# Traffic characterization using wavelet-based techniques

David Boulliant, Parag Pruthi and Adrian Popescu

Department of Telecommunications and Mathematics
University of Karlskrona/Ronneby
371 79 Karlskrona, Sweden

## ABSTRACT

Study of long-range dependence (LRD) properties in real traffic has received an increasing attention in traffic analysis. A wavelet-based tool for the analysis of LRD is presented in this paper together with a semi-parametric estimator of the Hurst parameter. The estimator has been proved to be unbiased under very general conditions and efficient under Gaussian assumptions. Analysis of the Bellcore Ethernet traces as well as of some VBR video traces using the wavelet-based estimator is reported.

**Keywords:** long-range dependence, Hurst estimation, wavelets, stationarity

## 1. INTRODUCTION

Recent traffic measurement studies from a wide range of working packet networks have convincingly shown the presence of self-similar (long-range dependence) properties in both local and wide area traffic traces. LRD processes are characterized (in the case of finite variance) by self-similarity of aggregated summands, slowly decaying covariances, heavy-tailed distributions and a spectral density that tends to infinity for frequencies approaching zero. This discovery calls to question some of the basic assumptions made by most of the research in control, engineering and operations of broadband integrated systems. At the time being, there is mounting evidence that self-similarity is of fundamental importance for a number of teletraffic engineering problems such as traffic measurements and modeling, queueing behavior and buffer sizing, admission control, congestion control, etc.

These impacts have highlighted the need for precise and computationally feasible methods to estimate diverse LRD parameters. Especially real-time estimation of measured data traces and off-line analysis of enormous collected data sets call for accurate and effective estimation techniques.

A wavelet-based tool for the analysis of LRD is presented in this paper together with a semi-parametric estimator of the Hurst parameter $H$. The complexity of this estimator is rather low, it is mainly given by the computation of a discrete wavelet transform. The estimator has been proved to be unbiased under very general conditions and efficient under Gaussian assumptions. Excellent agreements with the theoretical results have been observed for both $H$ estimates as well as for the corresponding confidence intervals. Analysis of the Bellcore Ethernet traces as well as of some VBR video traces using the wavelet-based estimator is also reported.

## 2. ESTIMATION OF LONG-RANGE DEPENDENCE

Several methods are available today to estimate the LRD parameter and/or the intensity of LRD in time series. Many of them suffer however because of diverse limitations, e.g., sensitivity to the (assumed) underlying model process, non-stationarity, and deterministic trends.[1]

For instance, traditional aggregation-based estimators like R/S analysis and variance-time analysis may get seriously biased because of poor statistics available from a single realization of the LRD process.[2] On the other hand, asymptotically unbiased estimators derived from Gaussian Maximum Likehood Estimation MLE (Whittle-type estimators) show good statistical performance but they have the drawback of being parametric estimators that require parametrized model processes to be known in advance.[1] This poses large difficulties in terms of exact implementation for large data sets because of high computational complexity.[2] Furthermore, if the assumed spectral density model is not the correct one, then this may result in a biased estimation as well.

---

Further author information:
D.B. (correspondence): Email: dbo@itm.hk-r.se; Telephone: +46 455 78151; Fax: +46 455 78057
P.P.: Email: parag@niksun.com
A.P.: Email: adrian@itm.hk-r.se; Telephone: +46 455 78061; Fax: +46 455 78057

# 3. WAVELET-BASED ESTIMATION

## 3.1. Discrete Wavelet Transform

The wavelet estimator is based on the discrete wavelet transform (DWT), which has the advantages of both aggregation-based and MLE estimators and also avoids their drawbacks. The underlying concept of the DWT is the so-called Multi-Resolution Analysis (MRA), which consists of splitting a sequence $x(t)$ into a (low-pass) approximation and (high-pass) details. These are associated with the coefficients $a_x$ and $d_x$, respectively[3]

$$x(t) = \sum_k a_x(J,k)\phi_{J,k}(t) + \sum_{j=1}^{J}\sum_k d_x(j,k)\psi_{j,k}(t) \tag{1}$$

