PERFORMANCE ANALYSIS OF A WAVELET-BASED HURST PARAMETER ESTIMATOR FOR SELF-SIMILAR TRAFFIC

Fei Xue and Ljiljana Trajković*
School of Engineering Science
Simon Fraser University
Vancouver, British Columbia, Canada
e-mail: {fxue, ljilja}@cs.sfu.ca

Keywords: Communications, Statistical analysis, Wavelet transform, Discrete simulation, Long-range dependence.

Abstract

In this paper we evaluate the performance of a wavelet-based estimator for the estimation of the Hurst parameter used in characterizing self-similar network traffic. We use genuine traffic traces and simulation tools to investigate the impact of short-range dependence on the accuracy of the estimated parameter. We briefly address the implication of our findings on traffic engineering. Our main conclusion is that this wavelet-based estimator can be used to analyze aggregated traffic data. Nevertheless, the estimator proves to be unsuitable for the medium- and high-bursty video traffic because of the traffic's complex correlation structure.

1 Introduction

Recent studies based on high quality traffic measurements in various packet networks have shown that traffic processes exhibit self-similarity, i.e., long-range dependence (LRD) [2, 7, 9]. It is commonly accepted that the definition of LRD is the slow power-law decrease of the autocorrelation function of a wide-sense-stationary process $X(t), t = 0, 1, ...$, i.e.,

$$r(k) \sim c_r k^{-(2-2H)}, \qquad k \to \infty, \tag{1}$$

where $c_r > 0$ and $0.5 < H < 1$. The key parameter in this description is the Hurst parameter H that captures the long-term correlation structure. In the case $0.5 < H < 1$, long-range dependence occurs and the process exhibits self-similarity. An equivalent property of an LRD process is that for a range of frequencies λ close to 0, the spectral density $f(\lambda)$ of $X(t)$ has the following form:

$$f(\lambda) \sim c_f |\lambda|^{1-2H}, \qquad \lambda \to 0, \tag{2}$$

where $c_f > 0$ and $0.5 < H < 1$. A reliable estimation of the Hurst parameter is of crucial importance in traffic engineering. Recent studies [2, 4, 8] have already revealed that self-similar traffic has a significant impact on network performance.

Wavelet analysis is a useful tool for analyzing the scale-dependent properties of data [3, 10]. Since self-similar processes possess similar statistical characteristics over a large range of scales, wavelet analysis appears to be a natural tool to test the fractal behavior of a long-range dependent process [1].

In Section 2, we survey the fundamentals of wavelet analysis and describe the wavelet-based estimator that we evaluated. Section 3 presents the experimental results using genuine traffic traces. In Section 4, we give simulation results for large sets of sample paths based on FARIMA$(1, d, 0)$ models. We conclude with Section 5 and discuss the implications of our findings on traffic engineering.

2 Wavelet analysis

2.1 Discrete wavelet transform

Wavelet transform is the representation of a signal $X(t)$ by means of its inner products with a set of basic wavelet functions $\psi_{a,t}(u) = \psi_0(\frac{u-t}{a})/\sqrt{a}$, called the wavelets. They are scaled and shifted versions of an adequately chosen mother wavelet ψ_0. $X(t)$ is then transformed into a scale-time wavelet domain (a,t), $a \in R^+$ and $t \in R$. Wavelet transform provides a tool for simultaneously observing a time se-

*This research was supported by the NSERC Grant 216844-99 and the BC Advanced Systems Institute Fellowship.

ries at a full range of different scales a, while retaining the time dimension t of the original data. Multiresolution analysis (MRA) theory shows that no information is lost if we sample the continuous wavelet coefficients at a sparse set of points in the scale-time plane known as the dyadic grid. This grid, defined by $(a, t) = (2^j, 2^j k)$, $j, k \in N$, leads to the discrete wavelet transform. Detailed background on wavelets and MRA is given in [3, 10].

