Proceedings of the 27th Annual IEEE Conference on Local Computer Networks, pp. 160-165, Tampa, Florida, November 2002

A NOVEL APPROACH TO THE ESTIMATION OF THE HURST PARAMETER IN SELF-SIMILAR TRAFFIC

Houssain Kettani and John A. Gubner
Electrical and Computer Engineering Department
University of Wisconsin
Madison, WI 53706
kettani@cae.wisc.edu
gubner@engr.wisc.edu

Abstract

We present a new method to estimate the Hurst parameter of the increment process in network traffic – a process that is assumed to be self-similar. The confidence intervals and biasedness are obtained for the estimates using the new method. This new method is then applied to pseudo-random data and to real traffic data. We compare the performance of the new method to that of the widely-used wavelet method, and demonstrate that the former is much faster and produces much smaller confidence intervals of the Hurst parameter estimate. We believe that the new method can be used as an on-line estimation tool for H and thus be exploited in the new TCP algorithms that exploit the known self-similar and long-range dependent nature of network traffic.

1 Introduction

It is now generally accepted that network traffic exhibits the features of long-range dependence and self-similarity [8, 10, 11, 12, 13, 4, 14]. The parameter that measures these features is known as the Hurst parameter, H, and many methods for estimating H have been proposed. For example, the following methods are described in the text by Beran [3]: R/S method, variance-time analysis, Higushi's method, correlogram method, periodogram method and Whittle estimator. More recent methods are residuals of regression method due to Peng *et al.* [16] (see also [17]) and the wavelet method due to Abry and Veitch [1]. By far, the wavelet method is the most widely used.

Almost all of the above methods are based on asymptotics rather than the exact form of the covariance function. However, since LAN traffic is generally accepted as *exactly* second-order self-similar [8], which specifies the form of the covariance function, we propose a new method for estimating H that exploits this structure. The new method is much faster and yields smaller confidence intervals than the wavelet method.

The remainder of the paper is organized as follows. Section 2 presents mathematical definitions and properties that will be used throughout the paper. In Section 3, we present and describe the proposed method for estimating the Hurst parameter. We then apply this novel method to artificial and real traffic data in Section 4 to check its performance, and compare it with that of the wavelet method. Lastly, we present concluding remarks and further research directions in Section 5.

2 Preliminaries

Let X_i denote the number of bits, bytes or packets seen during the ith interval. We say that X_i is *second-order stationary* if its mean $E(X_i)$ does not depend on i and if the autocovariance function

$$E[(X_i - E(X_i))(X_j - E(X_j))]$$

depends on i and j only through their difference $k = i - j$, in which case we write

$$\gamma(k) = E[(X_{i+k} - E(X_{i+k}))(X_i - E(X_i))].$$

We then put

$$\sigma^2 = \gamma(0) = E[(X_i - E(X_i))^2],$$

and

$$\rho(k) = \frac{\gamma(k)}{\sigma^2},$$

to denote the variance and autocorrelation function of the process X_i, respectively.

A second-order stationary process is said to be *exactly second-order self-similar* with Hurst parameter $0 < H < 1$, if

$$\gamma(k) = \frac{\sigma^2}{2}(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}),$$

or equivalently,

$$\rho(k) = \frac{1}{2}(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}). \quad (1)$$

If X_i is a Gaussian process, it is known as *fractional Gaussian noise*.

3 The New Method

Since X_i is exactly second-order self-similar, we have from (1) that

$$\rho(1) = 2^{2H-1} - 1,$$

We can solve for H to get

$$\hat{H} = \frac{1}{2}[1 + \log_2(1 + \rho(1))].$$

Given observed data $X_1, \ldots, X_n$, let

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\hat{\gamma}_n(k) = \frac{1}{n}\sum_{i=1}^{n-k}(X_i - \hat{\mu}_n)(X_{i+k} - \hat{\mu}_n),$$

$$\hat{\sigma}_n^2 = \hat{\gamma}_n(0)$$

and

$$\hat{\rho}_n(k) = \frac{\hat{\gamma}_n(k)}{\hat{\sigma}_n^2}, \quad (2)$$

denote the *sample mean*, the *sample covariance*, the *sample variance* and the *sample autocorrelation*, respectively.