The parameters $\{\phi_{j,k}(t) = 2^{-j/2}\phi_0(2^{-j}t - k)\}$ and $\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)\}$ are the set of shifted and dilated functions of the scaling function $\phi_0$ and the wavelet $\psi_0$. The discrete wavelet transform consists of the collection of coefficients $\{d_x(j,k), j = 1, \ldots, J, k \in Z\}$ and $\{a_x(J,k), k \in Z\}$ as defined by the inner products

$$d_x(j,k) = \langle x, \psi_{j,k}\rangle \tag{2}$$

$$a_x(j,k) = \langle x, \phi_{j,k}\rangle \tag{3}$$

These coefficients are located on a dyadic grid. The focus is then placed on details which are described by the wavelet coefficients. When going from high resolution to lower resolution, the MRA gives rise to details at larger time scales. This can be interpreted in the frequency domain as band-pass filtering, going from high to low frequencies with constant relative bandwidth. On the other hand, spectral estimators (based on periodograms) may easily get strongly biased due to the fact that constant bandwidth mismatches the power-spectrum to be analyzed.[1] In contrast, the wavelet constant relative bandwidth manages to provide a perfect match.

## 3.2. Estimation of LRD with Wavelets

Consider applying the wavelet transform to a wide-sense stationary stochastic process that is of type LRD. The autocovariance function of such a process shows the well-known power-law behavior for large time lags $\tau$, and is dependent on the Hurst parameter $H \in [0.5, 1)$[3]

$$\gamma_x(\tau) \sim c_\gamma \tau^{-(2-2H)}, \tau \to \infty \tag{4}$$

The process has also a power spectrum that diverges when the frequency $\nu$ approaches zero

$$\Gamma(\nu) \sim c_f|\nu|^{-(2H-1)}, \nu \to 0 \tag{5}$$

where $c_\gamma$ and $c_f$ are constants.

Some of the most important statistics of the wavelet coefficients are the mean value

$$E[d_x(j,k)] = 0 \tag{6}$$

and the second order moment

$$E[|d_x(j,k)|^2] = \int \Gamma(2^{-j}\nu)|\Psi_0(\nu)|^2 d\nu = c_f|2^{-j}|^{(1-2H)} \int |\nu|^{(1-2H)}|\Psi_0(\nu)|^2 d\nu \tag{7}$$

The coefficient $|d_x(j,k)|^2$ measures the amount of energy in the analyzed signal about the time instant $2^j k$ and about the frequency $2^{-j}\nu_0$, where $\nu_0$ is a reference frequency dependent of $\Psi_0$. In practice, the following estimation is used[4]

$$E[|d_x(j,k)|^2] = \frac{2^j}{n_0}\sum_k |d_x(j,k)|^2 \tag{8}$$

where $n_0$ is the data length. This estimation corresponds in fact to the mean energy in the frequency band $j$, and the normalized sum of squares is generally an unbiased estimator of the second order moment. Furthermore, assuming that Eqs. (4) and (5) hold for all time lags and frequencies, an unbiased estimator $\hat{H}$ for the parameter $H$ could be designed from a simple linear regression using

$$log_2(\frac{2^j}{n_0}\sum_k |d_x(j,k)|^2) = (2\hat{H}-1)j + \hat{c} \tag{9}$$

where $\hat{c}$ estimates $log_2(c_f \int |\nu|^{(1-2H)}|\Psi_0(\nu)|^2 d\nu)$. This is valid if the integral

$$\int |\nu|^{(1-2H)}|\Psi_0(\nu)|^2 d\nu \tag{10}$$

converges.

### 3.3. Bias and Efficiency

The wavelet estimator is non-biased under the conditions that Eq. (5) holds for all frequencies and the integral (10) converges. The first condition is easily satisfied because we have the freedom to select the frequency range over which the power-law holds. The second condition is also satisfied in the case we consider a wavelet with a large number $N$ of vanishing moments.[3] By this, the divergent behavior of the power-law spectra close to the frequency origin is balanced. The relation between number of vanishing moments and the Hurst parameter is given by

$$N > H - 1 \tag{11}$$

The wavelet estimator has been shown to be unbiased under very general conditions. It provides an analysis of very general data sets that are more independent of the correlation structure of the data than for instance the Whittle estimator. There are no restrictions in selecting the scales over which the power-law regime holds, which reduces the difficult task of selecting cutoff scale/frequency. The wavelet coefficients are almost uncorrelated, leading to an asymptotic normal distribution of the coefficients at a given scale from which a confidence interval can be deduced (under Gaussian assumptions). Accordingly, estimations based on weighted (least-square) linear regression can be performed.[4] Actually, there are solutions reported even without the asymptotic simplification, with the consequence of more effective weights and smaller confidence intervals.[5]

Finally, another important advantage of the wavelet-based estimators is given by the possibility to investigate if the process has multi-fractal or mono-fractal features, i.e., if the Hurst parameter alone can describe the correlation structure of the process.