An efficient implementation of the discrete wavelet transform is given by the filter-bank-based pyramidal algorithm. In this algorithm, the mother wavelet ψ_0 is represented in terms of the so-called scaling function ϕ_0:

$$\psi_0 = \sqrt{2} \sum_n (-1)^n h(1-n) \phi_0(2t-n). \tag{3}$$

ϕ_0 is defined by a two-scale difference equation, known as the dilation equation:

$$\phi_0 = \sqrt{2} \sum_n h(n) \phi_0(2t-n). \tag{4}$$

The discrete-time filters $h(n)$ and $g(n) = (-1)^n h(1-n)$ have the low-pass and high-pass characteristics, respectively. Given a time series $X(k)$ of length N and the initial approximation coefficients $a_x(0, k) = X(k)$, we can compute the so-called detail coefficients $d_x(j, k)$ at scale 2^j $(1 \le j \le log_2 N)$ by convolving the approximation coefficients with filters $h(n)$ and $g(n)$:

$$\begin{aligned} d_x(j,k) &= g * a_x(j-1,k) \\ a_x(j,k) &= h * a_x(j-1,k). \end{aligned} \tag{5}$$

2.2 Wavelet-based Hurst parameter estimator

A wavelet-based Hurst parameter estimator, proposed in [1], is based on a spectrum estimator designed by performing a time average of $|d_x(j,k)|^2$ at a given scale:

$$\Gamma_x(2^{-j}\nu_0) = \frac{1}{n_j} \sum_k |d_x(j,k)|^2, \tag{6}$$

where ν_0 is an arbitrary reference frequency selected by the choice of the mother wavelet ψ_0, and n_j is the number of wavelet coefficients at octave j. The linear relationship between $log_2(\Gamma)$ and level j for a range of scales $[j_1,\ j_2]$ indicates the presence of an LRD behavior. An estimator $\hat{H}$ for the Hurst parameter H can be designed by performing a linear regression of $log_2(\Gamma)$ on scale level j in the range $[j_1, j_2]$:

$$log_2(\Gamma_x(2^{-j}\nu_0)) = (2\hat{H} - 1)j + \hat{c}. \tag{7}$$

It is important to determine the range of time scales $[j_1,\ j_2]$ when using an asymptotic estimator. However, it is still an open question [1] how to choose a priori $[j_1, j_2]$. In this paper we investigate the choice of scales when estimating Hurst parameter.

We implemented the estimator in MATLAB and selected Daubechies wavelets as the mother wavelets.

3 Experiments with traffic traces

In order to test the wavelet-based estimator, we used two Ethernet traces (*pAug.TL*, *pOct.TL*) from the traffic traces collected at Bellcore in the late 1980's and early 1990's [7]. We considered samples of the rate processes obtained from these traces. Each sample corresponds to the number of packet arrivals over 10 ms.

Fig. 1 illustrates the wavelet analysis for trace *pOct.TL*. We have plotted on a log-log scale the spectrum Γ as a function of scale 2^j, i.e., $log_2(\Gamma)$ is plotted as a function of j. If the traces are self-similar, a plot

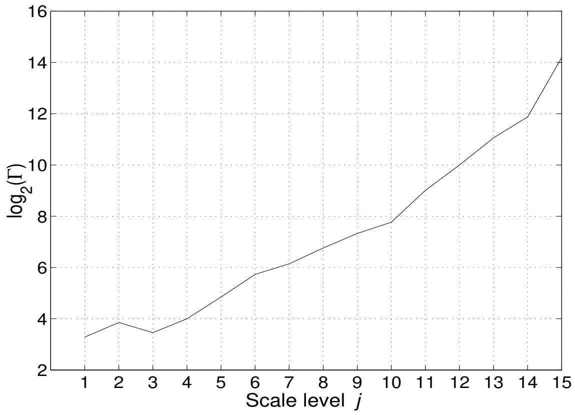

Figure 1: The $log_2(\Gamma)$ vs. j plot of the trace pOct.TL.

of $log_2(\Gamma)$ vs. j will show a linear relationship over all time scales. Fig. 1 shows an approximate linear relationship between $log_2(\Gamma)$ and level j for sufficiently large time scales. Therefore, these traffic traces are asymptotically self-similar and long-range dependent processes. The notable deviation from linearity for small value of j is caused by the short-range dependence (SRD) of the process.

We apply the estimator over a selected range of scales $[j_1,\ j_2]$ and obtain the estimated H values, given in Table 1. These values are consistent with results presented in [7], i.e., the estimated H of these two traces range from 0.8 to 0.9. As shown in Table 1, the wavelet estimator is sensitive to the selected

$[j_1,\ j_2]$	$\hat{H}$	95% Confidence interval
[7, 11]	0.818	[0.780, 0.856]
[7, 12]	0.804	[0.771, 0.838]
[8, 11]	0.819	[0.757, 0.881]
[8, 12]	0.817	[0.764, 0.870]

Table 1: Estimated Hurst parameters for *pOct.TL*.

range $[j_1,\ j_2]$. This range should correspond to the location where the process exhibits asymptotic LRD behavior. From the statistical viewpoint, network traffic can be regarded as an asymptotic LRD process, rather than an exact one. In reality, it possesses short-term correlation in addition to the long-range dependent structure. The SRD component is the reason that the linear dependence between $log_2(\Gamma)$ and level j holds only over a limited range. Therefore, in order to determine the correct range $[j_1,\ j_2]$, it is of fundamental importance to study the effects of SRD on the wavelet-based estimator.

