Estimation of Long Range Dependence Using Wavelets

David Boulliant, Parag Pruthi and Adrian Popescu
E-mail: {dbo, ppr, adrian}@itm.hk-r.se
Phone: +46 455 78151 or +46 455 78061
Fax: +46 455 78057
Department of Telecommunications and Mathematics
University of Karlskrona/Ronneby
371 79 Karlskrona, Sweden

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 LRD) 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. The estimator has been proved to be unbiased under fractional Brownian motion fBm and Gaussian assumptions. Analysis of the Bellcore Ethernet traces using the wavelet-based estimator is also reported.

2 Estimation of Long-Range Dependence

Several methods are available today to estimate LRD parameter and/or the intensity of LRD in a 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 [6].

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 [6]. 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.

3 Wavelet-Based Estimation

The wavelet estimator is based on the discrete wavelet transform (DWT). The underlying concept of the DWT is the so-called Multi-Resolution Analysis (MRA), which consists of splitting the sequence into a (low-pass) approximation and a (high-pass) detail. This estimator has the advantages of both aggregation-based and MLE estimators and avoids their drawbacks.

In the wavelet case, the focus is 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. In contrast, the wavelet constant relative bandwidth manages to provide a perfect match.

Furthermore, 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 scales/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. Accordingly, weighted least-square estimations can be performed. Actually, there are solutions reported even without the asymptotic simplification, with the consequence of more effective weights and smaller confidence intervals [7].

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 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 [7]. This property 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 rises the problem of having to test for the stationarity assumptions in the process. This may require additional knowledge of the underlying functionality of the process. It may be possible to handle this additional requirement by selection of relevant time scales where the process is stationary, but the relevant time scales are not known apriori.

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

In practice, for removing the effects of polynomial trends, one must make repeated estimations 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 [1]. 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 [3]. 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 (even less than FFT) 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. The solution to this problem is simple, we simply remove the corrupted wavelet coefficients from the data set, and by this the boundary effects. Especially in the case of large data sets, wavelet coefficients can be removed without loss of accuracy. This method is also valid for non-stationary processes [4]. This is to compare with another method existent in the literature (connecting the sequence ends), which is not applicable for this case [3].

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 [3]. The sequence can be split 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. Excellent agreements with the theoretical results have been obtained for both H estimates as well as for the corresponding confidence intervals.

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.

$\hat{H}_w$	$\hat{\sigma}_w$	$\hat{H}_s$	$\hat{\sigma}_s$
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 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 (shown inside parenthesis). The estimations where performed through scales (2, 7) using weighted and standard linear regression. The Haar wavelet was used ($N = 1$).

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

Table 2: H estimations using Daubechies wavelet with $N = 2$ and 3, for trend elimination supperimposed on a fGn trace 40 repetetive times with $H = 0.8$ and $n_0 = 4096$. As being expected, $N = 1$ is not enough to eliminate the linear trend.

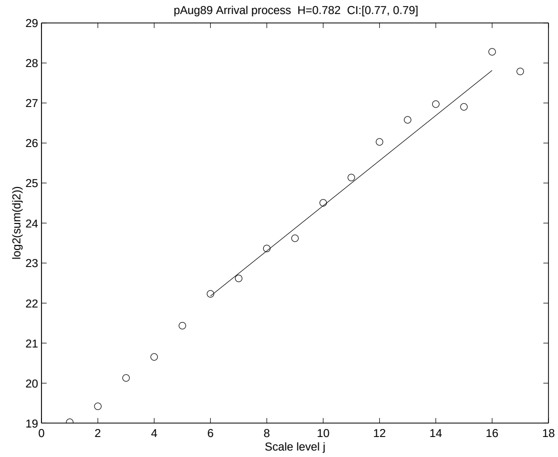

Figure 1: H estimation of Bellcore traces, namely the discrete arrival process with $H = 0.782 \pm 0.011$. A clear power-law behavior is observed over almost all scales. A Daubechies wavelet was used with $N = 2$.

Process	$N = 1(6,16)$	$N = 2(6,16)$	$N = 3(6,14)$
Arrival	1.02 ± 0.011	0.782 ± 0.011	0.794 ± 0.012
Point	0.826 ± 0.011	0.830 ± 0.011	0.838 ± 0.012

Table 3: H estimations using Daubechies wavelet with $N = 1, 2$ and 3, for trend elimination superimposed on a fGn trace 40 repetitive times with $H = 0.8$ and $n_0 = 4096$. 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.