We then put

$$\hat{H}_n = \frac{1}{2}[1 + \log_2(1 + \hat{\rho}_n(1))], \quad (3)$$

to denote the estimated Hurst parameter of the process X_i.

Theorem: Let X_i be an exactly second-order self-similar Gaussian process, i.e., fractional Gaussian noise. Assume in (1) that $0 < H \leq \frac{3}{4}$. Then for large sample size n, $\hat{\rho}_n(1)$ is approximately N(μ_n, σ_n^2), where

$$\mu_n = \rho(1) - (1 - \rho(1))n^{2H-2},$$

and

$$\sigma_n^2 = \frac{1}{n}\{(1 + 3\rho^2(1)) + 2\sum_{k=1}^{\infty}\Big[(1 + 2\rho^2(1))\rho^2(k)$$

$$+\rho(k-1)\rho(k+1) - 4\rho(1)\rho(k-1)\rho(k)\Big]\}, \quad (4)$$

if $H \in (0, \frac{3}{4})$ and

$$\sigma_n^2 = \frac{\log n}{n}[2H(2H-1)(1 + \rho(1))]^2, \quad (5)$$

if $H = \frac{3}{4}$.

Proof: This is a special case of Hosking's result [6].

A plot of $n\sigma_n^2$ in (4) (which does not depend on n) as a function of the Hurst parameter H, summing over

$k = 1$ to 10^7 (instead of $k = 1$ to ∞ as in (4)) is given in Figure 1. Now that we know $\hat{\rho}_n(1)$ is N(μ_n, σ_n^2),

$$P(|\frac{\hat{\rho}_n(1) - \mu_n}{\sigma_n}| \leq 1.96) = 0.95,$$

i.e.,

$$\mu_n - 1.96\sigma_n \leq \hat{\rho}_n(1) \leq \mu_n + 1.96\sigma_n$$

holds with 95% probability. Using (3),

$$h_- \leq \hat{H}_n \leq h_+,$$

where

$$h_\pm = \frac{1}{2}\{1+\log_2[1+\rho(1)-(1-\rho(1))n^{2H-2}\pm 1.96\sigma_n]\}, \quad (6)$$

also holds with 95% probability.

Note that the Theorem may hold even if X_i is not fractional Gaussian noise (see discussion in [6]). For example, the Theorem still holds when X_i is fractional ARIMA (see [7] for similar analysis for such process). In the case when $H > \frac{3}{4}$, the limiting distribution in the Theorem exists, but it is not normal. The cumulants of this distribution are given in [6, Theorem 6]. For practical proposes, in constructing the confidence intervals of the estimate $\hat{H}_n$ for $H > \frac{3}{4}$, we propose using the mean and variance of the limiting distribution, assume normality and proceed as in the case of $H \in (0, \frac{3}{4})$. The validity of this approach will be investigated in Section 4. Thus, in such case, the value σ_n^2 in (6) is taken to be

$$\sigma_n^2 = (1 - \rho(1))^2 n^{4H-4}\kappa_2, \quad (7)$$

where κ_2 corresponds to the value of the variance of the limiting distribution. For instance, $\kappa_2 = 1.832, 0.518, 0.157$ and 0.003 for the values $H = 0.80, 0.85, 0.90$ and 0.95, respectively.

3.1 Comments on the Confidence Intervals

For known H, the 95% confidence interval of the estimate $\hat{H}_n$ is $[h_-, h_+]$, with h_- and h_+ as in (6), with σ_n as in Figure 1 if $H \in (0, \frac{3}{4})$, as in (5) if $H = \frac{3}{4}$

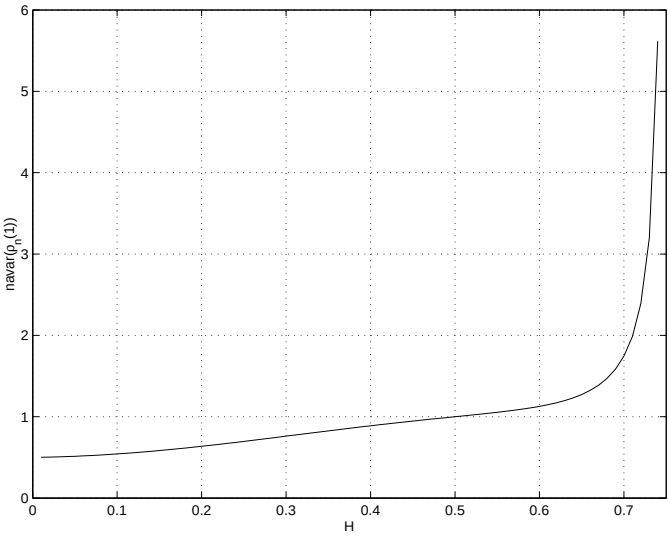

Figure 1. A plot of $n\sigma_n^2$ in (4) as a function of the Hurst parameter H, in the second-order self-similar case.

