Computer Science Journal of Moldova, vol.**12**, no.2(35), 2004

On Estimating Long Range Dependence of Network Delay

Michael S. Borella

Abstract

We analyse 12 traces of round-trip Internet packet delay. We find that these traces, when viewed as time series data, often exhibit Hurst parameter (H) estimates greater than 0.5, indicating long-range dependence. Several traces, however, are not well modelled with a constant H. We discuss in detail our analytical methods and the robustness of empirical estimators of H under conditions of non-negligible packet loss. We also apply a newly-developed wavelet estimator of H and find that (1) it explains why periodogram-based estimators can produce inconclusive results, and (2) it produces results that can be used to estimate short range as well as long range dependence. Our overall results indicate that Internet delay is bursty across multiple time scales, which implies that end-user quality of service in the Internet is likely to be impacted by long periods of very large and/or highly variable delays.

Keywords: traffic model, Internet, time scales, periodogram

1 Introduction

It is now widely accepted that network traffic exhibits long-range dependence (LRD). However the practical implications of this discovery are not yet completely understood and research to date has shown that most empirical data does not perfectly fit any theoretical model. But it is known that traditional Poisson models of network traffic cannot capture the behaviour of LRD traffic. Often LRD data is also self-similar, indicating that it possesses similar statistical properties on multiple

time scales. Recently, a large amount of research has focused on the analysis of self-similar phenomena found in network traffic, as well as techniques for synthetically generating self-similar data sets for trace-driven simulation. These studies coincide with a dramatic change in the focus of Internet architecture and protocol research. The goal of the next generation Internet will be to seamlessly incorporate the services of current telephony and television networks into data networks. In order to do so efficiently, we must understand the dynamics of not only data traffic but also that of packet voice and video streams.

Traditionally, three parameters are considered to impact multimedia quality of service ($Q \circ S$): delay, delay variance (jitter) and loss rate. It is well known that LRD arrival patterns (or any linearly correlated arrival process with strong low-frequency components) has a deleterious impact on required buffer lengths as compared to non-correlated arrivals [1]. However, the overall interaction between delay, jitter, loss, LRD and $Q \circ S$ is not well understood. Thus, there is a need to evaluate the accuracy and applicability of LRD models.

In this paper we describe several empirical measurements of speech transmission over the Internet with parameters not unlike those of the industry-accepted ITU recommendation G.723.1. We find that although packet inter-departure times are deterministic, arrivals at the receiver exhibit LRD in most cases. We focus on obtaining estimates of the Hurst parameter (H) for delay traces by using the variance-time, R/S, periodogram, Whittle and wavelet estimators. Our major results and contributions are summarised as follows:

- Internet packet delay exhibits LRD; however, the degree of LRD, as indicated by H, varies significantly on both a temporal and spatial basis.

- Static estimation of LRD is accomplished for our data sets by using the Whittle estimator to fit a delay time series to Fractional Gaussian Noise or the wavelet estimator.

- Periodogram-based estimators, including the Whittle estimator, can produce inconclusive results when the spectral power of the

observed data set does not scale logarithmically with frequency; however, the wavelet estimator does not exhibit this problem.

- The wavelet estimator involves a spectral decomposition of network delay that produces two or three distinct frequency scaling for estimation of H. This indicates different dependency structures for high frequency and low frequency components and can be used to model both long range and short range dependence.

- Delays cannot always be well modelled with a constant H over long periods of time - a significant portion of our experiments find that H changes dramatically between consecutive 5-15 minute periods.

- Independent packet losses have a relatively minor impact on the robustness of the estimators.

This paper is organised as follows. Section 2 discusses the theoretical foundations of LRD, gives examples of LRD stochastic processes, and describes the LRD estimators in detail. Section 3 introduces the software used to make the delay measurements and describes the conditions under which each measurement was made. Section 4 describes the analysis of the delay measurements and our resulting point estimates of H, as well as situations in which H is not well-modelled as constant. Section 5 presents an analysis of the robustness of the LRD estimators in the presence of simulated packet loss. Section 6 introduced the wavelet estimator and applies it to our measurements. Section 7 presents our conclusions, as well as a discussion of the future directions of this research.

2 Long Range Dependence

In an LRD time series, the first and second order properties do not deteriorate when the series is viewed on different time scales. This phenomenon of scale invariance has been popularised in the study of visually appealing fractals - geometric shapes which, when magnified,

display the same or similar patterns as the unmagnified shapes. This section provides the reader with a summary of the mathematical underpinnings of LRD stochastic processes and some previous results in the modelling of LRD network traffic.

2.1 Definitions

Let $X_t = Y_t - Y_{t-1}$ represent a zero-mean, wide-sense stationary stochastic process. An LRD process exhibits the property

$$Y_t =_d t^H Y_1, 0 \leq H \leq 1 \tag{1}$$

which means that, in a distributional sense, Y scales with t^H, for some constant H. It is relatively straightforward [2] to show that this implies that the autocorrelation of the increment process X_t at lag k is given by

$$\rho(k) = \frac{1}{2}\Big[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\Big] \tag{2}$$

Additionally, it can be shown that $\lim_{k\to\infty} \rho(k) = H(2H-1)k^{2H-2}$ which implies that, for $0.5 < H \leq 1$, $\sum_{k=0}^{\infty} \rho(k) = \infty$. This latter result indicates that a self-similar process has infinite memory if H is strictly greater than 0.5. For $H = 0.5$, the process is memoryless, i.e., $\rho(k) = 0$ for $k \neq 0$. Many traditional network traffic models rely on the assumption of memorylessness, due to the analytical tractability of Poisson models. Models, which utilise a limited amount of memory, such as Markov, auto regressive and moving average processes cannot capture the asymptotic properties of infinite memory processes [3].

2.2 LRD Stochastic Processes

LRD has been observed in many physical systems, from rainfall patterns to coastline dimensions. However, there are few analytical techniques, which can model the behaviour of real-world LRD systems. Two theoretical models, which exhibit LRD, are Fractional Gaussian Noise (FGN) [4] and Fractional ARIMA(p, d, q) [5], [6] or simply FARIMA. FGN is a generalisation of traditional Gaussian noise,

which allows correlation between successive values. FGN exhibits LRD for $0.5 < H \leq 1.0$ but does not incorporate short-range dependence (SRD)[1]. The ARIMA(p, d, q) family of processes [7] are popular SRD times-series modelling tools, where p is the order of the autoregressive component, q is the order of the moving average component and d is the degree of differencing. Fractional ARIMA(p, d, q) processes generalise this family, and for $0 < d \leq 1/2$ are LRD as well as SRD.

2.3 Empirical Estimators

From the previous, we conclude that H is a compact representation of the LRD of a series. Then, given a series, how do we estimate H? In general, this is a troublesome and often difficult task. The empirical series should be asymptotically wide-sense stationary – that is, the mean and variance are time invariant. The series should also be large, preferably a few hundred thousand observations or more. Unfortunately, long-term measurements of real traffic tend to be non-stationary due to daily patterns of use. Several methods of estimating H have been evaluated in [8]. We've chosen four estimators: variance-time plot, R/S analysis, periodogram analysis and Whittle's estimator (See [2] for a detailed discussion). These tools, along with careful analysis, are shown to produce reasonable estimates of H. We provide a brief discussion of each estimator below.