## 4. STATIONARITY

One of the typical features of stationary LRD processes is related to the fact that local trends and cycles seem to appear in the process evolution, which may disappear after some time.[5] This evolution can be misleading in the process of estimation, and makes it difficult to distinguish a stationary process with LRD from a non-stationary one. This also raises the problem of the need to test for stationarity assumptions in the process, which may require additional knowledge of the underlying functionality of the process. It may be also possible to handle this requirement by the selection of relevant time scales where the process is stationary, but the relevant time scales are not known apriori.

The wavelet-based technique provides a simple method of testing the stationarity assumptions. The fact that the wavelet coefficients are almost uncorrelated opens for the possibility to partition the sequence, and to perform estimations on each subsequence. Every estimation will be almost independent, and statistics such as mean and variance are meaningful. These statistics can finally be compared with the $H$ estimation on the entire sequence. For instance, Fig. 1 shows an example of variation of $H$ estimation for a trace of fractional Gaussian noise (fGn),

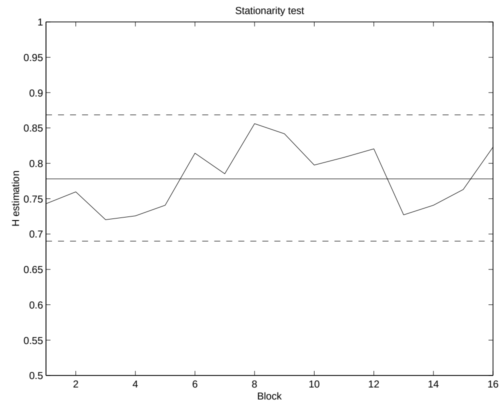


**Figure 1.** A trace of fGn ($n_0 = 16384$ and $H = 0.8$) is partitioned into 16 subsequences with different $H$ estimations. All partial $H$ estimations are found to be placed within the confidence interval (dotted lines), which is centered around the $H$ estimation over the entire trace (solid line). The Haar wavelet was used.

partitioned into 16 blocks. The confidence interval (dotted lines) is centered around the estimation of the entire trace (solid line), and all partial estimations are found to be placed inbetween. This trace appears to be a stationary one.

When dealing with the problem of deterministic trends, the wavelet-based analysis has proven to provide a simple solution. Every wavelet has a property called *vanishing moments*, which is defined as the degree $N$ of the polynomial for which the inner (scalar) product with the wavelet is zero. Every polynomial trend, of degree $P$, superimposed on the process can therefore be eliminated, conditioned on enough number of vanishing moments[4]

$$N \geq P + 1 \tag{12}$$

Furthermore, corruption trends consisting of periodic oscillations may also be reduced by increasing the number of vanishing moments. The effect of the trend can in this case be more concentrated at a single scale (frequency), enabling so inclusion of a specific scale in the estimation. The wavelet tends to behave like an ideal band-pass filter.[5]

In practice, for removing the effects of polynomial trends, repeated estimations must be made, with increasing number of vanishing moments. After a given order of vanishing moments, the estimation results will essentially be equal. Furthermore, periodic trends can be detected by graphical inspection, and searching for diversions from the regression line at single scales.

The stationarity of the process is tested by partitioning the data set into small sets, for which $H$ estimations are performed as if they were independent. This is a test motivated by the quasi-decorrelation of the wavelet coefficients, as described in.[3] For every segment an $H$ estimation, together with a common confidence interval, are obtained. Should the process be stationary, the mean of the $H$ estimations should not differ in any larger extend from the $H$ estimate taken over the entire data set.

