INTERNET TRAFFIC AND PACKET ROUND TRIP DELAY SELF-SIMILARITY

Taye Abdulkadir, Mohammed Abdo and G. Devarajan
Department of Electrical and Computer Engineering
Addis Ababa University

Abstract

Statistical analysis was carried out on the Internet traffic data grabbed at Internet Service Provider gateway. The analysis reveals the "bursty" nature, which was explained through Self-Similarity and Long Range Dependence(LRD). In this work, it has been shown that the Internet traffic and Packet Round Trip Delay visualized as a time series are statistically self-similar. The autocorrelation function decays reveals the data under investigation exhibits Long Memory or Long Range Dependence. The degree of self-similarity of Packet Round Trip Delay as measured by Hurst Parameter is presented and was found that it is directly related to network traffic load.

Key words: Self-Similarity, LRD, Internet Traffic, Internet Packet Round Trip Delay.

INTRODUCTION

Leland, et al [2] present a preliminary statistical analysis of the fractal nature of a High quality *Ethernet* Traffic data collected from Bellcore Morristown Research and Engineering Center and comment in detail the presence of "Burstiness" across an extremely wide range of time scales. This fractal like behaviour of aggregated Ethernet LAN Traffic is different from conventional telephone traffic. Subsequent research work on high traffic networks such as Internet also leads to having similar fractal like behaviour.

Self-similarity and *fractals*, notions pioneered by Benoit B. Mandelbrot [3], describe the phenomenon where a *certain* property of an object for example, *a natural image, the convergent sub domain of certain dynamical systems, a time series* (the *mathematical* object of our interest)- is *preserved* with respect to *scaling* in space and/or time. If an object is Self-similar or fractal, its parts when magnified resemble -in the suitable sense - the shape of whole. *Stochastic Self-similarity* admits the infusion of non-determinationism as necessitated by measured traffic traces but, nonetheless, is the property that can be illustrated visually. *Unlike deterministic fractals*, the objects do *not* possess exact resemblance of their part with the whole at finer details. Indeed, for measured traffic traces, it would be too much to expect to observe exact, determinist self-similarity given the *stochastic* nature of many network events (e.g. source arrival behaviour) that collectively influence actual network traffic. Second order statistics are statistical properties that *capture* burstiness or variability, and the *Autocorrelation* function is a yardstick with respect to which scale invariance can be fruitfully defined.

This nature of *time-invariant burstiness* is completely different from the traditional *tele-traffic models,* such as the *Poisson Process,* which has a "smoothed-out" burst structure as the time aggregation increases; that is, after a certain time scale there are no surprises due to spikes of Traffic. For actually measured traffic the correlation in traffic can extend to a wide range of different time, or mathematically, the correlation function of realistic traffic decays with lag time in the way of *power law*, which is the property of the so called *Long Range Dependence (LRD)*; while for traditional models of generated traffic, its correlation function decays exponentially fast; resulting in *Short Range Dependence*. [1]

The main objectives of this paper are: *(a)* To study the statistical nature of Internet Packet Traffic and Round Trip Delay Processes and *(b)* To illustrate some of the differences between self-similar models and the standard models for packet traffic considered in the literature. Accordingly, this paper has the following sections.

THE INTERNET

The word *internet (*also *internetwork)* is simply a contraction of the phrase *interconnected network.* However, when written as a capital "I" the *Internet* refers to a *worldwide* set of interconnected networks, so the Internet is an internet, but the reverse does not follow. One of the greatest things about the Internet is that nobody really owns it. It is a global collection of networks, both big and small. These networks connect together in many different ways to form the single entity that we know as the *Internet*.

Current data networks typically use packet switching as a means of dynamically allocating network resources on a demand basis. Packet switching had been widely used because it facilitates the *interconnection* of networks with different architectures, and it provides flexible *resource allocation* and good *reliability against node and link failure*. Packets of a single traffic stream may take different routes and reach the intended destination. This process of assigning packets to available routes is called *Routing* and the devices facilitating this operation are called *Routers*. Routing of the Internet has two features: *Flexibility and Scalability*. The Internet provides the *Dynamic* Routing based on the exchange of the *Routing Information* among Routers.

Each Internet communication consists of a *transfer* of information from one computer to another. When a file is transferred, it is not sent across the Internet as a continuous block of bits, rather the file is broken up into pieces called *packets*, and each packet is sent individually. Many different *protocols* collectively carry out the transfer. The two core protocols are TCP, the *Transmission Control Protocol* and IP, the *Internet Protocol*.