2.3.1 Variance -Time Plot

Consider the sample mean of n observations of increment process X.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(Y_n - Y_0) \tag{3}$$

From Equation 1, $\bar{X} = n^{H-1}(Y_1 - Y_0) = n^{H-1}X_1$. Therefore $Var\left[\bar{X}\right] = n^{2H-2}\sigma^2$, where σ^2 is the variance of X. Taking the logarithm of both sides gives us

[1] A short-range dependent process exhibits a summable autocorrelation function, which implies that observations sufficiently distant in time are uncorrelated.

$$\log(Var\lfloor \bar{X} \rfloor) = (2H-2)\log(n) + 2\log(\sigma) \approx (2H-2)\log(n) \qquad (4)$$

For a given time series, we can take a number of non-overlapping subseries of length n, calculate the sample mean of these series and the resulting variance of the sample means. As n increases, this variance should decrease proportionally to Equation 4. Plotting $\log(Var\,[\bar{X}])$ versus $\log(n)$ allows us to approximate $2H-2$ (and therefore H) with a least-squares regression. In practice, we find that for small n, points on the variance-time plot are influenced by short-range dependence.

Likewise, for large n, the points may be noisy if too few subseries are used. Therefore, we discard the very low and very high ends of the plot and fit the regression line through the asymptotic slope of the remaining points.

2.3.2 R/S Analysis

The *adjusted range* of process Y is defined to be

$$R(t,k) = \max_{0\le i\le k}\left[Y_{t+i} - \frac{1}{k}Y_{t+k}\right] - \min_{0\le i\le k}\left[Y_{t+i} - \frac{i}{k}Y_{t+k}\right] \qquad (5)$$

Intuitively, $R(t,k)$ estimates the variability of the process over k units of time by relating the amount of time that has passed (i) with the range of values that the process takes on. In order to compare the adjusted range of two or more processes, which differ in magnitude, we define the rescaled adjusted range of Y to be $R/S = R(t,k)/S(t,k)$ where $S(t,k)$ is the sample standard deviation of $X_t, \ldots, X_{t+k}$. It has been shown (see [2] for details) that if X_t^2 is stationary and Y exhibits long-range dependence, then

$$\lim_{k\to\infty} R/S \approx k^H \qquad (6)$$

Therefore, plotting $\log(R/S)$ versus $\log(k)$ for reasonably large k allows us to estimate H. Similar to the case of the variance-time plot,

we discard the very low and very high ends of the plot due to short-range dependence and too few samples, respectively, and perform the regression on the asymptotic slope of the remaining points.

2.3.3 Periodogram Analysis

Based on the discrete Fourier transform, the periodogram is an estimate of the power spectral density of a discrete process. For a time series of n observations, the mean squared amplitude of frequency λ is estimated by

$$I(\lambda) = \frac{1}{n}\left|\sum_{j=1}^{n}(X_j - \bar{X})e^{2\pi i\lambda/n}\right|^2 \tag{7}$$

For an LRD process

$$I(\lambda) \approx \frac{1}{\lambda^H} = \lambda^{1-H} \tag{8}$$

for small λ (this asymptotic identity can, with some effort, be proven from Equation 2), and H can be estimated by a least squares regression on $\log(I(\lambda))$ versus $\log(\lambda)$. Since Equation 8 holds only for low frequencies, it is recommended [8] that the regression line be fit to the lower 10% of the frequencies in the spectrum. In practice, we have found that while the periodogram estimator produces reasonable estimates of H for synthetically generated theoretical processes (such as FGN or FARIMA), for real data (which will almost always be messier) it tends to produce higher estimates of H than other estimators. Also, the periodogram technique is certainly not robust on smaller sets (a few hundred samples) of such real data, as in many cases it estimates H to be greater than 1.0 when other estimators report $H < 1.0$.

2.3.4 Whittle's Estimator

While these heuristic estimators are useful in "eyeballing" LRD, they are not statistically rigorous nor do they provide confidence intervals. The Whittle estimator is an asymptotic estimate of H, and provides a

confidence interval, but also requires the empirical series to be a Gaussian process and that the underlying form of the series be provided. This form can be based on Fractional Gaussian Noise [4], which has long-range, but no short-range dependence, or a Fractional ARIMA process [5], which has both long-range and short-range dependence. Using the Whittle estimator, we find the value of z which minimises the function

$$W(z) = \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda; z)} d\lambda \tag{9}$$

where $I(\lambda)$ is the periodogram and $f(\lambda; z)$ is the spectral density at frequency λ. For FGN, $z = H$, while for FARIMA, z is a vector containing the autoregressive, moving average and back-shift components of the process. For a detailed discussion, see chapter 5 of [2]. These tools have been shown to produce reasonable estimates of H; however, one must be careful during their use and when interpreting their results. In particular, caution is necessary when using the Whittle estimator. While it provides an estimate of H and a confidence interval, it also requires the empirical series to be a Gaussian process and that the underlying form of the series be provided.

This form can be based on any LRD process, usually FGN or FARIMA. In order to use the Whittle estimator on processes that may be non-Gaussian, we note that a long-range dependent process asymptotically converges to Fractional Gaussian Noise when aggregated. Thus, we can aggregate the original series in non-overlapping blocks of size m

$$X_i^{(m)} = \frac{1}{m}(X_{tm-m+1} + X_{tm-m+2} + \ldots + X_{tm}) \tag{10}$$

We then determine the Whittle estimator of the series for various $X^{(m)}$. As m increases, the Whittle estimator should converge to a robust estimate of H, but if $X^{(m)}$ contains fewer than about 100 points, estimates become noisy and tend to have very large (and therefore less meaningful) confidence intervals.

Table 1. Details of Internet delay experiments: date, starting time, means, standard deviation, median, inter-quartile range, and loss rate. All delay quantities are in millisecond. A date annotated with $a*$ indicates that the experiment was performed either on a weekend or early morning, times at which Internet load is usually light.

Trace	Date	Start	Mean	Stdev	Median	IQR	Loss
UIC – 1	5/24/97*	9:3 AM	20. 28	20. 28	9. 42	10. 28	0. 34 %
UIC – 2	5/31/97*	9:3 AM	22. 14	25. 33	10. 69	15. 06	0. 28 %
UIC – 3	6/2/97	1:09 PM	27. 25	24. 39	17. 75	23. 28	0. 46 %
UIC – 4	6/17/97	7:40 PM	53. 69	39.02	44. 90	57. 81	0. 67 %
UCD – 1	6/18/97	8:46 PM	105. 89	33. 02	89. 76	39. 25	1. 70 %
UCD – 2	6/19/97*	1:00 AM	83. 59	15. 23	80. 30	4.39	0. 24 %
UCD – 3	6/23/97	9:38 AM	93. 56	21. 55	87. 26	12. 33	1. 74 %
UCD – 4	6/23/97	5:55 PM	91. 27	19. 70	86. 78	11. 09	0. 49 %
SIT – 1	6/24/97	1:31 PM	75. 05	28. 77	64. 81	29. 94	1. 41 %
SIT – 2	6/26/97	10:50 AM	94. 96	36. 66	84. 59	61. 84	3. 84 %
SIT – 3	6/27/97	1:38 PM	91. 38	36. 72	81. 78	54. 54	2. 52 %
SIT – 4	6/29/97*	5:06 PM	75. 10	34. 27	59. 86	37. 01	1. 14 %