## 5. IMPLEMENTATION OF THE ESTIMATOR

A wavelet-based tool for the analysis of LRD has been implemented to allow for semi-parametric estimation of $H$ parameter for stationary data and data with stationary increments.[4] The complexity of this estimator is rather low, it is mainly given by the calculation of a discrete wavelet transform (DWT). This is done in $O(n)$ operations (less than for Fast Fourier Transform FFT) and using a fast pyramidal filter-bank based algorithm followed by a time-averaging unit. The wavelet coefficients are (recursively) computed by means of cascade filters operating in a dyadic grid. Starting from an initial sequence (with a highest resolution), successive sequences of approximations (aggregation) and details at lower resolution are computed. Furthermore, since the wavelet is associated with a discrete filter, some considerations must be given to the fact that the data set is finite. Boundary effects may arise because of the mismatch in time domain between the filter and the sequence. These effects may lead to some corrupted wavelet coefficients, which are located at the edges. This effect is directly linked to $N$, since a large $N$ results in a filter with large number of coefficients.

The solution to this problem is simple, we simply discard the corrupted wavelet coefficients from the data set, and accordingly the boundary effects are removed. A drawback however is given by the reduced number of usable wavelet coefficients in the linear regression. On the other hand, a larger $N$ reduces the variance of the estimated $H$. A good practical compromise seems to be[5]

$$N \simeq H + 1 \tag{13}$$

Especially in the case of large data sets, wavelet coefficients can be discarded without loss of accuracy.

This method is also valid for non-stationary processes with stationary increments, like fractional Brownian motion (fBm).[6] In this case the Fourier transform of the autocovariance function becomes

$$\Gamma(\nu) \sim |\nu|^{-(2H+1)}, \nu \to 0. \tag{14}$$

When estimating the $H$ parameter of fBm with Whittle's technique one must first compute the increments. On the contrary, with the wavelet estimator one can work directly with the process itself. In this case, Eq. (11) becomes

$$N > H \tag{15}$$

Furthermore, slighty different results may be obtained when computing with the discrete wavelet transform, which depends on where the filter starts acting on the sequence. The reason for this is that the wavelet coefficients constitute a subsample of the original sequence, located on a dyadic grid. Due to this, $H$ estimates may slighty differ in specific cases, the statistical estimates however are the same.

The advanced filter-bank algorithm has significant advantages in terms of memory usage.[4] The sequence can be partitioned into smaller blocks of treatable sizes, and separate computations can be carried without reduction in the range of scale analyzed. High sampling rates may be used to accurately capture fine details of the sequence. The results are finally "combined" to obtain the final estimate.

The estimator has been validated using traces of fractional Brownian motion fBm (generated with the Random Midpoint Displacement RMD method) and also fractional Gaussian noise fGn.[4] Excellent agreements with the theoretical results have been obtained for $H$ estimates as well as for the corresponding confidence intervals (Tables 1 and 2). Influences from deterministic trends (polynomial and periodical) have been eliminated. The estimator is unbiased and of minimum or close to minimum variance for the scale parameter.

**Table 1.** Performance statistics for the wavelet estimator. A fGn trace of 4096 samples is generated 100 repeated times per trial. The statistics show very good agreement with theoretical predictions (parenthesis). The estimations were performed through scales (2, 7) using weighted and standard linear regression. The Haar wavelet was used ($N = 1$).

| $\hat{H}_w$ | $\hat{\sigma}_w$ | $\hat{H}_s$ | $\hat{\sigma}_s$ |
|---|---|---|---|
| 0.497(0.5) | 0.029(0.029) | 0.498(0.5) | 0.019(0.019) |
| 0.593(0.6) | 0.031(0.029) | 0.597(0.6) | 0.019(0.019) |
| 0.699(0.7) | 0.027(0.029) | 0.700(0.7) | 0.017(0.019) |
| 0.788(0.8) | 0.031(0.029) | 0.791(0.8) | 0.021(0.019) |
| 0.895(0.9) | 0.027(0.029) | 0.900(0.9) | 0.021(0.019) |

**Table 2.** $H$ estimations using Daubechies wavelet with $N = 2$ and 3, for trend elimination supperimposed on a fGn trace 40 repetitive times with $H = 0.8$ and $n_0 = 4096$. As being expected, $N = 2$ is not enough to eliminate a quadratic trend.

| Trend | $N = 2$ | $N = 3$ |
|---|---|---|
| $a \cdot t$ | 0.776 | 0.801 |
| $a \cdot t^2$ | 1.824 | 0.806 |
| $a \cdot \sin(2\pi \frac{4}{n_0} t)$ | 0.774 | 0.795 |
| $a \cdot t^{-\frac{1}{4}}$ | 0.773 | 0.798 |