4 Experiments with FARIMA processes

In order to investigate the impact of SRD on the performance of this estimator, we use the fractionally differenced autoregressive integrated moving-average (FARIMA) process [5, 6]. This process possesses the LRD and SRD properties simultaneously. We apply wavelet-based estimator [1] on large sets of data generated by FARIMA$(1, d, 0)$ model in order to evaluate its performance.

4.1 FARIMA process

FARIMA processes are the generalizations of standard ARIMA processes when the degree of differencing d is allowed to be a fraction in the range $(-0.5, 0.5)$. The FARIMA(p, d, q) process $X(t), t = ..., -1, 0, 1, ...$, where p and q are non-negative integers and $d \in (-0.5, 0.5)$, is described by the difference equation

$$\Phi(B) \triangle^d X(t) = \Theta(B)a(t), \tag{8}$$

where in the backward-shift operator B

$$\begin{aligned} \Phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p \\ \Theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q. \end{aligned} \tag{9}$$

are polynomials of order p and q, respectively. $\triangle^d$ is the fractional differencing operator, and $a(t)$ is a white noise process with constant variance σ^2. When $\Phi(z)$ and $\Theta(z)$ have all the roots outside the unit circle and have no common roots, the process $X(t)$ is stationary and invertible. In the case $0 < d < 1/2$, long-range dependence occurs and the process exhibits self-similarity with Hurst parameter $H = d + 0.5$.

FARIMA(p, d, q) process can be decomposed into two basic components: ARIMA(p,q) and FARIMA$(0, d, 0)$ processes, where the former describes the short-range and the latter describes the long-range behavior.

4.2 Performance results

We selected the FARIMA$(1, d, 0)$ model to generate traces that exhibit LRD and SRD properties simultaneously. The parameter d is the indicator for the presence of LRD. The autoregressive coefficient ϕ_1 can be used as an asymptotic measure of the SRD structure. FARIMA$(0, d, 0)$ process can be regarded as a special form of FARIMA$(1, d, 0)$ with $\phi_1 = 0$.

$\phi_1 = 0.3$		$\phi_1 = 0.5$	
$[j_1,\ j_2]$	$\hat{H}$	$[j_1,\ j_2]$	$\hat{H}$
[2, 7]	0.786	[3, 10]	0.764
[3, 5]	0.736	[4, 8]	0.698
[3, 6]	0.694	[4, 9]	0.695
[3, 7]	0.707	[4, 10]	0.693

Table 2: Estimated Hurst parameters for FARIMA$(1, d, 0)$ with $d = 0.15$ $(H = 0.65)$, $\phi_1 = 0.3$ and 0.5.

We generated FARIMA$(1, d, 0)$ traces with four values of d (0.15, 0.25, 0.35, 0.45) and four values of ϕ_1 (0, 0.1, 0.3, 0.5). Each trace contains 16,384 samples. Traces with identical d's have distinct short-range correlation structures, although they share the same long-range dependent behavior. We classify these traces into four sets, according to the parameter d of the underlying model. A summary of the wavelet analysis for each set is shown in Fig. 2. As seen from Fig. 2, traces with large d (0.35 and 0.45) have rather similar properties. The differences are insignificant from the statistical point of view. In contrast, the traces with small d do not have the same linear relationship between $log_2(\Gamma)$ and j.

We estimated the Hurst parameter of these traces over various scales $[j_1,\ j_2]$ and summarized the findings by observing four cases:

1. *Traces FARIMA$(1, d, 0)$ with* $\phi_1 = 0$: The estimator can capture the value of H consistently.

2. *Traces FARIMA(*$1, 0.15, 0$*) with* $\phi_1 = 0.3$ *and* 0.5: Because it is difficult to determine the scale range where the linear behavior holds, the estimated Hurst parameters deviate from the targeted values, as shown in Table 2. This visible deviation can also be found for trace FARIMA$(1, 0.25, 0)$ with $\phi_1 = 0.5$.

