# An Investigation Using Wavelet Analysis to Detect A Change In The Characteristics Of Self-Similar Traffic

S.A Kalim, Dr Lionel Sacks

Department of Electronic Engineering
University College London

**Abstract:** At present threshold values are based on the mean and standard deviation of link utilisation are used as trigger to find more stable route in the network. The Aim of this paper is to use a technique based on the wavelet method (of Abry and Veitch 1998 ) to detect a real time change in the characteristics of self-similar traffic and this can be used as a threshold trigger to provide optimum path selection for routing protocols such as OSPF (Open Shortest Path First).

## 1. Introduction

Recent measurements of traffic in local area networks have shown that network traffic is scale invariant or bursty across a wide range of time scales. This phenomenon has been observed across diverse networking contexts from Ethernet to ATM, VBR video, and WWW traffic [1]. A number of studies on network performance have shown that self-similarity has a negative impact on network performance leading amplified queuing delay and packet loss [2].

From the prospective of a queue a distinguishing characteristic of long-range dependent traffic is that the queue length distribution decays hyperbolically rather than exponentially, later would be the case for an exponential distribution. These performance effects to some extent can be curtailed by estimating the Hurst parameter by wavelet analysis on real time network traffic. For this purpose we use the matlab code provided by Darryl Veitch available from the web.

The central limit theorem states that a distribution with self-similar characteristics has finite mean but infinite variance and standard deviation. For routing purposes threshold values based on the mean and standard deviation of arrival process modelled on Poisson or exponential distributions are used to detect change in traffic behaviour. For self-similar traffic false increases in the threshold value will trigger false action to be taken for example selecting a more stable path. Hence we need to investigate the characteristics of self-similarity to detect a significant change in the characteristic of the traffic and use that as a threshold for the purpose of effective path selection.

The Hurst parameter measures the degree of self-similarity therefore it is important to consider it in the control and management of a network. The use of the estimation of the Hurst parameter as a metric could be useful in obtaining more stable routes and in buffer management.

## 2. Self-Similarity And Heavy-Tails

Self-similarity is defined as the scale-invariance property of the traffic. The term self-similar was first used by Mandelbrot [3] during the 1960's to designate those processes that are scalable over time without losing their scalable properties. A stationary process is long range dependent (LRD) if its autocorrelation function r(k) is non sum-able therefore the definition of long-range dependence applies only to infinite time series [4].

A discrete-time wide sense stationary process with mean $\mu$ and variance $\sigma$ squared and autocorrelation function,

$$\Gamma_x(k), k \geq 0,$$

Which decays slowly at large lags and is given by;

$$\Gamma_x(k) \sim c_\Gamma K^{-b}, k \to \infty$$

Where Cr is a positive constant and $0<\beta<1$

For each m = 1, 2, 3,…, let denote a new wide sense stationary time series obtained by averaging the original time series x over non-overlapping blocks of size m.

Samples are given by;

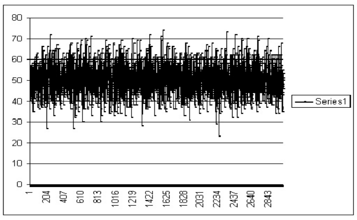
$$x_i^{\ m} = \frac{1}{m}(x_{im-m+1} + \ldots.. + x_{im}), i \geq 1$$

The process x is considered second-order self-similar with a self-similarity Hurst parameter,

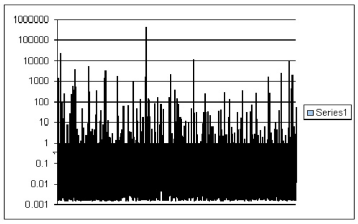
$$H = 1 - \beta/2$$

The Hurst parameter represents the degree of self-similarity in the observed traffic. When the value of the Hurst parameter is between 0.5 and 1 the traffic is said to be self-similar values of H closer to 1 indicate a high degree of self-similarity.

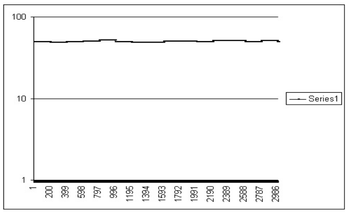
**Non Self-Similar Traffic** **Self-Similar Traffic**



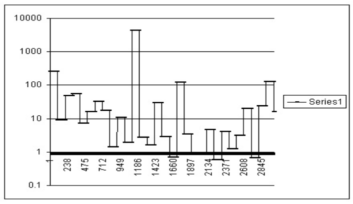

**X10**




**X10**




**X 100**




**X 100**

Figure: (A) Figure: (B)

The above diagram is the histogram of the arrival rate. It shows that on a large time scale Poisson or non self-similar traffic in figure (A) averages out. On the other hand Pareto or self-similar traffic in figure (B) remains bursty over a wide range of time scale.

## 3 Heavy tails

The simplest heavy tailed distribution is the Pareto distribution with probability density function

$$f(x) = \frac{a}{k}\left(\frac{k}{x}\right)^{a+1}$$

The parameter $\alpha$ is the shape parameter. As the value of alpha decreases the more of the probability mass is located in the tail of the distribution.