3 Measurements

3.1 Software

We have developed a set of tools to measure packet delay in the Internet. Our goal was to be able to measure the delay that a real-time, application-layer service would experience. Thus we wrote a distributed application layer program consisting of a client and a server to be run on two different hosts. The client transmits a series of UDP packets to the server. The application allows the user to specify the interval between transmissions by the client, and the number of bytes in each application-layer packet. Each packet contains a sequence number (for packet loss measurements) and is time-stamped at the moment that

the data leaves the application layer (t_1). When the server receives a packet, it adds a time-stamp (t_2) to the data field, and then retransmits the packet back to the client. Upon reception of the retransmitted packet, the client adds a third time-stamp (t_3) and stores the packet. Thus, round-trip delay is measured by ($t_3 - t_1$).

3.2 Experiments

Table 1 describes measurements of Internet delay using our software. We will refer to each measurement session as a *trace*. Each trace consists of 64K delay measurements using 80-byte UDP packets with a fixed 30 ms inter-departure time (thus, each trace lasted about 33 minutes).

These parameters are similar[2] to those of the ITU recommendation G.723.1 for the encoding and segmentation of a digital voice stream over packet-switched networks [9]. The transmitting host was always at DePaul University in Chicago. The server hosts were University of Illinois, Chicago (UIC), University of California, Davis (UCD) and Stevens Institute of Technology (SIT) in Hoboken, New Jersey. The routes from DePaul to UCD and SIT both traverse one or more national backbone providers. The route from DePaul to UIC uses a single service provider, CICNET. Thus we are able to analyse the delays exhibited over both wide-area and metropolitan networks.

First and second order characteristics of our traces are also shown in Table 1. The results agree quite well with previous research [10]. The traces to UIC exhibit a much smaller mean and median delay than the traces to UCD and SIT, presumably due to the close topological distance between DePaul and UIC.

However, surprisingly, the UCD traces exhibit the lowest standard deviation and interquartile range (IQR)[3], while the SIT traces exhibit the highest values of these two statistics. The extreme variability of the

[2]The G.723.1 recommendation specifies a 20-byte or 24-byte payload with 30 ms inter-departure times.

[3]The inter-quartile range of a data set is the difference between the values at the 75th and 25th percentiles.

delays is most evident for the UIC traces, for which the IQR exceeded the median delay four out of five times. Due to the low delays over the UIC route, relatively minor changes in traffic patterns or system loads at the endpoints may have a proportionally large impact on the second-order statistics of delay.

4 Estimation of LRD

Some preparation of our data sets was necessary before the techniques discussed in Section 2.3 could be used to estimate H.

First, we performed a log transform of the raw delay values, which adjusts the marginal distribution of the data so that it is closer to Gaussian (see Section 2.3), while preserving H [2]. We then checked the data for a linear trend (there were none that were significant). Finally, we visually examined a plot of the data for any obvious non-stationarities, such as long plateaus of very high delays (again, there were none)[4].

However, as discussed in Section 4.4, more rigorous analysis later indicated that several of the traces are not well modelled with a constant H, which indicates non-stationarity in terms of the LRD structure of these traces.

Tables 2-4 contain estimates of H for our traces using the four empirical estimators. Whittle estimates include aggregation and 95% confidence intervals and were performed using both FGN and FARIMA. An entry of "NC" indicates that the estimator used in that case did not converge. We will analyse the results of each type of estimator separately below.

4.1 Heuristic Estimates

The variance-time, R/S and periodogram estimates were found by following the procedures discussed in Section 2.3.

[4] In general, determining whether a trace is stationary or not, is not always possible. In some cases, simple "eyeball" tests will suffice when there is an obvious non-stationarity. See Chapter 7 of [2] for details.

Table 2. Estimates of H for UIC traces.

Estimator (m)	UIC-1	UIC-2	UIC-3	UIC-4
Variance-time	0.7137	0.8823	0.8619	0.6595
R/S	0.7503	0.8770	0.8595	0.6578
Periodogram	0.9208	0.9560	0.9179	0.9524
Wh. FGN	0.9059 ± 0.0051	0.9095 ± 0.0052	0.8711 ± 0.0052	0.9990 ± 0.0051
Wh. FGN (100)	0.7475 ± 0.0507	0.8572 ± 0.0514	0.9376 ± 0.5129	0.5571 ± 0.0486
Wh. FGN (200)	0.7224 ± 0.0715	0.8812 ± 0.0730	0.9468 ± 0.0734	0.5928 ± 0.0709
Wh. FGN (300)	0.7279 ± 0.0876	0.9358 ± 0.0898	0.9074 ± 0.0897	0.6374 ± 0.0861
Wh. FGN (400)	0.7259 ± 0.1013	0.9573 ± 0.1042	0.9194 ± 0.1037	0.7106 ± 0.1010
Wh. FGN (500)	0.6979 ±0.1124	0.9770 ± 0.1164	0.8885 ± 0.1154	0.7604 ± 0.1136
Wh. FGN (600)	0.7382 ± 0.1242	0.9595 ± 0.1274	0.8805 ± 0.1265	0.8139 ± 0.1255
Wh. FARIMA (0,d,0)	0.9931	0.9892	0.9459	1.1188
Wh. FARIMA (1,d,1)	0.9132	0.9304	0.9850	1.0171
Wh. FARIMA (2,d,2)	0.9844	NC	NC	NC

Table 3. Estimates of H for UCD traces.

Estimator (m)	UCD-1	UCD-2	UCD-3	UCD-4
Variance-time	0.7014	0.7417	0.8642	0.6972
R/S	0.8448	0.7535	0.7985	0.7416
Periodogram	0.9083	0.7900	0.8784	0.8143
Wh. FGN	0.9994 ± 0.0051	0.8189 ± 0.0051	0.8782 ± 0.0052	0.7986 ±0.0051
Wh. FGN (100)	0.8957 ± 0.0516	0.7192 ± 0.0505	0.8767 ± 0.0515	0.6967 ± 0.0503
Wh. FGB (200)	0.7627 ± 0.0719	0.7436 ± 0.0717	0.9156 ± 0.0732	0.6854 ± 0.0710
Wh. FGN (300)	0.7239 ± 0.0876	0.7311 ± 0.0876	0.9582 ± 0.0901	0.6708 ± 0.0867
Wh. FGN (400)	0.7359 ± 0.1015	0.7431 ± 0.1016	0.9504 ± 0.1040	0.6602 ± 0.1001
Wh. FGN (500)	0.6857 ± 0.1122	0.7358 ± 0.1133	0.9803 ± 0.1164	0.6559 ± 0.1119
Wh. FGN (600)	0.7263 ± 0.1239	0.7997 ± 0.1252	0.9696 ± 0.1275	0.6525 ± 0.1222
Wh. FARIMA (0,d,0)	1.1181	0.8904	0.9568	0.8623
Wh. FARIMA (1,d,1)	0.9192	0.9145	0.9530	0.8931
Wh. FARIMA (2,d, 2)	NC	0.8060	0.8618	0.7889

Table 4. Estimates of H for SIT traces.