# 6. MEASUREMENTS

## 6.1. Ethernet Data

Estimations of some of the well-known Bellcore Ethernet traces[7] (pAug89) have been done. The traces consist of one million timestamps in seconds (and fractions thereof) and frame sizes in bytes.[8] The focus of our estimations is laid on the discrete version of the arrival process and the corresponding point-process, using a sampling interval of 6 ms. As shown in Figs. 2 and 3, both processes are LRD with almost similar $H$ values (0.830 and 0.782 $\pm 0.011$ when $N = 2$). The estimations show no relevant dependence of $N$, except for a possible linear trend in the arrival process, indicating lack of smoothing trends (Table 3).

The stationary condition is also tested (Fig. 4). The partial estimations are centered around the $H$ estimation for the entire trace, and also within the (asymptotic) confidence interval. Hence the trace could be treated as stationary. The results are well matching similar estimations reported in the literature.[7]

**Table 3.** $H$ estimations using Daubechies wavelet with $N = 1, 2$ and 3, for trend elimination. In the case when $N = 3$, the number of available scales is less due to the boundary effect discussed above. The estimations were performed through scales (6,16) and (6,14), and using weighted linear regression.

| Process | $N = 1(6, 16)$ | $N = 2(6, 16)$ | $N = 3(6, 14)$ |
|---|---|---|---|
| *Arrival* | $1.02 \pm 0.011$ | $0.782 \pm 0.011$ | $0.794 \pm 0.012$ |
| *Point* | $0.826 \pm 0.011$ | $0.830 \pm 0.011$ | $0.838 \pm 0.012$ |

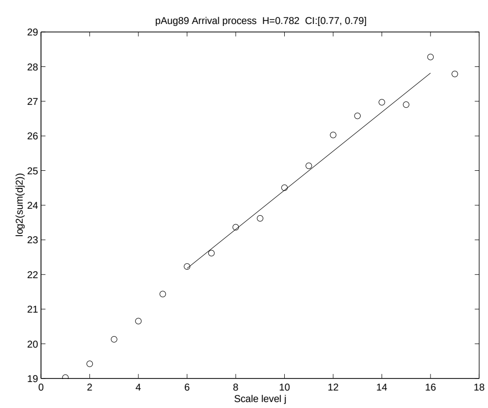


**Figure 2.** $H$ estimation of Bellcore traces, namely the discrete arrival process with $H = 0.782 \pm 0.011$. A clear power-law behavior is observed over almost all scales. A Daubechies wavelet was used with $N = 2$.

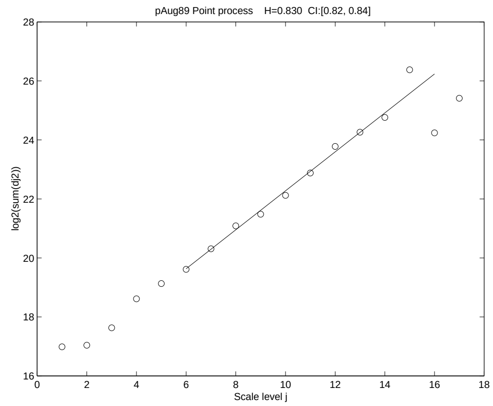


**Figure 3.** $H$ estimation of Bellcore traces, namely the discrete point process with $H = 0.830 \pm 0.011$. A clear power-law behavior is observed over almost all scales, starting at approximatly 20 ms. A Daubechies wavelet was used with $N = 2$.

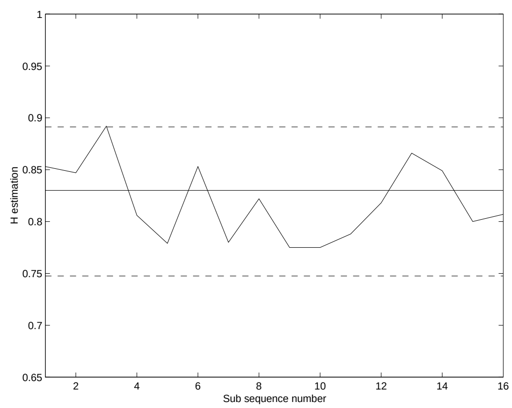


**Figure 4.** Repeated $H$ estimations on 16 subsequences. The confidence interval (dotted lines) is found to be centered around the $H$ estimation for the entire trace (solid line), and all separate H estimations are found within. A Daubeches was used with $N = 2$.