and as in (7) if $H \in (\frac{3}{4}, 1)$. Let w_n denote the width of such intervals, i.e.,

$$w_n = h_+ - h_-.$$

A log-log plot of w_n versus the number of samples n is given in Figure 2 for different values of H. It is remarkable to see the plot resembling a straight line for each value of H. Thus, the width w_n can be written as

$$w_n \approx an^{-b}, \quad (8)$$

where a and b are constants for fixed H. The values of these constants are given in Table 1. It is interesting to note that the width w_n is upper bounded by the w_n at $H = 0.74$. Hence in the case when H is not known (which is the typical case with real data), we choose the confidence interval centered around $\hat{H}_n$ with width

$$w_n = \frac{5}{\sqrt{n}}. \quad (9)$$

3.2 Summary of the Algorithm

In what follows, we present a summary of the new method:

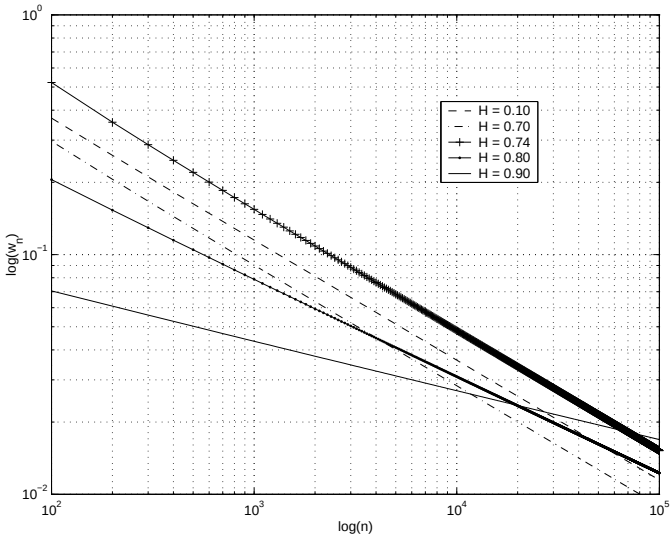

Figure 2. A plot of the width of the 95% confidence intervals for different H values

- Let $X_1, X_2, \ldots, X_n$ be a realization of a Gaussian second-order self-similar process,
- Compute $\hat{\rho}_n(1)$ as in (2),
- Compute $\hat{H}_n$ as in (3), which is the estimated Hurst parameter,
- The 95% confidence interval of H is centered around the estimate $\hat{H}_n$ with width as in (9).

4 Illustrative Examples

For each value of $H = 0.10, 0.20, \ldots, 0.90$, we generate 100 realizations of a fractional Gaussian noise. The length of each realization is $n = 4000$ points. For a given estimation method, we obtain 100 estimated values of H. Call these estimates $\hat{H}_n^{(k)}, k = 1, 2, \ldots, 100$. We compute their sample mean. We also provide the theoretical and empirical 95% confidence intervals of the estimates $\hat{H}_n$ and $\hat{H}_n^w$ from the proposed method and the wavelet method, respectively. The result of the application of the new method and the wavelet method to these data sets is given in Tables 2 and 3, respectively.

H	a	b
0.10	3.65	0.50
0.20	3.44	0.50
0.30	3.28	0.50
0.40	3.08	0.50
0.50	2.85	0.50
0.60	2.65	0.50
0.70	2.92	0.50
0.80	1.28	0.40
0.90	0.18	0.21

Table 1. The values of the constants a and b in (8) for each value of H.

From both Tables 2 and 3, it is observed that the confidence intervals obtained through the new method CI_o are narrower than those obtained through the wavelet method. The width of the empirical confidence intervals for the optimization method is about 0.05 versus 0.11 to 0.13 for those obtained through the wavelet method.