Characteristics of Internet Traffic

Internet engineering and management depend on an understanding of the characteristics of network traffic. *Statistical models*, which can generate traffic that mimics closely the observed behavior on live Internet wires, are needed. Models can be used on their own for some tasks and combined with network simulators for others, but the challenge of model development is immense. Internet traffic data are ferocious. Their statistical properties are complex, databases are very large, Internet network topology is vast, and the engineering mechanism is intricate and introduces feedback into the traffic.

Packet header collection and organization of the headers into connection flows yields data rich in information about traffic characteristics and serves as an excellent framework for modeling. Many existing statistical tools and models, especially those for time series, point processes, and marked point process, can be used to describe and model the statistical characteristics, taking into account the structure of the Internet, but failed to exhaustively represent the properties of the Internet traffic's inherent nature. Therefore, new tools and models are needed. Internet traffic data are exciting because they measure an intricate, fast-growing network connecting up the world, transforming culture, politics, and business. A deep analysis of Internet traffic can contribute substantially to network performance-monitoring, equipment planning, quality of service, security, and the engineering of Internet communications technology. Further analysis can be made on traffic measurements to produce statistical models.

Modeling Internet traffic data will require new approaches, new tools, and new models for time series data. Long-range Dependence is pervasive in Internet traffic data, but the pervasiveness had to be discovered. After the discovery of long-range dependence, Internet traffic can be studied through the vehicle of *Self-Similar* processes, invoking the creative work of Mandelbrot [3]. Traffic models for voice traffic, developed over the years to serve the telephone network, did not apply as might have been hoped because voice traffic does not give rise to the same traffic characteristics as Internet data traffic, which is *burstier*.

Packet Delays in the Internet

A packet round-trip delay is the sum of delays on each subnet link traversed by the packet. Each link (or hop) delay in turn consists of four components, including *processing delay, Queueing delay, transmission delay and propagation delay*. Fixed the packet length and the route, the packet round-trip delay only changes with the Queueing delay, which in Internet is changed with the fluctuation of traffic. Internet is expanding dramatically fast, as it is the most complicated collection of Networks connected together. Many well developed and currently developing applications run across Internet to go around the global network. Some applications (audio, video) are sensitive to the performance of the whole Internet, or precisely, the Packet Delay in Internet. When data networks become heavily utilized, the combined bandwidth demand from all sources occasionally exceeds the network capacity. For data, this is not a problem. It simply means that the data does not arrive as quickly; it is delayed by a few additional milliseconds. For data traffic, the motto is "Better late than never", so it does not have to pay the penalty of retransmission.

Video and voice traffic, on the other hand, must get a fairly regulated number of packets through to the destination in a timely manner. If video/voice packets are late enough to have missed their "play" window, they are useless. Hence the motto for video/voice traffic is "Better never than late." The network should drop traffic that is late so it does not consume additional scarce network resources.

The capability to provide resource assurance and service differentiation in a network is often referred to as quality of service (QoS). QoS techniques are designed to balance the needs of voice, video, and data across the network. QoS reserves a portion of the network bandwidth for the predictable use of the voice/video and lets the data traffic consume the remainder. By ensuring that bandwidth is available when needed for voice/video traffic QoS techniques can reduce or eliminate audio pops, video artifacts, and other performance problems.

SELF-SIMILARITY & LRD

The term "Traffic Theory" originally encompassed all mathematics applicable to the design, control and management of the Public Switched Telephone Networks (PSTN); Statistical Inference, Mathematical Modeling, Optimization, Queuing and Performance Analysis. Later, its practitioners would extend this to include data networks such as the Internet.

Traditional tele traffic theory as applied to POTS (Plain Old Telephone Service) has arguably been one of the *most successful* applications of mathematical technique in industry. It has led to first-rate telephone networks, quality of service we fully rely on and take for granted. Among the main reasons for this tremendous success of tele-traffic theory and practice in traditional telephony are the highly *static* nature of conventional PSTNs and a well defined and ever present notion of limited variability, a trade mark of homogenous systems where one talk about *"typical" users* and *"generic" behaviour* and where *averages* describe the system performance adequately.