## 6.2. VBR Video Data

Estimations of the LRD behavior have been done for a VBR video trace recorded from the movie "Star Wars" and coded using an MPEG-1 encoder.[9] The trace is available over the Internet[10] as a data file that consists of 171,000 integers representing the number of bits per video frame. The movie length is of about 2 hours.

The estimations show that the process is LRD with $H$ values that are similar to those reported in the litterature, i.e., approximately $H = 0.91$ with a 95% confidence interval of 0.05 (Figs. 5 to 7).

Furthermore, the trace can be considered as being stationary according to Fig. 8. However, some of the partial estimations are larger than one ($H \geq 1$). Although this indicates the presence of non-stationarities, this could also be attributed partly to extremely strong short-term correlations. Previous studies of VBR video traffic have shown the presence of significant spectral power in the high-frequency range.[9] According to Eq. 5, a power spectrum with $H \geq 1$ can be associated with a non-stationary process. However, further investigation is required here to determine whether the resulting observations can in fact be attributed solely to non-stationarity or can be partly attributed to estimator sensitivity in the presence of strong short-lag correlations. Similar observations are also made in Figs. 5 to 7, where the asymptotic power-law behavior is observed from scale 8 and beyond. On a shorter trace, the power-law behavior is not observed, and the trace can be interpreted as non-stationary.

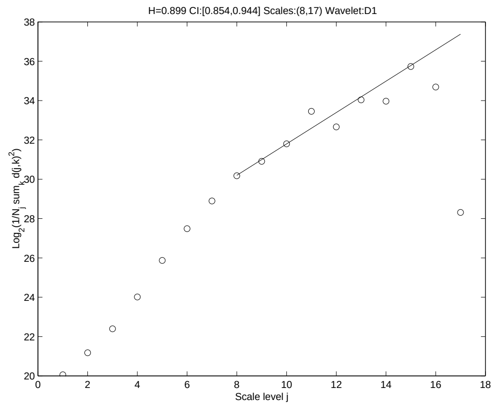


**Figure 5.** Estimation of $H$ and a 95%-confidence interval on the *Starwars* trace, using the Haar wavelet and weighted linear regression. $H = 0.899$ CI:[0.854, 0.944]

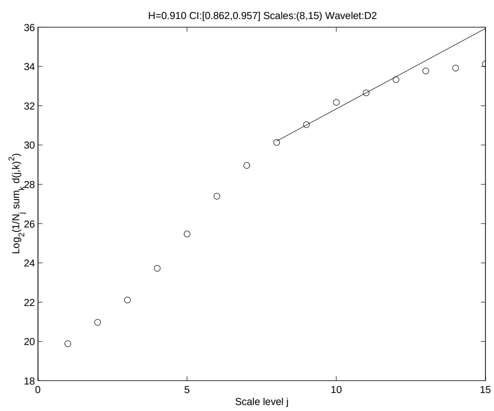


**Figure 6.** Estimation of $H$ and a 95%-confidence interval on the *Starwars* trace, using the Daubechies 2 wavelet and weighted linear regression. $H = 0.910$ CI:[0.862, 0.957]

## 7. CONCLUSIONS

In this paper we report a wavelet-based tool for the analysis of LRD traffic to allow for semi-parametric estimation of the $H$ parameter. Validation has been done using fBm and fGn models, and the obtained estimations show excellent agreements with the theoretical results.

There are several interesting areas to continue this work. A first one is related to the selection of the scales over which the estimation is done. There is an imperious need for a well founded statistical basis for the selection of the scaling range. This range, unknown a priori, should correspond to the LRD (scaling) regime of the sequence.

The next topic of interest is related to the wavelet-based analysis of VBR traces. Due to the statistical complexity of the VBR video traffic, with correlations on several time scales, we expect the wavelet-based tool to be very useful.

It is also of interest to use wavelet-based techniques for the generation of LRD traffic traces which can be used for simulation studies or lab testing of network components. Efficient generation techniques are necessary to speed up simulation times, especially when simulating extremely high-speed networks. Real load testing of network components also requires real-time generation of true LRD traffic at line rates (e.g., OC-12 to OC-48).