The theoretical and empirical confidence intervals are almost the same for the new method. This similarity holds even for $H \geq 0.80$, where we assumed normality although we knew that the distribution of $\hat{\rho}_n(1)$ is not normal. On the other hand, for the wavelet method, the empirical confidence intervals are considerably wider than theoretical ones, with the difference in width getting as high as 0.11 versus 0.07 for $H = 0.90$.

For $H < 0.80$, we see that the mean of the estimated Hurst parameter obtained by the new method $\hat{H}_o$ is the same as the true value H. For $H < 0.40$, the mean of the estimated Hurst parameter obtained by the wavelet method $\hat{H}_w$ is far from the true H and the latter does not fall in the 95% confidence interval. For $0.40 \leq H \leq 0.70$, the mean of $\hat{H}_w$ is closer to H and the theoretical confidence intervals contain the true value.

For $H = 0.80$, the new method under estimates the true value, with the mean of the estimates $\hat{H}_n$ is 0.79. The wavelet method, on the other hand, over estimates the true Hurst parameter value by the same

H	Mean of $\hat{H}_n$	Theoretical CI_o	Empirical CI_o
0.10	0.10	[0.06,0.14]	[0.07,0.13]
0.20	0.20	[0.17,0.23]	[0.17,0.23]
0.30	0.30	[0.27,0.33]	[0.27,0.32]
0.40	0.40	[0.37,0.43]	[0.37,0.43]
0.50	0.50	[0.48,0.52]	[0.47,0.52]
0.60	0.60	[0.58,0.62]	[0.58,0.63]
0.70	0.70	[0.67,0.73]	[0.68,0.72]
0.75	0.74	[0.72,0.77]	[0.72,0.76]
0.80	0.79	[0.77,0.81]	[0.77,0.82]
0.90	0.87	[0.86,0.90]	[0.85,0.90]

Table 2. Results of empirical and theoretical study of the new method using 100 independent realizations.

H	Mean of $\hat{H}_n^w$	Theoretical CI_w	Empirical CI_w
0.10	0.00	[-0.10,-0.02]	[-0.05,0.06]
0.20	0.16	[0.07,0.15]	[0.10,0.22]
0.30	0.28	[0.21,0.28]	[0.21,0.34]
0.40	0.39	[0.33,0.40]	[0.32,0.44]
0.50	0.50	[0.44,0.52]	[0.44,0.56]
0.60	0.60	[0.55,0.62]	[0.55,0.67]
0.70	0.70	[0.65,0.73]	[0.65,0.75]
0.75	0.76	[0.73,0.85]	[0.70,0.82]
0.80	0.81	[0.75,0.83]	[0.75,0.87]
0.90	0.91	[0.86,0.93]	[0.84,0.95]

Table 3. Results of empirical and theoretical study of the wavelet method using 100 independent realizations.

quantity, namely the mean of the estimates $\hat{H}_n^w$ is 0.81. For $H = 0.90$, the estimates produced by the wavelet method over estimate the true value by the same quantity, namely the mean of the estimates $\hat{H}_n^w$ is 0.91. The new method, on the other hand, under estimates H, with the mean of the estimates $\hat{H}_n$ is 0.87. In this case, On average, $\hat{H}_n^w$ are closer to the true value than $\hat{H}_n$. However, the empirical confidence intervals of the wavelet method are much larger than those of the new method, with the width of the former is almost double the latter. It is also worth noting that for $H > 0.20$, the confidence intervals of the estimates obtained by the new method are contained in those of the wavelet method.

The number of *flops* is 2.5×10^4 for the new method and 1.3×10^7 for the wavelet method. Thus, the former is 520 times faster than the latter. In general, it is apparent that the new method gives more accurate and reliable results and is much faster than the wavelet method.

We next consider real data to test both methods. For this purpose, we investigate Ethernet measurements for a local area network traffic at Bellcore, Morristown, New Jersey [8]. From this data we extract a data with length $n = 8475$ representing the amount of traffic observed each $100ms$. Passing the Bellcore data through the wavelet method gives $\hat{H}_n^w = 0.79$ with 95% confidence interval $CI_w = [0.75, 0.82]$. The new method, on the other hand, results in the value $\hat{H}_n = 0.81$ with confidence interval $CI_o = [0.78, 0.84]$. The variance-time, R/S, and periodogram methods resulted in $\hat{H} = 0.80, 0.79, 0.82$, respectively [8].