$d = 0.35$		$d = 0.45$	
$[j_1, j_2]$	$\hat{H}$	$[j_1, j_2]$	$\hat{H}$
[4, 8]	0.872	[4, 9]	1.013
[4, 9]	0.868	[5, 7]	1.056
[5, 8]	0.870	[5, 8]	1.037
[5, 9]	0.891	[5, 9]	1.025

Table 3: Estimated Hurst parameters for a FARIMA$(1, d, 0)$ process with $\phi_1 = 0.5$, $d = 0.35$ and 0.45 ($H = 0.85$ and 0.95).

3. *Traces FARIMA(*$1, 0.35, 0$*) with* $\phi_1 = 0.5$ *and FARIMA(*$1, 0.45, 0$*) with* $\phi_1 = 0.5$: Although the estimated H parameters are not accurate, as shown in Table 3, the inaccuracy can not be explained by the lack of linear behavior. In contrast, we can observe from Fig. 2 that there are apparent linear relationships. When both long- and short-range correlation structures are present in a finite-length trace, especially in the case when both structures have a strong presence, they are hard to distinguish. Hence, the estimated H parameters tend to be larger because of the combined effects of the LRD and SRD properties.

We confirmed this finding by analyzing a JPEG coded "*Star Wars*" traffic trace at frame level [2]. This trace is of a medium burstiness. In general, a video traffic in a medium- or a high-burstiness class exhibits a complex temporal behavior. An earlier study has revealed that the estimated H of this trace is between 0.84 to 0.97. The study used a periodogram-based estimator with various aggregation levels [2]. As shown in Table 4, H parameters estimated by a wavelet-based estimator over four scale ranges are larger than 1. The corresponding wavelet analysis plot is shown in Fig. 3. This effect indicates that the wavelet-based estimator cannot perform well enough to capture the H of a process with complex correlation structure.

4. *Remaining traces*: Simulated results show that the embedded SRD slightly affects the performance of the wavelet-based estimator, especially for data with larger d. For traces with smaller d, as for example FARIMA$(1, 0.15, 0)$ with $\phi_1 = 0.1$, and FARIMA$(1, 0.25, 0)$ with $\phi_1 = 0.1$ and 0.3, the esti-

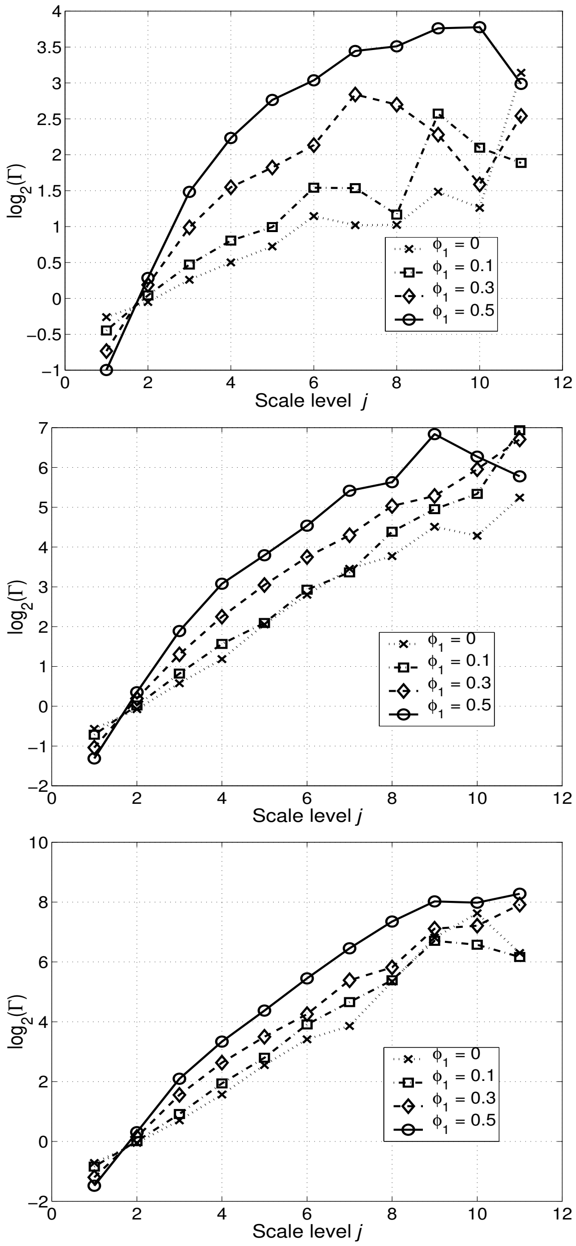

Figure 2: The $log_2(\Gamma)$ vs. j plots for FARIMA $(1, 0.15, 0)$ set (top), FARIMA$(1, 0.35, 0)$ set (middle) and FARIMA$(1, 0.45, 0)$ set (bottom).

