Wavelet-based analysis of simulated network traffic

Mitko Gospodinov, Evgeniya Gospodinova

Abstract: *In the paper is applied wavelet-based Hurst parameter estimator for analysis of simulated network traffic, based on the fractional Gaussian noise. It was made a comparative analysis between the results obtained of wavelet-based estimator and wide using estimators as R/S statistic, variance-time plot and periodogram. The Hurst parameter obtained of wavelet-based estimator has the least value of relative inaccuracy compared to the other estimators. The simulated network traffic has a high degree of accuracy and this generator can be used in practical computer simulation studies.*
Key words: *Long-Range Dependence, Hurst parameter, wavelet-based estimator.*

Introduction

The high-speed network traffic investigations performed recently demonstrated that this is a Long-Range Dependence (*LRD*) process [2], [9]. The *LRD* process exhibits similar behavior observation on various time scales. This process is characterized by *Hurst* parameter (*H*) which measures the degree of *LRD.* The value of *H* is between *0.5* and *1*. The correct and efficient estimation of *H* is important for the statistical analysis of the process. There are several Hurst parameter estimators, such as R/S plot, variance-time plot, periodogram and wavelet-based [5]. The wavelet transform with their natural scale invariance and low computational cost is suitable for analyzing of *LRD* process [3]. In this paper is used wavelet-based *H* estimator for analysis of simulated network traffic, based on the fractional Gaussian noise (FGN) process [4], [7].

1. Wavelet-based estimator of the Hurst parameter

1.1. Wavelet transform

The wavelet transform is mathematical tool for representing signals as sum of "small waves". It is a better substitute of the Fourier transform. The Fourier transform is used to transform a signal from the time domain to the frequency domain. The signal is transformed into a sum of sinusoid of different frequencies. The Fourier transform cannot present information about the time. The wavelet transform is capable of providing the time and frequency information of a signal simultaneously. This transform is analyzed non-stationary signals with sudden peaks, without losing information on low or high frequency and time domain [2].

The basic algorithm of the wavelet transform is shown in Figure 1. This simple filter algorithm performs a one-dimensional one-scale wavelet transform on any one-dimensional input sequence. It uses the pyramidal algorithm shown in Figure 2. This algorithm is based on two filters (G_0 and G_1) that are derived from the scaling function and mother wavelet chosen for the transformation. G_1 is high-pass wavelet filter and G_0 is the complementary low-pass wavelet filter. The outputs are the low-pass residue for the G_0 filter branch, represented by *Approx,* and the high-pass sub-band for the G_1, branch, represented by *Detail* [1].

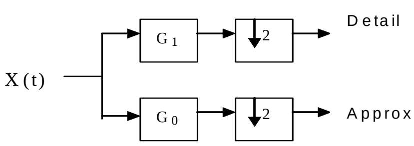

Figure 1. Basic algorithm for the wavelet transform

The computational cost of the pyramidal algorithm is O(n), which is lower than the cost of the FFT algorithm O(nlogn) [6].

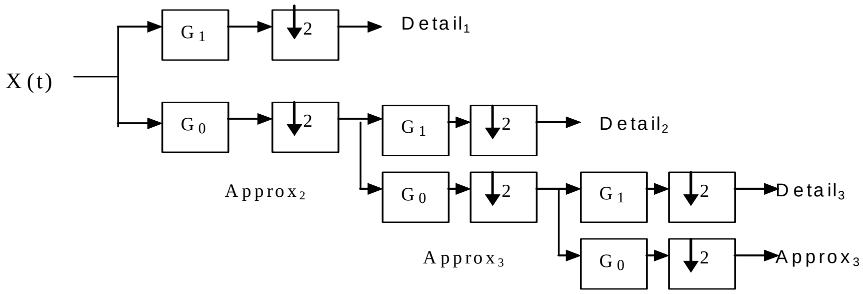

Figure 2. Recursive pyramidal algorithm for the multi-scale wavelet transform.

1.2. Multi-Resolution Analysis

The wavelet transform is based on the concept of Multi-Resolution Analysis (MRA). MRA considers the information at different resolutions or scales. Fourier transforms are constant resolution based, because they involve single forward/reverse transforms that convert data to/from a different representation.