Estimator (m)	SIT-1	SIT-2	SIT-3	SIT-4
Variance-time	0.7909	0.7617	0.7469	0.8033
R/S	0.8271	0.7504	0.7501	0.9108
Periodogram	0.9333	0.8764	0.8787	0.9646
Wh. FGN	0.9959 ± 0.0052	0.9475 ± 0.0052	0.9996 ± 0.0051	0.9816 ± 0.0051
Wh. FGN (100)	0.8637 ± 0.0515	0.7675 ± 0.0508	0.7318 ± 0.0506	0.8436 ± 0.0514
Wh. FGN (200)	0.8174 ± 0.0725	0.7492 ± 0.0717	0.7842 ± 0.0722	0.8761 ± 0.0729
Wh. FGN (300)	0.8586 ± 0.0892	0.7670 ± 0.0881	0.8228 ± 0.0889	0.8382 ± 0.0890
Wh. FGN (400)	0.8708 ± 0.1003	0.7429 ± 0.1016	0.8233 ± 0.1027	0.8310 ± 0.1028
Wh. FGN (500)	0.8631 ± 0.1151	0.8289 ± 0.1147	0.8323 ± 0.1147	0.7802 ± 0.1140
Wh. FGN (600)	0.8690 ± 0.1263	0.7726 ± 0.1248	0.8878 ± 0.1266	0.8223 ± 0.1256
Wh. FARIMA (0.d,0)	1.0955	1.0395	1.1286	1.0769
Wh. FARIMA (1,d,1)	0.9914	0.9116	0.8970	0.9876
Wh. FARIMA (2,d,2)	0.9982	0.7988	0.7426	NC

In general, we find that the variance-time and R/S techniques agree quite well. These estimators were within 0.01 of one another for half of the traces. However, the periodogram estimates were 0.03-0.30 greater than those given by the variance-time and R/S procedures. The difficulty with the periodogram seems to be the rather arbitrary choice in fitting the regression line through the lowest 10% of the frequencies. Although this '10% rule' has been used often (see [3] and [8], among others), we noted that the spectral densities of our traces were often non-linear on a log-log plot, which increases the sensitivity of the estimator to the number of frequencies used.

4.2 FGN Whittle Estimates

Using the Whittle estimator requires some care. In Section 2.3 we noted that the FGN Whittle estimator assumes that the data it operates upon has a Gaussian marginal distribution. In practice we find that many real-world data sets are non-Gaussian. In order to derive accurate point estimates of H with the FGN Whittle estimator, we aggregate our traces for levels of $m = 100$, 200, 300, 400, 500, 600. At $m = 600$ we discontinue aggregation because the number of points in $X^{(600)}$ is about 100. Point estimates of H are found by following the heuristic described in [1]; that is, as m increases, the Whittle estimator may initially fluctuate, but will eventually converge. The earliest point of this convergence and its associated confidence interval are used as point estimates. From Tables 2-4, we note that the estimates of H produced by running the FGN Whittle estimator on the unaggregated traces are, in most cases, significantly different from the value that the Whittle estimator converges to when the traces are aggregated. For UIC-4, the Whittle estimator did not converge. For $m = 100$ it indicates that the trace exhibits little, if any, LRD. However, as m increases, so do the estimates of H. Further analysis (see Section 4.4) indicates that the trace is well modelled with $H \approx 0.55$. As of the time of this writing, we have no explanation for the anomalous behaviour of this trace. By observation, we have found that that marginal distribution of Internet delays exhibits a light lower tail, but heavy upper tail.

For $m = 1$, the marginal distribution of delay in the traces is quite far from Gaussian. As m increases, the marginal distribution approaches a Gaussian distribution. Naturally, we should make sure that the distribution is "close enough" to Gaussian so that the Whittle estimator is accurate. For UCD-2, the aggregated Whittle estimator converges to $H \approx 0.74$ while the variance-time and R/S estimates are $H = 0.7417$ and $H = 0.7535$, respectively. Although the marginal distribution is not exactly Gaussian it is "close enough" in the case of UCD-2 to agree with estimators that do not require Gaussian marginal distributions. Since the marginal distributions of the rest of our traces at the level of aggregation chosen for the point estimate of H are no less Gaussian (as visually determined by a quantile-quantile plot) than UCD-2 for $m = 300$, we conclude that the Whittle estimator is a reasonable estimator of H for our traces, despite their deviations from a purely Gaussian distribution.

Table 5. FGN Whittle point estimates with 95% confidence intervals compared to heuristic estimators.

Trace	FGN Whittle & CI	Var-time	R/S	Periodogram
UIC-1	0.7224 ± 0.0715	Yes	Yes	No
UIC-2	0.9358 ± 0.0898	Yes	Yes	Yes
UIC-3	0.9074 ± 0.0897	Yes	Yes	Yes
UIC-4	0.5571 ± 0.0486	No	No	No
UCD-1	0.7239 ± 0.0876	Yes	No	No
UCD-2	0.7436 ± 0.0717	Yes	Yes	Yes
UCD-3	0.9582 ±. 0901	No	No	Yes
UCD-4	0.6708 ± 0.0867	Yes	Yes	No
SIT-1	0.8586 ± 0.0892	Yes	Yes	Yes
SIT-2	0.7492 ± 0.0717	Yes	Yes	No
SIT-3	0.8228 ± 0.0889	Yes	Yes	Yes
SIT-4	0.8382 ± 0.0890	Yes	Yes	No

Table 5 compares point estimates of H using the FGN Whittle estimator to estimates resulting from the heuristic estimators. The Whittle estimates are shown with 95% confidence intervals. For each

heuristic estimate, we indicate (with "Yes" or "No") whether or not it fell within the Whittle estimator's confidence intervals. The variance-time estimator fell within the confidence interval for 10 of 12 traces while the R/S estimator fell within the confidence interval for 9 of 12 traces. The periodogram estimator only fell within the confidence intervals for 6 of 12 traces. Given these results, we recommend that if the heuristic estimators produce estimates of H which are outside of the 95% confidence interval of Whittle's estimator, the data should be more carefully analysed for non-stationarities. This is pursued further in Section 4.4.

4.3 FARIMA Whittle Estimates

The FARIMA Whittle estimates of Tables 2-4, seem to indicate that Internet delay will not easily fit a FARIMA model. In particular, we find that many of our $(0, d, 0)$ estimates are greater than 1.0 and all but 2 are greater than 0.9. Similar, though less extreme, results were obtained for $(1, d, 1)$ estimates. These results are not surprising given that the marginal distributions of the traces were non-Gaussian. FARIMA $(2, d, 2)$ models converged for only 7 out of the 12 traces, and 5 of these were within, or at least very close to, the 95% confidence interval of the FGN Whittle estimator. We also attempted to model the traces using $(1, d, 2)$ and $(2, d, 1)$ processes (not shown), but the H estimates were always very close to those of FARIMA $(1, d, 1)$. Although FARIMA processes are more powerful modelling tools than FGN because they contain SRD as well as LRD, our experience indicates that they tend to provide aggressive estimates of H. We may not have found the optimal combination of aggregation and parameters, but we know no better technique to find such a combination other than exhaustive search.