In short, it is clear that the new method and the wavelet method give close estimates with both real and artificial data. It is also noted that in all cases considered in this section, the new method's estimates fall in the 95% confidence intervals of the wavelet method. Moreover, the confidence intervals of the new method are contained in these of the wavelet method. Finally, the new method was shown to be much faster than the wavelet method.

5 Summary and Concluding Remarks

In this paper, we have presented a new tool to estimate the Hurst parameter in local area network traffic. The confidence intervals and biasedness of the estimates obtained by this new method are obtained. This new method is then applied to pseudo-random data and to real LAN traffic data. We compare the performance of the new method to that of the widely-used wavelet method. We demonstrated that the former is much faster and produces smaller confidence intervals

of the Hurst parameter estimates. Furthermore, the confidence intervals of the estimates from the new method were shown to be contained in the confidence intervals of the wavelet method. Moreover, the new method was found to be much faster than the wavelet method.

In view of the above, we believe that this method can be used as an on-line estimation tool for H and thus be exploited in the new TCP algorithms that exploit the known self-similar (and therefore long-range dependent) nature of network traffic. We mention for example, TCP – Traffic Prediction proposed in [5].

References

[1] P. Abry and D. Veitch (1998). Wavelet Analysis of Long-Range Dependent Traffic. *IEEE Transactions on Information Theory*, 44(1):2–15.

[2] T. W. Anderson (1958). *The Statistical Analysis of Time Series*. John Wiley & Sons, New York, New York.

[3] J. Beran (1994). *Statistics for Long-Memory Processes*. Chapman & Hall, New York, New York.

[4] M. Grossglauser and J. Bolot (1996). On the Relevance of Long-Range Dependence in Network Traffic. *Computer Communication Review*, 26(4):15–24.

[5] G. He, Y. Gao, J. C. Hou and K. Park (2002). A Case for Exploiting Self-Similarity of Internet Traffic in TCP Congestion Control. *Submitted to IEEE/ACM Transactions on Networking*

[6] J. R. M. Hosking (1996). Asymptotic Distribution of the Sample Mean, Autocovariances, and Autocorrelations of Long-memory Time Series. *Journal of Econometrics,* 73:261–284.

[7] H. Kettani (2002) A Novel Approach to the Estimation of the Long-Range Dependence Parameter. *Ph.D. Thesis, University of Wisconsin - Madison.*

[8] W. Leland, M. Taqqu, W. Willinger, and D. Wilson (1994). On the Self-Similar Nature of Ethernet Traffic (Extended Version). *IEEE/ACM Transactions on Networking*, 2(1):1–15.

[9] B. B. Mandelbrot (1983). *The Fractal Geometry of Nature*. W. H. Freeman and Co., New York, New York.

[10] R. Morris and D. Lin (2000). Variance of Aggregated Web Traffic. In *Proceedings of IEEE INFOCOM 2000*.

[11] K. Park, G. Kim, and M. Crovella (1996). On the Relationship Between File Sizes, Transport Protocols, and Self-Similar Network Traffic. In *Proceedings of the 4th International Conference on Network Protocols*.

[12] K. Park, G. Kim, and M. Crovella (1997). On the Effect of Traffic Self-Similarity on Network Performance. In *Proceedings of the SPIE International Conference on Performance and Control of Network Systems*.

[13] K. Park and W. Willinger (2000). Self-Similar Network Traffic: An Overview *Self-Similar Network Traffic and Performance Evaluation*. K. Park and W. Willinger (editors), John Wiley & Sons, New York, New York.

[14] V. Paxson and S. Floyd (1995). Wide-Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244.

[15] V. Paxson (1995). Fast Approximation of Self-Similar Network Traffic. *technical report LBL-36750/UC-405*.

[16] C. K. Peng, S. V. Buldyrev, M. Simons, H. E. Stanley and A. L. Goldberger (1994). Mosaic Organization of DNA Nucleotides. *Physical Review E*, 49:1685–1689

[17] M. S. Taqqu, V. Teverovsky and W. Willinger (1995). Estimators for Long-range Dependence: An Empirical Study *Fractals*, 3(4):785–788.

[18] P. Whittle (1953). Estimation and Information in Stationary Time Series *Arkiv for Matematik*, 2:423–434.