	-1.54	0.067	0.259
0.70	0.692	-1.14	0.078	0.279	0.689	-1.57	0.077	0.277
0.75	0.732	-3.63	0.088	0.297	0.738	-1.60	0.087	0.294
0.80	0.769	-3.88	0.097	0.311	0.773	-3.44	0.090	0.302
0.85	0.805	-5.29	0.107	0.327	0.805	-5.29	0.092	0.303
0.90	0.846	-6.00	0.117	0.342	0.853	-5.44	0.095	0.305
0.95	0.876	-7.79	0.126	0.355	0.875	-7.90	0.098	0.307









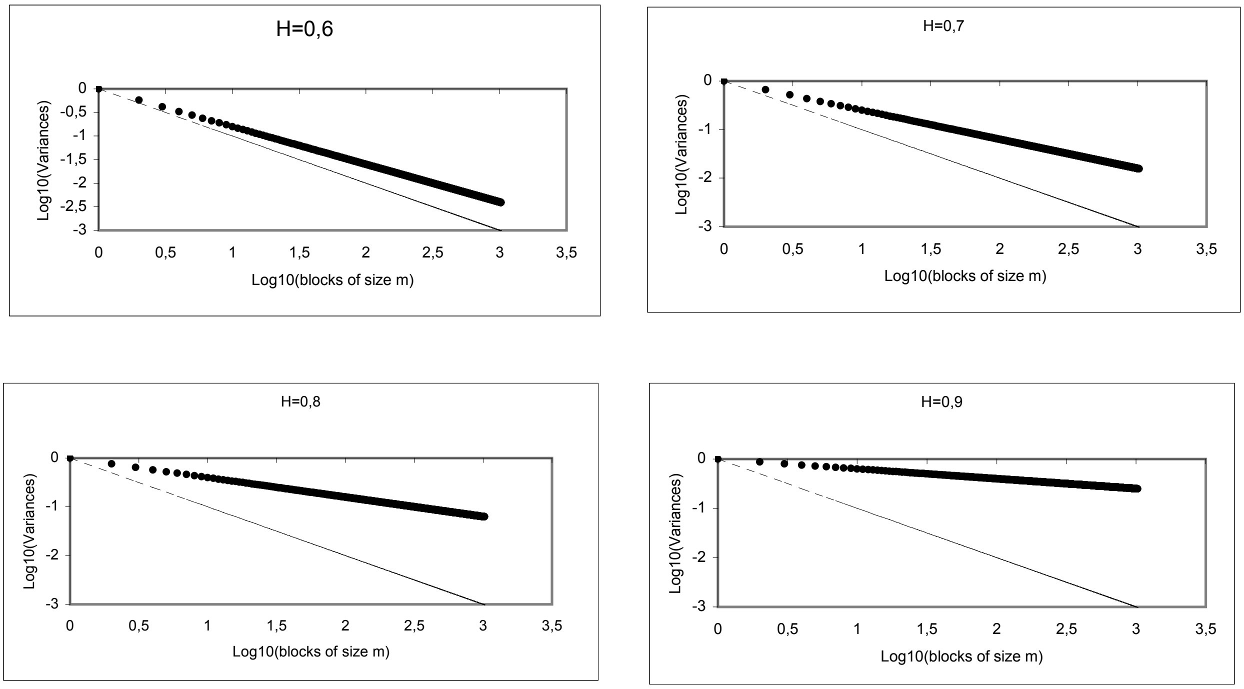


Figure 3. Variance-time plot for FGN method (H = 0.6, 0.7, 0.8, 0.9)

Table 1 presents the results of mean, standard deviation, and mean squared error for each value of *H,* estimated using each method. The parameter's dimensions are a little higher for the *R/S* statistics method than the variance-time plot. The results show that the generator produces self-similar *FGN* sequences with high degree of accuracy.

Conclusions

In this paper are used two graphical techniques for estimating Hurst parameter of pseudo-random self-similar sequences, based on the FGN method. The analysis show that the FGN method always produces self-similar sequences, with relative inaccuracy of the resulted H below 8%, if 0.5≤ H ≤0.95.

The analysis of generating FGN data shows that this generator should be recommended for practical simulation studies of high-speed telecommunication networks (Ethernet, ATM, VBR video traffic, Web traffic, Telnet and FTP), since it is very accurate.

This paper is a part of doctor's work developed under scientific research program by the research team of Central Laboratory of Mechatronics and Instrumentation at Bulgarian Academy of Sciences.

References

[1] Beran, J., R. Sherman. "Long-range dependence in variable-bit-rate video traffic". IEEE Transactions on Communications, vol.43, 1995.

[2] Leland, W., M.Taqqu. "On the self-similar nature of Ethernet traffic". IEEE/ACM Transactions on Networking, vol.47, 1993.

[3]Taqqu, M., V.Teverovsky. "Is network traffic self-similar or multifractal?". Fractals, vol.5,1997.

[4]Willinger, W., M.S.Taqqu. "Self-similarity through high-variability:Statistical analysis of Ethernet LAN traffic at the source level". IEEE/ACM Transactions on Networking, vol.5, 1993.

[5]Господинов, М., Е.Господинова. "Самоподобно поведение на Интернет трафика". Национална конференция "Роботика и мехатроника '2004", Дряново, 2004.

ABOUT THE AUTHORS

Assoc. Prof. Mitko Gospodinov, PhD, Central Laboratory of Mechatronics and Instrumentation at Bulgarian Academy of Sciences, Fhone: +359 887 426 490, E-mail: mitgo@abv.bg

Assistant Prof. Evgeniya Peneva Gospodinova, Central Laboratory of Mechatronics and Instrumentation at Bulgarian Academy of Sciences, Phone: +359 886 626 911, E-mail: jenigospodinova@abv.bg