4.4 Testing for Stationary H

For some traces, the variation of the H estimates is quite large. In [12] it is shown that for stationary time series, $\bar{H}^{(100)}$ has the same asymptotic distribution as $H^{(100)}$; thus, in cases where the entire trace is well-modelled with a constant H, we expect that $\bar{H}^{(100)} \approx H^{(100)}$.

Given the anomalous case of UIC-4 and the high (> 0.9) Whittle estimates for UIC-2, UIC-3, and UCD-3, we turn our attention towards determining how the estimates of H for an entire trace compare to estimates of H for disjoint subsets of that trace. We divide $X^{(100)}$ for each trace into five subsets of 128 points each, and run the FGN Whittle estimator on each subset.

Table 6 shows $H_i^{(100)}, 1 \leq i \leq 5$ for each subset, along with $\bar{H}^{(100)} = \frac{1}{5}\sum_{i=1}^{5} H_i^{(100)}$ and the estimate of H from the entire trace.

Table 6. FGN Whittle estimates for 5 disjoint subsets of 128 points, for $X^{(100)}$.

Trace	$H_1^{(100)}$	$H_2^{(100)}$	$H_3^{(100)}$	$H_4^{(100)}$	$H_5^{(100)}$	$H^{(100)}$	H
UIC-1	0.7954	0.8027	0.76661	0.6457	0.6731	0.7366	0.7475
UIC-2	0.5642	0.6471	0.6461	0.6164	0.9107	0.6769	0.8572
UIC-3	0.7129	0.9411	0.7297	0.8877	0.9702	0.8483	0.9376
UIC-4	0.5620	0.5000	0.5982	0.5465	0.5975	0.5608	0.5571
UCD-1	0.9350	0.8158	0.8560	0.8973	0.8480	0.8704	0.8957
UCD-2	0.7398	0.7396	0.7205	0.5446	0.5217	0.6532	0.7192
UCD-3	0.8064	0.7552	0.5852	0.8309	0.9130	0.7782	0.8767
UCD-4	0.6705	0.7252	0.7206	0.7364	0.6248	0.6955	0.6967
SIT-1	0.9017	0.8582	0.8674	0.8180	0.8529	0.8596	0.8637
SIT-2	0.7321	0.7522	0.8831	0.7585	0.6632	0.7578	0.7675
SIT-3	0.6784	0.7322	0.6479	0.7392	0.7072	0.7072	0.7318
SIT-4	0.8211	0.8548	0.9114	0.8285	0.6427	0.8117	0.8436

Considering Table 6, we find that in cases where the $H_i^{(100)}$'s do not fluctuate dramatically, this property holds quite well. In particular, we conclude that UIC-1, UIC-4, UCD-4, SIT-1, and SIT-2 are reasonably well modelled with constant H as determined by the FGN Whittle estimator on the entire trace. The other seven traces are not well modelled with constant H, and UIC-2, UIC-3, and UCD-3 are particularly pathological. Also note that these three traces all exhibit H estimates greater than 0.9. For the seven traces with significantly non-homogeneous H, the Whittle estimator over the entire trace gives us an overly generous estimate of H, possibly due to these non-stationarities. Regardless of the test used, it is clear that in some cases H may change dramati-

cally over relatively small (5 minute) periods of time, while in others H may be well-modelled as constant for 15 minutes or more. While this result suggests that our traces may contain non-stationarities, it also indicates that the appropriate time scale over which to analyse network delays is not clear. Our analysis tools are most robust when analysing stationary data, and by reducing the time scale of our analysis (analysing smaller time periods) our data may exhibit stationarity. However, current LRD estimators require a relatively large number of samples, and our sampling rate is limited by hardware clock resolution.

4.5 Network Load and H

Previous research regarding LRD network traffic has concluded that H is correlated with network load. For example, in [3], Ethernet traffic is found to have a higher H during busy periods, and in [13], VBR video is found to exhibit higher H during high-activity scenes. Given these results, we had expected to find a clear correlation between the magnitude of delay for a trace (as an indicator of network load) and H. However, when comparing the mean and median delay of each trace from Table 1 to the point estimates of H from Table 6, we found that, in both cases, H was negatively correlated with delay (-0.14 and -0.15, respectively)! Packet loss rate was positively correlated with H, but not significantly so (0.06). We also compared the $\bar{H}^{(100)}$ values from Table 5.

We found that for these values, mean and median delays were positively correlated with H (0.16 and 0.10, respectively). However, if we disregard the traces for with H fluctuated the most (UIC-2, UIC-3, UCD-2, UCD-3, SIT-3 and SIT-4), we find that mean and median delays were much more highly correlated with our point estimates of H (0.35 and 0.30, respectively). Thus, we can conclude that the magnitude of H does seem to be correlated with network load for traces with stationary H.

5 Estimator Robustness and Loss

When evaluating an estimator of H on our traces, we ignored all packet loss events. Although our traces do not exhibit packet loss rates greater than 4%, there is evidence that loss rate of 10% - 20% and higher may be common on some Internet paths [14]. It is unclear what effect this packet loss has on the estimation of H. When such censored data is converted into the frequency domain, some spectral leakage would be expected. However, the impact of leakage on periodogram-based estimates of H (including the Whittle estimator) is not certain. A robust analysis must be able to compensate for "gaps" in the data.

Table 7. Estimates of H for FGN ($H = 0.7$) series with varying loss rates.

Estimator	*No loss*	*5% loss*	*10% loss*	*20% loss*	*33% loss*	*50% loss*
Variance-time	0.6700	0.6752	0.6662	0.6659	0.6678	0.6507
R/S	0.7006	0.7035	0.7030	0.6950	0.6876	0.6708
Periodogram	0.6997	0.7032	0.6937	0.6904	0.6813	0.6747
Weighted Periodogram	0.6991	0.6964	0.6947	0.6894	0.6806	0.6636
Whittle	0.7015	0.6965	0.6926	0.6842	0.6706	0.6528

Table 8. Estimates of H for UIC-4 series with varying loss rates.

Estimator	*Unmodified*	*5% loss*	*10% loss*	*20% loss*	*33% loss*	*50% loss*
Variance-time	0.6595	0.6801	0.6832	0.6713	0.6682	0.6712
R/S	0.6578	0.6682	0.6595	0.6437	0.6336	0.6797
Periodogram	0.9524	0.9517	0.9452	0.9345	0.9068	0.8647
Weighted Periodogram	0.5533	0.5506	0.5559	0.5449	0.6029	0.6586
Whittle	0.5571	0.5571	0.5571	0.5571	o.5571	0.5571

We introduce a weighted periodogram, which has been shown to be robust [15] when applied to unevenly sampled data. Consider our

Table 9. Estimates of H for UCD-4 series with varying loss series.