The MRA consists of collection of nested subspaces {V_i, i∈ Z}, satisfying the following properties [1]:

1. $\bigcap V_i = \{0\}$, $\bigcup V_i$ is dense in Hilbert space $L^2(R)$;

2. $V_i \subset V_{i-1}$;

3. $x(t) \in V_i \leftrightarrow x(2^i t) \in V_0$;

4. The function $\phi_0(t)$ in vector V_0 is called the scaling function and the collection $\{\phi_0(t-j), j \in Z\}$ is an orthonormal basis for V_0.

The input signal is represented in terms of dilated versions of a prototype of high-pass wavelet function ($\psi_{i,j}$) and shifted version of a low-pass scaling function ($\phi_{i,j}$), based on the scaling function (ϕ_0) and the mother wavelet basis function (ψ_0). The relationship between these function are:

$$\phi_{i,j}(t) = 2^{-i/2}\phi_0(2^{-i}t - j), \quad j \in Z \tag{1}$$

$$\psi_{i,j}(t) = 2^{-j/2}\psi_0(2^{-j}t - j), \quad j \in Z \tag{2}$$

The approximation information of sequence *x* is given by:

$$approx_i(t) = \sum_j a_x(i,j)\phi_{i,j}(t) \tag{3}$$

Where the coefficient $a_x(i,j)$ is given by calculating the inner product of *x:*

$$a_x(i,j) = \langle x, \phi_{i,j} \rangle \tag{4}$$

The detail information (*detail$_i$*) of sequence *x* is given by:

$$\det ail_i(t) = \sum_j d_x(i,j)\psi_{i,j}(t) \tag{5}$$

Where the coefficient $d_x(i,j)$ is given by calculating the inner product of *x:*

$$d_x(i,j) = \langle x, \psi_{i,j} \rangle \tag{6}$$

MRA represent the information about sequence *x* as a collection of details and a low-resolution approximation:

$$x(t) = approx_N(t) + \sum_{i=1}^{N} \det ail_i(t) \tag{7}$$

$$= \sum_j a_x(N,j)\phi_{N,j}(t) + \sum_{i=1}^{N}\sum_j d_x(i,j)\psi_{i,j}(t).$$

The function ϕ_0 produces an approximation of signal *x* and it must be a low-pass filter. The mother wavelet function ψ_0 must be high-pass filter, and it performs a differential operation on the input signal to produce the detail version.

1.3. Wavelet-based Hurst parameter estimator

The wavelet-based Hurst parameter estimator is based on a spectral estimator obtained by performing a time average of the wavelet *detail* coefficients $|d_x(i,j)|^2$ at a given scale [8]:

$$S_x = \frac{1}{n_i}\sum_j |d_x(i,j)|^2 \tag{8}$$

Where n_i is the number of wavelet coefficients at scale *i*, i.e., $n_i=2^{-i}n$, where *n* is number of data points.

The estimator first performs Discrete Wavelet Transform (DWT) on the input signal, employing wavelets from the Daubechies family. In the paper is used wavelet Daubechies4, which has 4 vanishing moments. After computing the DWT, the estimator calculates the estimates of $log_2E[d(i,j)]^2$ and variance of these estimates and performs a linear regression. The estimator estimates the slope β by performing linear regression of $log_2(S)$ for range $[i_1,i_2]$. The *H* is calculated by formula: $H=0.5(1+\beta), 0<\beta<1.$ The linear relationship between $log_2(S)$ and scale level *i* over a range $[i_1,i_2]$ indicates the presence of a LRD behavior.

The estimator is based on the following idealizations:

1. The *process X(t)* and its wavelet coefficients are Gaussian;
2. For fixed *i, d(i,j)* are independent, identically distributed variables;
3. The processes $d(i_1,j)$ and $d(i_2,j)$ for $i_1 \neq i_2$ are independent.

2. Simulation results

In this paper is used the wavelet-based Hurst parameter estimator for analyzing the accuracy of simulated network traffic, based on fractional Gaussian noise process. The simulations are implemented in C++ language. For each of Hurst parameter *= 0.6, 07, 08* and *0.9* are generated *100* sample sequences of 2^{15} *(32768)* numbers, starting from different random seed.

Table 1 shows the relative inaccuracy of mean values of Hurst parameter for different sample size of self-similar FGN process. The relative inaccuracy ΔH is calculated using the formula: $\Delta H=\frac{\hat{H}-H}{H}*100\%$, where *H* is the input value and $\hat{H}$ is an empirical mean value. The relative inaccuracy decreases with the increase in sample size and it is minimal at sample size 2^{15}. This sequence length is chosen for simulating network traffic.