## 4. The Wavelet Method

The wavelet method is used to estimate the Hurst parameter unlike other estimation methods it offers a semi-parametric estimator which can achieve much more efficiency in terms of a statistical and complexity. To estimate the Hurst parameter we use a technique called the wavelet method [5] and is implemented in Matlab.

In the wavelet method the differences in aggregated series is analysed if $Y^{j}$ is aggregated series than,

$$Y_{2k}^{j+1} = (Y_{2k}^{j} - Y_{2k-1}^{j})1/\sqrt{2}, k = 1,2,.....,n/2j \quad \text{where j = 1,2,....}$$

Since expectation of Y is zero. In frequency domain the variance is equivalent to the signal energy in a frequency band depending on j $E_j Vs 2^j$ is plotted on log-log scale. Linearity is checked for all scales j. The Hurst parameter is then calculated by calculating the slope of the line.

## 5. Modelling Wide Area Ethernet Traffic

The paper by Murad Taqqu, Walter Willinger [6] describes how by integrating simple ON-OFF processes results in self-similar behaviour. The integration of these individual ON-OFF sources allow for the explanation of observed self-similarity in wide area traffic.

In this case the traffic source is transmitting packets at constant rate in the ON period and is idle in the OFF period. The ON-OFF periods are taken from a heavy tailed distribution such as Pareto distribution. Which has infinite variance and finite mean.

When a large number of these sources are aggregated it results in traffic having self-similar characteristics [LTWW94].

The model is implemented in NS (Network Simulator) a large number of sources are used. The queuing discipline used is Drop-Tail the first packet in the queue is served first and the last one is served last. UDP is used as the transport layer. The queue state probability is calculated from simulation for different values of the shape parameter alpha.

## 6. Results

6.1 Effect of change in the shape parameter on the queue state

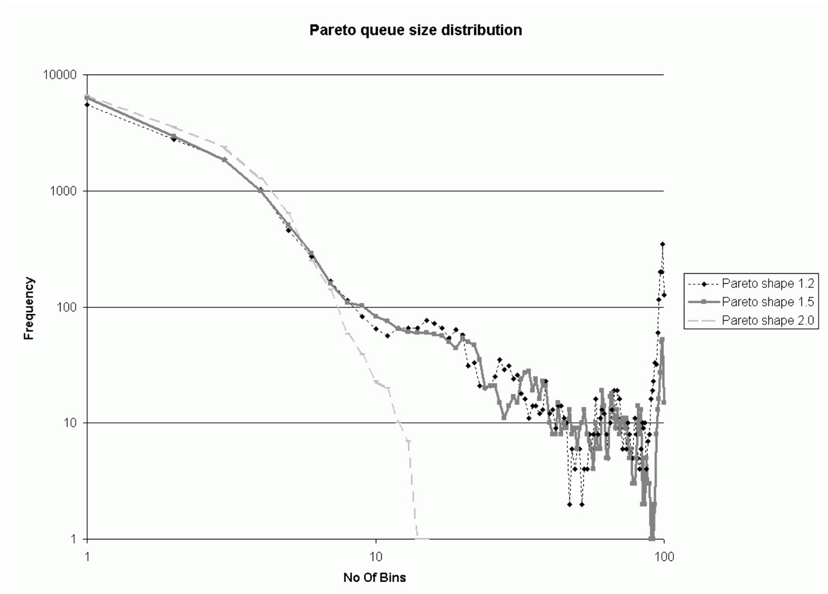



Figure No1: Log-log graph of frequency against number of bin for the queue

These results were obtained from the above experiment and show as the shape parameter of the Pareto distribution increases from 1.2 to 2.0 the distribution becomes more like SRD (short-range dependent process). For when the shape parameter is close to one the distribution has infinite variance and finite mean. The queue size decays slowly in an, hyperbolic fashion indicating the presence of long-range dependence.

The above graph indicates that although there is high probability of small number of packets remaining in the queue and that there is significant probability of having large number of packets in the queue. This is the spiky nature invariant property of such traffic that makes it hard to dimension buffers because at some point they will over flow.