Estimator	*Unmodified*	*5% loss*	*10% loss*	*20% loss*	*33% loss*	*50% loss*
Variance-time	0.6972	0.6574	0.6405	0.6427	0.6538	0.6328
R/S	0.7416	0.6790	0.6768	0.6770	0.6806	0.6646
Periodogram	0.8143	0.8103	0.8205	0.8082	0.7861	0.7087
Weighted Periodogram	0.6276	0.6606	0.6512	0.6378	0.6122	0.6264
Whittle	0.6708	0.6708	0.6708	0.6708	0.6708	0.6708

Table 10. Estimator of H for SIT-1 series with varying loss rates.

Estimator	*Unmodified*	*5% loss*	*10% loss*	*20% loss*	*33% loss*	*50% loss*
Variance-time	0. 7909	0.8024	0.7949	0.7947	0.7942	0.7699
R/S	0.8271	0.8288	0.8254	0.8257	0.8167	0.8745
Periodogram	0.9333	0.9166	0.9106	0.8969	0.8892	0.8764
Weighted Periodogram	0.8612	0.9412	0.9289	0.8975	0.8683	0.8557
Whittle	0.8586	0.8586	0.8586	0.8586	0.8586	0.8586

measurement process, described in Section 3. Although we schedule each packet transmission individually, CPU load at the transmitting host or media-access delays at the network-interface card may cause our transmission to be slightly delayed. And, as discussed above, packet losses put gaps in our delay time series. Our goal is to use this robust periodogram to prove one of two possibilities:

(a) The unweighted periodogram, and possibly some or all of the other estimators, are biased under packet loss and should be used with caution in these circumstances, or

(b) Our current set of LRD estimators do a reasonable job of estimating H when packets are lost, are therefore may be used with confidence in situations where packet loss is significant.

The weighted periodogram is defined as follows for a time series in which the jth sample occurs at time t_j and the sampled value is X_j

$$I_w(\lambda) = \frac{1}{2}\left(\frac{\left[\sum_j X_j \cos\lambda(t_j-\tau)\right]^2}{\sum_j \cos^2\lambda(t_j-\tau)}\right) + \\ + \frac{1}{2}\left(\frac{\left[\sum_j X_j \sin\lambda(t_j-\tau)\right]^2}{\sum_j \sin^2\lambda(t_j-\tau)}\right) \tag{11}$$

where λ is a frequency and

$$\tau = \tan^{-1}\left[\frac{\left(\sum_j \sin 2\lambda t_j\right) / \left(\sum_j \cos 2\lambda t_j\right)}{2\lambda}\right] \tag{12}$$

In order to evaluate the robustness of the LRD estimators under varying levels of packet loss, we synthetically generate series of 64K FGN[5] points with $H = 0.7$. We calculate H using each estimator for packet loss rates of 5% - 50%. Our loss model is very simple. To simulate a loss rate of $0 \leq p \leq 1$, we iterated through an FGN trace and discarded each point with probability p. Thus, losses were modelled as independent Bernoulli trials, and the length of loss bursts were geometrically distributed. Empirical studies of Internet packet loss have not settled on a model for loss characteristics, though there is evidence that UDP packet loss rates are not independent for closely spaced transmissions [17]. However, for our purposes of understanding how the LRD estimators respond to censored data, this model is sufficient.

Table 7 shows the results for synthetic FGN data sets generated with $H = 0.7$. The first column (*No loss*) shows the estimate of H determined by each estimator for the original data set (64K samples). Successive columns show estimates of H generated for the subsets of this data derived by randomly removing 5%, 10%, 20%, 33% and 50% of the data points. As data loss increases, there is a general trend away from the lossless H value, as might be expected. However, all estimators are reasonably robust, with the maximum variation being only 8% between the 0% and 50% loss estimates across all estimators.

[5]To generate FGN, we used the *fgn* tool presented in [16], which is available at http://www.acm.org/sigcomm/ITA/.

Although this is an encouraging result, we must keep in mind that FGN is an idealised process. We expect real network traffic to be messy. Tables 8- 10 show the same experiments performed on three of the more well-behaved traces (in terms of stationarity of H): UIC-4, UCD-4 and SIT-1. Since the original traces already exhibited a small amount of packet loss, we do not have a lossless control case to compare against. Also, since we tested the unaggregated traces, the Whittle estimator could not be trusted to produce an accurate result (see Section 4.2). Instead, we use the Whittle point estimate for each trace from Table 5.

For all three traces, the robustness of the variance-time estimator degraded somewhat with loss, and the R/S and the two periodogram estimators produced noisy, but reasonable estimates of H, even when 50% of the packets were lost. Thus, we will conclude that the LRD estimators are reasonably robust under loss. An interesting trend that can be seen in Tables 9 and 10 is that as the loss rate increases, the unweighted periodogram produces estimates of H closer to the Whittle point estimate. We hypothesize that our loss model preserved most of the low-frequency components of the traces, while perhaps eliminating many of the high-frequency components.

Finally, we note that for the unmodified UIC-4 trace, the weighted periodogram agrees very closely with the Whittle point estimate, while the other heuristic estimators do not. Recall that in Section 4.2 we noted that the Whittle estimator did not converge for this trace, though further analysis (see Section 4.4) indicated that the trace was well modelled with constant $H \approx 0.55$.

6 Wavelet Analysis

Wavelet analysis is a tool for simultaneously measuring the spectral and temporal components of a signal or time series. Not unlike Fourier analysis, wavelet analysis involves transforms. The continuous wavelet transform of a signal $x(t)$ is given by

$$C(x,\psi)(\tau,s) = \frac{1}{\sqrt{s}}\tau, s\left(\frac{t-\tau}{s}\right)dt \tag{13}$$

where τ is called the translation of the signal and s is called the *scale* of the signal. The function ψ is called the basis function or mother wavelet, and must be in the function space L^2 (L^2 contains all functions $f(t)$ that are square integrable; i.e., $\int f^2(t)dt < \infty$). Computing $C(x,\psi)(\tau,s)$ for particular values of τ and s is equivalent to applying ψ, shifted in time by τ and scaled in size by s, to the signal. For a constant s, a wavelet transform is not unlike a short-term Fourier transform, where ψ is the window function. As s increases, the size (in time) of ψ grows. Thus, for small and large s, high frequency and low frequency components of the signal can be efficiently identified, respectively.

The continuous wavelet transform produces an overcomplete planar representation of the signal. Let n be the number of coefficients of the transform when $s = 1$. By sampling the coefficients of the continuous wavelet transform at the points

$$D_{(x,\psi)}(i,j) = C_{(x,\psi)}\left(i2^j, 2^j\right) \quad 0 \le j < [\log_2 n], \quad 0 \le i < n/2^j \tag{14}$$

on this plane, we obtain a discrete representation of the coefficients without losing information [18]. This representation, called the discrete wavelet transform, allows us to determine a finite, yet complete, set of coefficients of the wavelet transform. This approach analyses $x(t)$ by logarithmically scaling ψ based on the octaves j.