Finally, analysis of multi-fractal properties of traffic represents an area of increasing interest. This analysis offers the choice to complement the knowledge gained at low frequencies in the spectrum (LRD) with analysis at the other end of the spectrum, and therefore reveal properties of the traffic at small time scales. Especially in the case of WAN

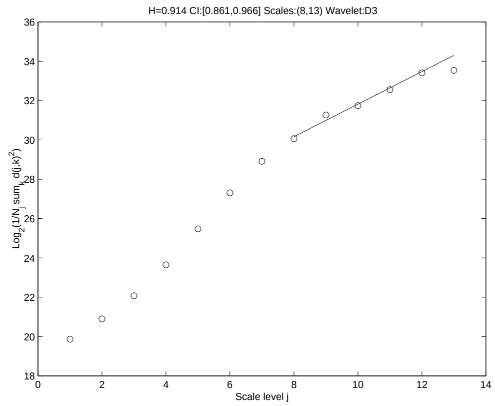


**Figure 7.** Estimation of $H$ and a 95%-confidence interval on the *Starwars* trace, using the Daubechies 3 wavelet and weighted linear regression. $H = 0.914$ CI:[0.861, 0.966]

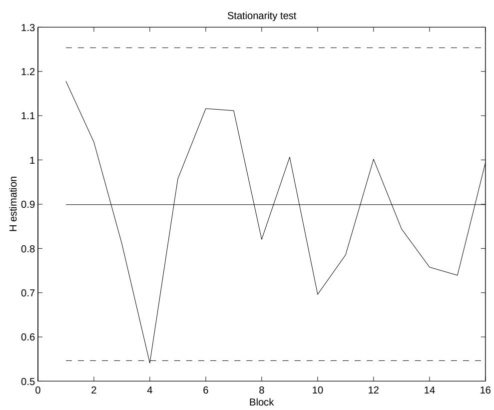


**Figure 8.** The stationary test using the Haar wavelet. The partial estimations are centered around the estimation on the entire trace, and within the confidence interval. Some of the partial estimations are larger than $H = 1$, which indicates the presence of non-stationarities on smaller timescales.

traffic traces[11] and VBR video traffic, multi-fractals have the potential to provide for a structural modeling approach able to capture, in a compact and parsimonious manner, the observed scaling phenomena at large and small time scales.

## Acknowledgment

We would like to thank Darryl Veitch for numerous useful discussions on this topic.

## REFERENCES

1. M. S. Taqqu, V. Teverovsky and W. Willinger, "Estimators for Long-Range Dependence: An Empirical Study," *Fractals,* **3**, No. 4, 1995.
2. J. Beran, *Statistics for Long-Memory Processes,* Chapman and Hall, 1994.
3. P. Abry and D. Veitch, "Wavelet Analysis of Long Range Dependent Traffic," *IEEE Transactions on Information Theory,* **44**, No. 1, pp. 2-15, January 1998.
4. D. Boulliant, *Estimating Long-Range Dependence Using Wavelets,* technical report, University of Karlskrona/Ronneby, December 1997.

5. D. Veitch and P. Abry, *A Wavelet Based Joint Estimator for the Parameters of Long-Range Dependence,* preprint, 1996, http://www.serc.rmit.edu.au/~darryl
6. P. Flandrin, "Wavelet Analysis and Synthesis of Fractional Brownian Motion," *IEEE Transactions on Information Theory,* **IT-35**, 1992.
7. W. Leland, M. Taqqu, W. Willinger and D. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," *IEEE/ACM Transactions on Networking,* **2**, No. 1, pp. 1-15, February 1994.
8. D. V. Wilson, *Bellcore Ethernet Traces,* http://ita.ee.lbl.gov/html/contrib/BC.html
9. M. W. Garrett and W. Willinger, "Analysis, Modelling and Generation of Self-Similar VBR Video Traffic," *Proc. ACM SIGCOMM'94,* pp. 269-280, London, UK, 1994.
10. M. W. Garrett, *VBR Video Traces,* ftp://ftp.bellcore.com/pub/vbr.video.trace/
11. A. Feldmann, A. C. Gilbert, W. Willinger and T. G. Kurtz, "The Changing Nature of Network Traffic: Scaling Phenomena," *Computer Communications Review,* **28**, No. 2, pp. 5-29, April 1998.