6.2 Effect of increase in load on the Hurst parameter

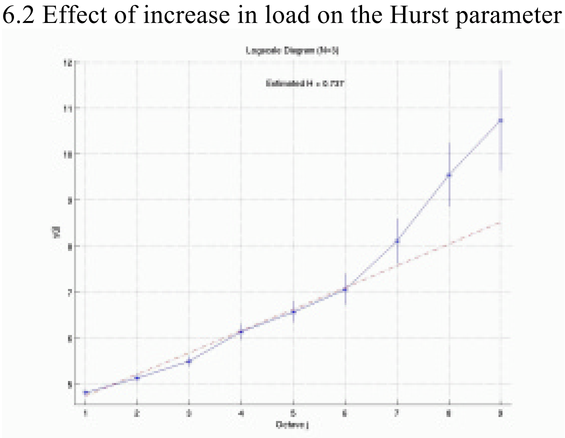


**Figure (1) Load 30Mb**

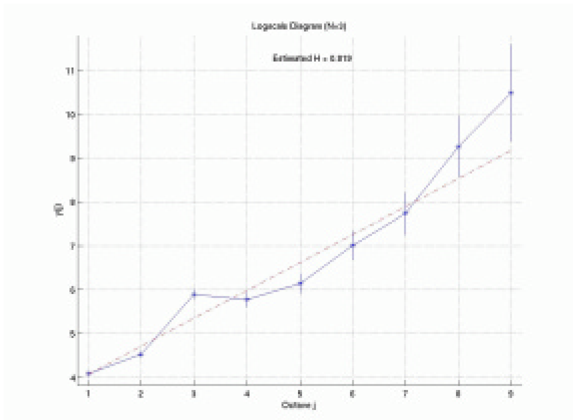


**Figure (2) Load 27Mb**

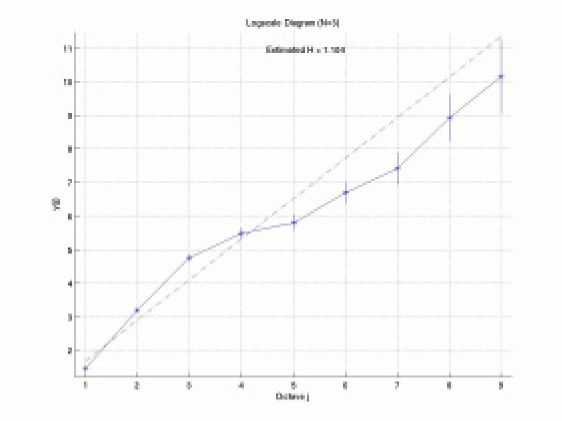


**Figure (3) Load 24Mb**

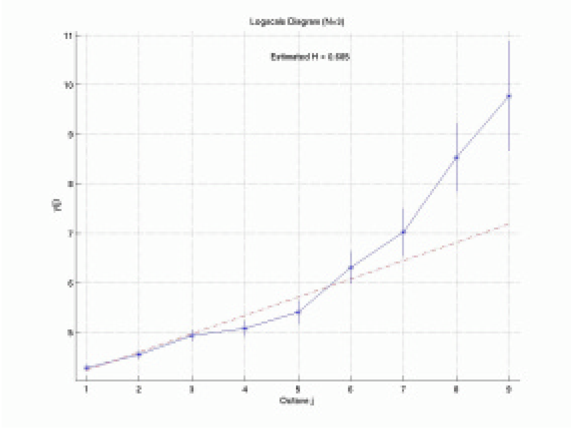


**Figure (4) Load 21Mb**

The graphs of figure (1) (2) (3) and (4) are produced by using the Wavelet method. [Abry & Veitch]

1) The above graphs are log scale. The line of regression of $Y^j$ against j for LRD process is plotted. The vertical bars at each octave give the 95% confidence intervals for the $Y^j$ . The analysed data is synthetic data generated from the aggregation of large no of ON-OFF Pareto sources.
2) As the load increases or decreases there is a change in the slope of the line of regression. Which is used to calculate the Hurst parameter. It can be observed from the graph of fig1 to fig4 as the load decreases the line of regression also shifts down wards.
3) We can see that the load has an effect on the slope of the line. The Hurst parameter is calculated from using the slope this demonstrates that changing the load has an effect of shifting the line and the slope. So the Hurst parameter is not directly proportional to the load.

The Hurst parameter indicates the degree of self-similarity in the traffic. The shift in the line indicates that more rare events will occur which will that increase the occupancy of the queues and hence packets will be dropped.

## 7. Future work

It will be interesting to investigate how the change in the behaviour of real time self-similar traffic can be used to find a threshold that can be used as trigger in routing protocols to provide better quality of services in order to support services, which require a high degree of quality of service guarantees.

The effect of TCP flow control actually exaggerates the multiplicative effect therefore it will be interesting to try this experiment with TCP.

## 8. Conclusion

Due to the self-similar nature of LAN and WAN traffic the usual metrics used for routing are probably not valid. Therefore we need to find a metric for routing based on real time wavelet estimates of the Hurst parameter to provide a threshold trigger for effective stable route selection in routing protocols.

## 9. References


[1] Ethernet to ATM, VBR video and WWW traffic [Crovella and Bestavros]

[2] Wide Area Traffic: The Failure of Poisson Modelling Vern Paxson and Sally Floyd 1995

[3] B.B Mandelbrot and J.W. Van Ness, “Fractional brownian motions, fractional noises and applications SIAM Rev, vol.10, pp. 422-437, 1968

[5] Fast, Approximate synthesis of fractional gaussian noise for generating self-similar network traffic Vern Paxson

[5] Wavelet Analysis of Long-Range dependent traffic by Abry and Veitch

[6] Proof of a Fundamental Result in Self-Similar Traffic modelling by Murad Taqqu, Walter Willinger and Rober Sherman.