6.1 A Discrete Wavelet Estimator

A technique for estimating H with the discrete wavelet transform was discussed in [19]. The mean-square power around octave j is given by

$$P_j = \frac{1}{n2^j}\left|D_{(x,\psi)}(i,j)\right|^2 \tag{15}$$

Thus, given that the power around frequency λ of an LRD process scales with the inverse of λ raised to the H (see Section 2.3.3), we

can estimate H by plotting j versus $\log_2 P_j$ and taking a least-squares regression through the resulting points. Since there are logarithmically fewer coefficients in each successive octave, we weight the point at octave j by the number of coefficients in this octave (i.e., $n/2^j$) when performing the regression. The slope, α, of the fitted line is the basis of the estimate of H; that is, $H = (\alpha+1)/2$. In [19] this estimator is shown to be asymptotically unbiased. Furthermore, it does not suffer from the limitation of the Whittle estimator; that is, the data set being analysed does not have to be Gaussian. Thus, aggregation is not necessary. Finally, it is possible to infer the strength of the dependence structure between different frequency ranges.

This allows us to compare the strength of both long and short-range dependence.

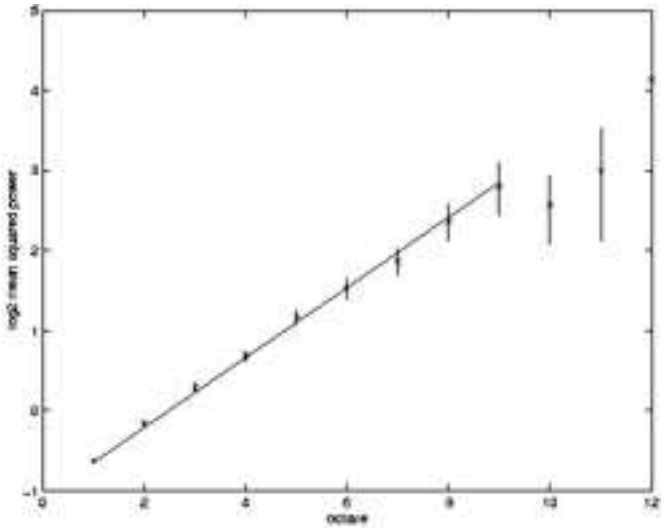

Figure 1. Wavelet estimator applied to FGN with $H = 0.7$

An example of the wavelet estimator[6], is shown in Figure 1 for a synthetically generated set of 64K observations of FGN with $H =$

[6]Throughout this paper we used the wavelet Daubechies-3 [20].

0.7. The logarithmically-transformed mean-squared power ($\log_2 P_j$) is plotted for each octave (j) with confidence intervals[7]. The weighted regression is taken over octaves 1 to 9. In general, we find that as j increases, the power estimates become noisy due to increasingly smaller sets of coefficients. When performing the regression, we only consider octaves for which the fitted line intersects the confidence intervals. The resulting estimate of $H = 0.71$ is consistent throughout both low (large j) and high (small j) frequencies.

6.2 Estimation of LRD and SRD of Network Delay

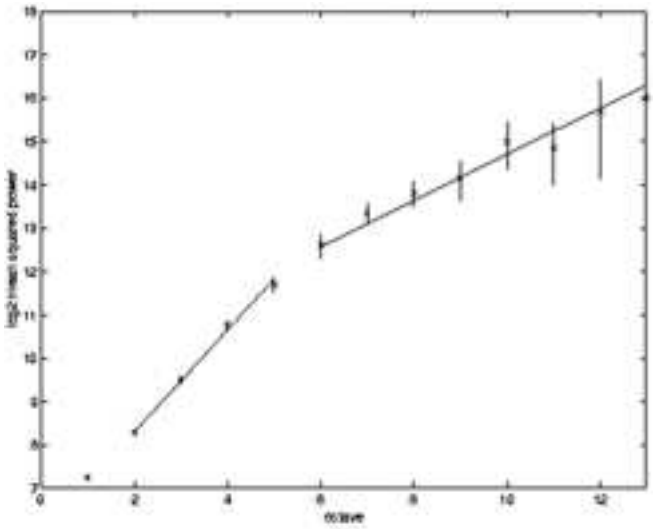

Figure 2. The Wavelet estimator applied to SIT-2.

We used the wavelet estimator on all 12 of our delay traces. We found that none of our traces are well modelled with H consistent throughout all frequency ranges. In particular, 10 out of 12 traces exhibited two distinct octave ranges (scalings) from which different

[7]The upper and lower confidence intervals are not always equidistant due to the logarithm scaling of the graphs. These confidence intervals are computed per [19].

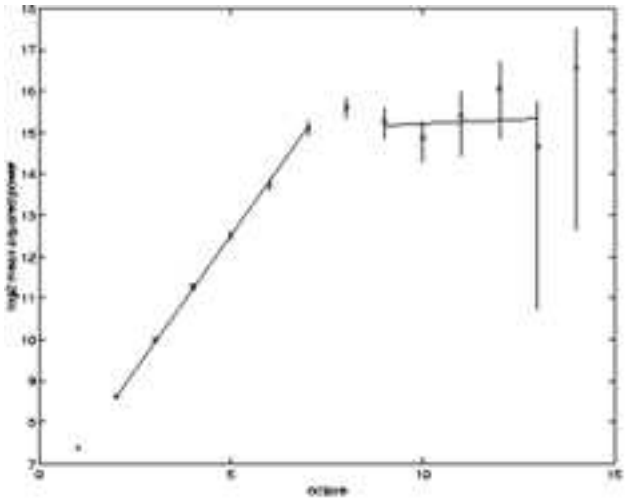

Figure 3. The Wavelet estimator applied to UIC-4

values of H were estimated, while the remaining 2 traces exhibited three such scalings. Similar non-linearity's have been seen in wavelet-based analyses of traffic volume [21].

Figure 2 shows the wavelet estimator applied to trace SIT-2. Two regressions are performed: for $j = 2 \ldots 5$ and $j = 6 \ldots 13$. For low frequencies ($j = 6 \ldots 13$), H is estimated to be 0.7664, which is consistent with the Whittle estimate of 0.7492. For high frequencies $j = (2 \ldots 5)$, H is estimated to be 1.0881. At these high frequencies (low octaves) the H estimate no longer indicates LRD, but does imply a more localised dependency structure.

The wavelet estimator also provides a visual explanation for some of the difficulties we experienced when using the Whittle estimator. In Section 4.4 we found that the Whittle estimate of H for trace UIC-4 varied dramatically with the level of aggregation. Figure 3 shows the wavelet estimator applied to trace UIC-4. As the octaves that we examine increase, we find that the H estimate is high (1.1333) for $j = 2 \ldots 7$, but low (0.5208) for $j = 9 \ldots 13$. For even greater octaves, the H estimate increases again. Recall that the Whittle estimator

Table 11. Wavelet estimates of H for various scalings.