6 Measurements

Estimations of the well-known Bellcore Ethernet traces [5] have been done. The traces consist of one million timestamps in seconds and frame sizes in bytes. The focus is laid on the discrete version of the arrival process and the corresponding point-process, using a sampling interval of 6 ms. Both processes are LRD with almost similar H values (0.830 and 0.782 ± 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 (figs. 1 and 2).

The stationary condition is also tested (fig. 3). The estimations are centered around the H estimation over the entire trace, and within the (asymptotic) confidence interval. Hence the trace could be treated as stationary. The results are well matching similar estimations reported in the literature [5].

7 Conclusions

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

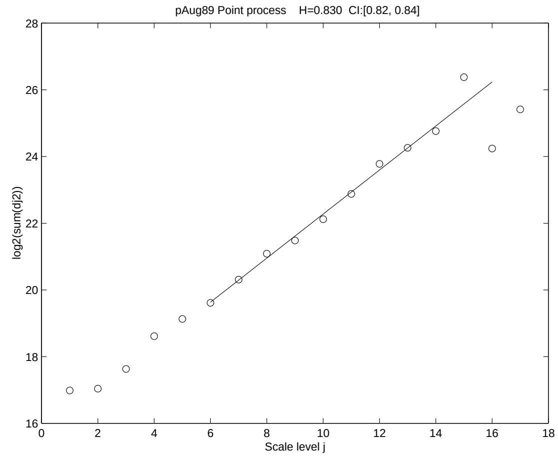

Figure 2: H estimation of Bellcore traces, namely the discrete point process with $H = 0.830 \pm 0.011$. A clear power-law behavior is observed over almost all scales, starting at approximatly 20 ms. A Daubechies wavelet was used with $N = 2$.

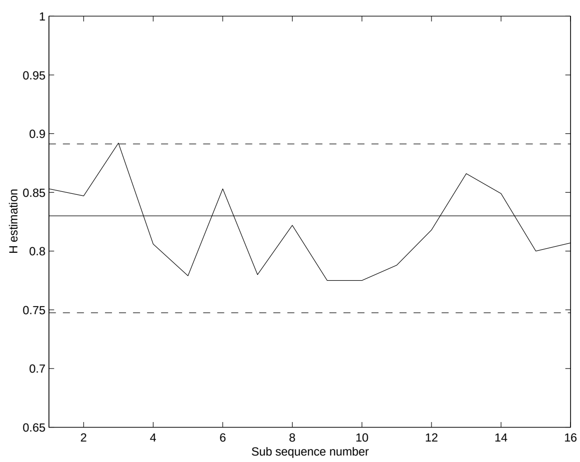

Figure 3: Repeated H estimations on 16 subsequences. The confidence interval (dotted lines) is found to be centered around the H estimation over the entire trace (solid line), and all separate H estimations are found within. A Daubeches was used with $N = 2$.

There are several relevant 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.

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.

Acknowledgment

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

References

[1] Abry, P. and Veitch, D., *Wavelet Analysis of Long Range Dependent Traffic,* IEEE Transactions on Information Theory, December 1997.

[2] Beran, J., *Statistics for Long-Memory Processes,* Chapman and Hall, 1994.

[3] Boulliant, D., *Estimating Long-Range Dependence Using Wavelets,* technical report, Univ. of Karlskrona/Ronneby, December 1997.

[4] Flandrin, P., *Wavelet Analysis and Synthesis of Fractional Brownian Motion,* IEEE Transactions on Information Theory, IT-35, 1992.

[5] Leland, W., Taqqu, M., Willinger, W. and Wilson, D., *On the Self-Similar Nature of Ethernet Traffic (Extended Version),* IEEE/ACM Transactions on Networking, February 1994.

[6] Taqqu, M.S., Teverovsky, V. and Willinger, W., *Estimators for Long-Range Dependence: An Empirical Study,* Fractals, Vol. 3, No. 4, 1995.

[7] Veitch, D. and Abry, P., *A Wavelet Based Joint Estimator of the Parameters of Long-Range Dependence,* preprint, 1996.