$[j_1, j_2]$	[3, 11]	[3, 12]	[4, 11]	[4,12]
$\hat{H}$	1.191	1.138	1.158	1.100

Table 4: Estimated Hurst parameters for the "*Star Wars*" traffic trace.

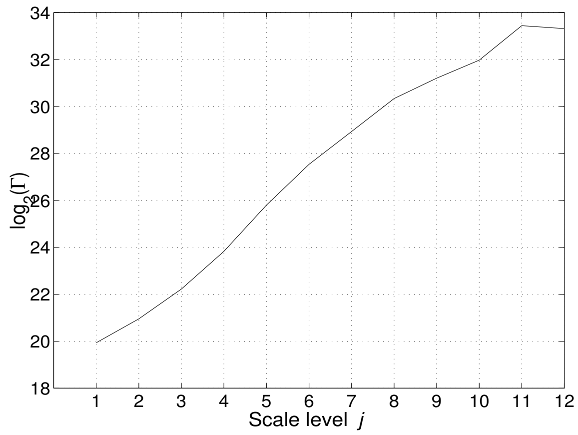

Figure 3: The $log_2(\Gamma)$ vs. j plot for "*Star Wars*" traffic trace.

mated H tends to be large. In order to obtain an accurate H, because of the weak impact of SRD in these cases, we propose to select the minimum of all possible estimated H parameters that were obtained for different ranges $[j_1, j_2]$. Table 5 presents the results based on this approach. These values of H are very close to the targeted ones. Since the SRD property of an aggregated traffic trace may not be as strong as in the case of a trace obtained from a purely video application, the suggested method is useful for the accurate estimation of Hurst parameters in measured packet traffic traces.

	$d = 0.15$ $\phi_1 = 0.1$	$d = 0.25$ $\phi_1 = 0.1$	$d = 0.25$ $\phi_1 = 0.3$
$[j_1, j_2]$	[3,7]	[3,7]	[4,9]
$\hat{H}$	0.663	0.755	0.742

Table 5: The estimated Hurst parameters using suggested method.

5 Concluding remarks

In this paper we evaluated the performance of a wavelet-based estimator and investigated the impact of SRD structure on the quality of this estimator.

Based on the simulation analysis, we observed that the wavelet-based estimator is not suitable for processes with strong SRD and either weak or strong LRD component. We confirmed this finding by analyzing a medium-bursty video traffic trace with complex correlation structure. This implies that we cannot apply the wavelet-based estimator to analyze the medium- and high-bursty video traffic.

For processes with weak LRD and either weak or medium SRD component, wavelet-based estimator overestimates the Hurst parameter. We proposed a method that can reduce this effect and may prove useful for the accurate estimation of H in measured traffic traces from packet networks.

For the remaining processes, we verified that the evaluated wavelet-based estimator is efficient and robust enough to accurately capture Hurst parameter.

References

[1] P. Abry and D. Veitch, "Wavelet analysis of long-range-dependent traffic," *IEEE Trans. Information Theory*, vol. 44, no. 1, pp. 2–15, 1998.

[2] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1566–1579, 1995.

[3] C. K. Chui, *An Introduction to Wavelets.* Boston: Academic Press, 1992.

[4] N. G. Duffield and N. O'Connell, "Large deviations and overflow probabilities for the general single-server queue, with application," *Proc. Cambridge Philosophy Society*, pp. 363–375, 1995.

[5] J. Hosking, "Fractional differencing," *Biometrika*, vol. 68, pp. 165–176, 1981.

[6] C. W. J. Granger and R. Joyeux, "An introduction to long-memory time series models and fractional differencing," *J. Time Series Analysis*, vol. 1, pp. 15–29, 1980.

[7] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, 1994.

[8] I. Norros, "On the use of fractional Brownian motion in the theory of connectionless networks," *IEEE J. Selected Areas Commun.*, vol. 13, pp. 953–962, 1995.

[9] V. Paxson and S. Floyd, "Wide-area traffic: the failure of Poisson modeling," *Proc. ACM SIGCOMM'94*, 1994, pp. 257–268.

[10] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software.* Wellesley, MA: A. K. Peters, 1994.