Trace	Scalings	$\boldsymbol{H}$	In Whittle CI?
UIC-1	(1, 5) (6, 13)	1.0272 0.7799	yes
UIC-2	(1, 3) (4, 12)	1.1009 0.8791	yes
UIC-3	(3, 5) (6,13)	1.1548 0.8256	yes
UIC-4	(2, 7) (9,13)	1.1333 0.5208	yes
UCD-1	(3, 5) (6, 8) (9,13)	1.0137 0.9849 0.6448	yes
UCD-2	(1, 3) (4, 12)	0.8064 0.7593	yes
UCD-3	(1, 4) (8,13)	1.0059 0.9389	yes
UCD-4	(2, 6) (7, 13)	0.9565 0.7045	yes
SIT-1	(2, 4) (5, 7) (8,13)	1.0821 0.8747 0.7819	yes
SIT-2	(2, 5) (6, 13)	1.0881 0.7664	yes
SIT-3	(2, 5) (6, 11)	1.1731 0.7046	no
SIT-4	(3, 6) (7,10)	1.0358 0.8728	yes

is periodogram-based. The frequency decomposition produced by the wavelet transform gives us an intuitive explanation of why the Whittle estimator behaves the way that it does. Loosely speaking, a spectral LRD estimator expects that the spectral power of the observed process increases logarithmically with octave. Trace UIC-4 exhibits a flatness of spectral power localised to a few particular octaves that causes the periodogram-based estimators to produce inconclusive results.

Table 11 shows wavelet estimates of H for each scaling of our 12 traces. We found that the low frequency scalings exhibited H estimates within the confidence interval of the Whittle estimator in 11 of the traces. For all traces, the high frequency scalings exhibited a stronger dependence structure than the low frequency scalings.

This latter result implies that characterising the dependence of network traffic with just the H parameter may be misleading. A more promising approach is to incorporate the H estimates for each scaling, thus preserving both the long range and short range dependence of the observed traffic.

Such a characterisation can also be used to synthetically generate realistic traffic of network delays. Given the power in each octave of the discrete wavelet transform, we can use an inverse wavelet transform [18] to construct a time series with a similar long range and short range dependence structure.

As in Section 4.4, we found that wavelet estimates of H change over the length of the observed time series. Recent research in [22] introduces a wavelet-based method for characterising the dynamics of LRD. As of the time of this writing we are still examining this method.

7 Conclusions

The main contribution of this paper is the result that in the majority of the cases that we have studied, there is strong statistical evidence that Internet packet delay is long-range dependent.

However the degree of this dependence may change dramatically over relatively small periods of time, which indicates that the appropriate time scale over which to analyse Internet delay is not clear.

Stochastic modelling of delay, as LRD is a non-trivial task and the tools at our command may not be sufficient for describing our observations. For example, we do not know whether fluctuations of H imply that a process is truly changing its character. We found that with the help of wavelet analysis, we can estimate both the long range and short range dependence structure of empirical network delay traces. By examining the power of different frequency octaves, we found that periodogram-based LRD estimators can produce confusing results when the observed power spectrum is not well modelled as logarithmically decaying with frequency. As a result we advocate the use of the wavelet estimator and wavelet analysis in general as an important tool for measurement and analysis of the dependence structure of network traffic. Our future research includes methods of synthetically generating realistic traces of network delays. The ability to generate these traces has many important uses, including the building of simulation environments for testing the $Q \circ S$ of real-time audio and video tools.

References

[1] Erramilli, O. Narayan, and W. Willinger: "Experimental queueing analysis with long-range dependent packet traffic", *IEEE/ACM Transactions on Networking*, vol. 4, pp. 209-223, Apr. 1996.

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

[3] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson: "On the self-similar nature of Ethernet traffic (extended version)", *IEEE/ACM Transactions on Networking*, vol. 2, pp. 1-15, Feb. 1994.

[4] B. B. Mandlebrot and J. W. Van Ness: "Fractional Brownian motions, fractional noises and applications", *SIAM Review*, vol. 10, pp. 422-437, Oct. 1968.

[5] C. W. Granger and R. Joyeux: "An introduction to long-range time series models and fractional differencing", *Journal of Time Series Analysis*, vol. 1, pp. 15-30, 1980.

[6] J. R. Hosking: "Modelling persistence in hydrological time series using fractional differencing", *Water Resources Research*, vol. 20, pp. 1898-1908, Dec. 1984.

[7] G. Box, G. M. Jenking, and G. Reinsel: *Time Series Analysis: Forecasting and Control.* Englewood Cliffs, NJ: Prentice Hall, 3 ed., 1994.

[8] M. S. Taqqu, V. Teverovsky, and W. Willinger: "Estimator for long-range dependence: An empirical study", *Fractals*, vol. 3, no. 4, pp. 785-798, 1995.

[9] International Telecommunication Union: "Recommendation G.723.1", 1996.http://www.itu.int.

[10] A. Mukherjee: "On the dynamics and significance of low frequency components of Internet load", *Internetworking: Research and Experience*, vol. 5, pp. 163-205, 1994.

[11] W. Willinger, M. S. Taqqu, W. E. Leland, and D. V. Wilson: "Self-similarity in high-speed packet traffic: Analysis and modeling of Ethernet traffic measurements, *Statistical Science*, vol. 10, no. 1, pp. 67-85, 1995.

[12] J. Beran and N. Terrin: "Estimation of the long-memory parameter based on a multi-variante central limit theorem", *Journal of Time Series Analysis*, vol. 15, pp. 269-278, 1994.

[13] J. Beran, R. Sherman, M. S. Taqqy, and W. Willinger: "Long-range dependence in variable-bit-rate video traffic", *IEEE Transactions on Communications*, vol. 43, pp. 1566-1579, Feb./Mar./Apr. 1995.

[14] D. Sanghi, A. K. Agrawala, O. Gudmundsson, and B. N. Jain: "Experimental assessment of end-to-end behavior on Internet", in *Proceedings, IEEE INFOCOM '93*, pp. 867-874, Mar. 1993.

[15] J. D. Scargle: "Studies in astronomical time series analysis II: Statistical aspects of spectral analysis of unevenly spaced data", *Astrophysical Journal*, vol. 263, pp. 835-853, Dec. 1982.

[16] V. Paxson: "Fast approximation of self-similar network traffic, Tech. Rep. LBL-36750, Lawrence Berkley National Lab., Apr. 1995.

[17] J.-C. Bolot: "Characterising end-to-end packet delay and loss in the Internet, *Journal of High Speed Networks*, vol. 2, pp. 305-323, 1993.

[18] C. S. Burns, R. A. Gopinath, and H. Guo: *Introduction to Wavelets and Wavelet Transforms.* Prentice Hall, 1998.

[19] P. Abry and D. Veitch: "Wavelet analysis of long range dependent traffic", *IEEE Transaction of Information Theory*, vol. 44, pp. 2-15, Jan. 1998.

[20] I. Daubechies: *Ten Lectures on Wavelets*, SIAM, 1992.

[21] A. Feldmann, A. Gilbert, P. Huang, and W. Willinger: "Dynamics of IP traffic: A study of the role of variability and the impact of control", in *Proceedings, ACM SIGCOMM '99*, 1999.

[22] Y. Wang, J. Cavanaugh, and C. Song: "Self-similarity index estimation via wavelets for locally self-processes", tech. Rep. Dept. of Statistics, University of Missouri, Columbia, 1998.

Michael S. Borella, Received October 27, 2004

CommWorks, A 3Com Company, 3800 Golf Rd.,
Rolling Meadows IL 60008, USA
Tel: (847) 262-3083, Fax: (847) 262-2255
E–mail: *mike_borella@3com.com*