Table 1 Relative inaccuracy of mean values of ΔH (%) for different sample size

Sample size	*Relative inaccuracy ΔH (%)*			
	H=0.6	*H=0.7*	*H=0.8*	*H=0.9*
2^{12} (4096)	0.2844	0.4605	0.1851	-1.2552
2^{13} (8192)	0.2807	0.4263	0.0771	-1.2523
2^{14} (16384)	0.2009	0.4190	0.0661	-1.3128
2^{15} (32768)	0.1770	0.4062	0.0268	-1.0639

Table 2 shows the relative inaccuracy when the mean values of $\hat{H}$ are obtained for 100 self-similar sequences.

Table 3 shows variances of $\hat{H}$ values for different scales and values of *H*. The variances gradually decrease as the values of i_1 increase. The results shows that the variance of $\hat{H}$ is the least biased at scales i_1= 8 and 9. Therefore, the scale i_1=9 is chosen for analyzing of simulated network traffic.

Table 2 Relative inaccuracy of mean values of ΔH (%) for different scales

Scale	*Relative inaccuracy ΔH (%)*			
(i_1,i_2)	*H=0.6*	*H=0.7*	*H=0.8*	*H=0.9*
(5,12)	-1.7936	-3.0274	-4.7117	-7.0652
(6,12)	-1.1259	-2.0074	-3.4078	-5.5752
(7,12)	-0.7429	-1.2483	-2.3118	-4.2275
(8,12)	-0.4802	-0.5752	-1.2456	-2.8429
(9,12)	+0.1770	+0.4560	+0.5751	-1.0639

Table 3 Variances of $\hat{H}$ values for different scales and *H* values

Scale	*Estimated variances*			
(j_1,j_2)	*H=0.6*	*H=0.7*	*H=0.8*	*H=0.9*
(5,12)	0.0034	0.0098	0.0188	0.0290
(6,12)	0.0029	0.0085	0.0164	0.0254
(7,12)	0.0023	0.0069	0.0135	0.0210
(8,12)	0.0009	0.0040	0.0087	0.0145
(9,12)	0.0007	0.0028	0.0062	0.0103

Figure 3 shows the linear relationship between *$log_2(S)$* and scale *i* over a range from *5* to *12.* This relationship indicates the presence of long-range dependence behavior of simulated network traffic. The estimator $\hat{H}$ for the Hurst parameter is determined by performing a linear regression of *$log_2(S)$* on scale level *i* in the range from *5* to *12*.

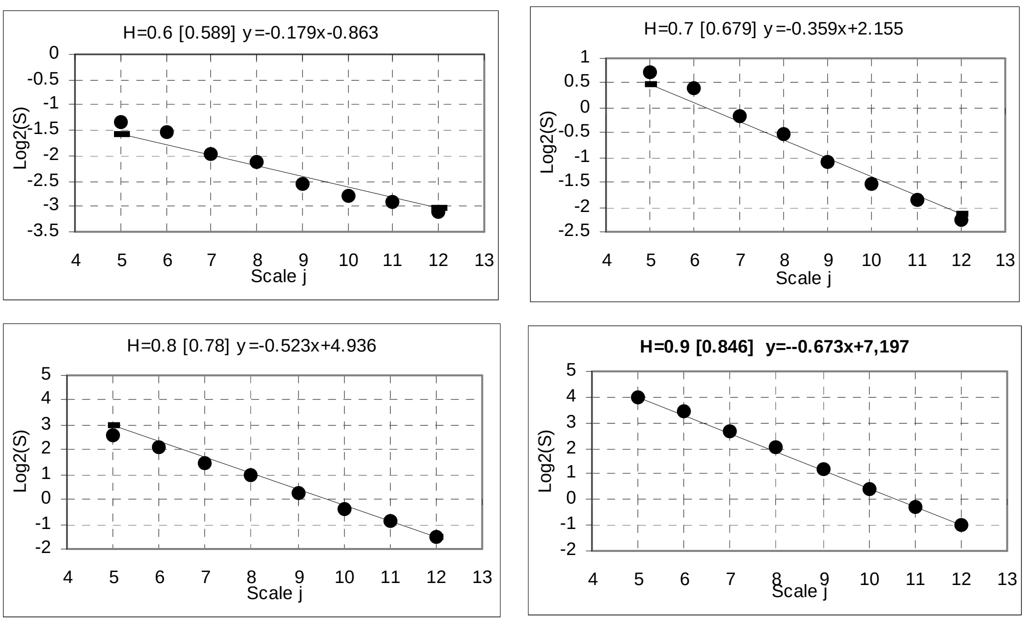

Figure 3. The regression curves for wavelet-based H estimates of self-similar *FGN* process, with *H=0.6, 0.7, 0.8* and *0.9*, for scale *(j_1,j_2)=(5,12).*

Figure 4 shows relative inaccuracy (ΔH) of mean values of the estimated *H* using different Hurst parameter estimation techniques: periodogram, variance-time plot, R/S statistics and wavelet-based estimator for different values of *H=0.6, 0.7, 0.8* and *0.9* of the simulated network traffic.

Basic on the results could be made the following conclusions:

- The periodogramm plots have decreasing slopes when *H* increase. The negative slopes of all plots for H are evidence of self-similarity. The estimated Hurst parameters have positively biased $\hat{H}$ values. For *0.6≤H<0.65* the confidence intervals of estimated Hurst parameter contain the approximately exact values.
- The variance-time estimator produces negatively $\hat{H}$ values when *H* increase with $|\Delta H|$<5%.

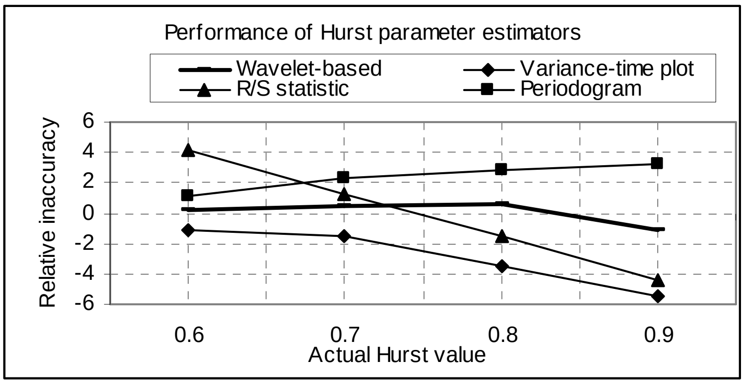

Figure 4. Bias performance of Hurst parameter estimators for self-similar fractal Gaussian noise process

- The R/S plot produces positively biases for H<0.75 and negatively biases for H>0.75 with $|\Delta H|$<4.12%.
- The results for the wavelet-based estimator show that for all input *H* values, confidence intervals are within the assumed theoretical values. For *H=0.6* the wavelet-based estimator is the most accurate. For *0.6≤H<0.9* the relative inaccuracies are less than *0.58%.*

Conclusions

The correct and efficient estimation of the Hurst parameter of high-speed network traffic is important for traffic analysis. The application of wavelet-based Hurst parameter estimator is implemented for analysis of simulated network traffic, based on the fractional Gaussian noise. Comparative analysis is made between the results obtained of wavelet-based estimator and wide using estimators as R/S statistic, variance-time plot and periodogramm. The conclusions based on the results are:

1. The *Hurst* parameter obtained of wavelet-based estimator has the least value of relative inaccuracy compared to the other estimators. For *H=0.6* the wavelet-based estimator is the most accurate. For *0.6 ≤ H < 0.9* the relative inaccuracies are less than *0.58%.*
2. The guarantees to simulate network traffic, based on the fractional Gaussian noise are with high degree of accuracy. This generator can be used in practical computer simulation studies, when long self-similar sequences of numbers are needed.

References

[1] Abry, P., D.Veitch. "Wavelet analysis of Long-Range-Dependent Traffic", IEEE Transactions on Information Theory, vol. 1998.

[2] Abry, P., D.Veitch. "A wavelet-based joint estimator of the parameters of long-range dependence". IEEE Transactions on Information Theory, vol. 45, 1999.

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

[4] Gospodinov, M., E. Gospodinova. "Generator of fractional Gaussian noise for modeling self-similar network traffic", CompSysTech'2005.

[5] Gospodinov, M., E. Gospodinova. "The graphical methods for estimating Hurst parameter of self-similar network traffic", CompSysTech'2005.

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

[7] Paxson, V. "Fast, Approximate Synthesis of Fractional Gaussian Noise for Generating self-similar network traffic", ACM SIGCOMM, 1997.

[8] Rincon, D., S.Sallent. "Characterizing fractal traffic with redundant wavelet-based transform". Submitted to IEEE Globecom, 2004.

[9] 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.

ABOUT THE AUTHOR

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