The static nature of traditional PSTNs contributed to the popular belief in the existence of "universal laws" governing voice networks, the most significant of which is the *presumed Poisson* nature of call arrivals at links in the network traffic is heavily aggregated. This law states that call arrivals are mutually *independent*, and that call inter arrivals times are all *exponentially* distributed, with one and the same parameter.

Failure of Poisson Modelling

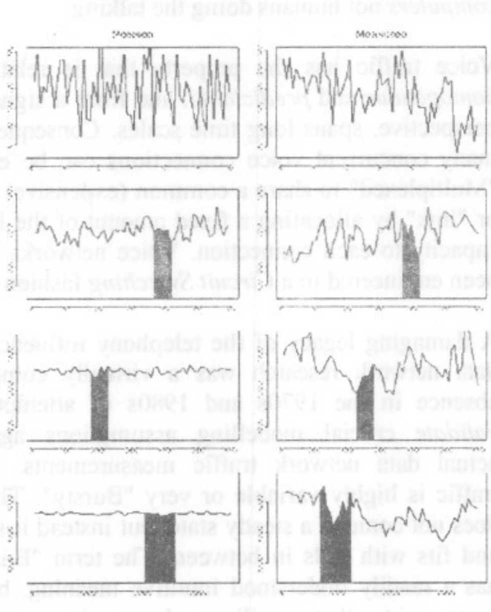
One might expect that the voice network modelling success story would enjoy another triumph when applied to data networks, and indeed this has been attempted. But in fact much of the voice traffic modelling has proven nothing short of disastrous when applied to data networks; for the simple but profound reason that the rules all change when it is *computers* not humans doing the talking.

Voice traffic has the property that is relatively *homogenous* and *predictable*, and from a signaling perspective, spans long time scales. Consequently, many concurrent voice connections can be easily "Multiplexed" to share a common (expensive) wire or "link" by allocating a fixed amount of the link's capacity to each connection. Voice networks have been engineered in a *Circuit Switching* fashion.

A damaging legacy of the telephony influence on data network research was a virtually complete absence in the 1970s and 1980s of attempts to *validate* crucial modelling assumptions against actual data network traffic measurements. Data traffic is highly variable or very "Bursty". That it does not come at a steady state, but instead it starts and fits with lulls in between. The term "Bursty" has a readily understood intuitive meaning, but it turns out that nailing down its precise, mathematical meaning has profound implications for developing *mathematical model* of Network Traffic.

The relevant mathematics for POTS is one of *limited variability* in both time (traffic processes are either independent or have temporal correlations that decay exponentially fast) and in space- i.e. the distribution of traffic related quantities have exponentially decaying tails, but for data networks, the mathematics is one of *high* or *extreme variability*. [5]

Figure 1 is a visual demonstration of the failure of Poisson modelling to capture the burstiness present in actual network traffic. The difference between the Poisson Model and the measured traffic is obvious and striking: as the time scale increases the Poisson Traffic "smoothes out", becoming quite tame, while the measured traffic shows no such predilection. The difference is crucial from an Engineering perspective: Traffic that behaves as shown in Fig. 1(a) can be easily engineered for. Above a certain time scale there are no surprises – every thing boils down to knowing the long term arrival rate; no need for big buffers in routers or switches, no reason for being conservative in choosing safe operating points for engineering back bone trunks, and why even think of user perceived Quality of Service as being a relevant issue? In stark contrast, measured traffic like that showed in Fig. 1(b) is "wild", remains so even on quite coarse time scales, and plays havoc with conventional traffic engineering.



(a) *(b)*

Figure 1(a) Synthesized traffic from a Poisson model Vs. (b) Internet traffic to which its mean and variance fit, viewed over three orders of aggregation. Source: reference [6]

Welcome Fractals

Many networking experts argue that the only way to gain an in-depth understanding of data traffic is –simply put- doing away with teletraffic tradition and starting from scratch. Interestingly, *Mathematics*, which has been largely responsible for the success story of the teletraffic theory for the voice network, has recently provided strong ammunition in supporting of the Networking experts' arguments. However, as voice traffic turns out to differ drastically from data traffic, so do the underlying mathematical ideas and concepts.

Statistically, Long Range Dependence, i.e Autocorrelations that exhibit power law decay, captures temporal high variability in traffic processes. On the other hand, extreme forms of spatial variability can be described parsimoniously using *heavy tailed distribution* with infinite variance, i.e, probability distribution F with the property that for large x values

$$1 - F(x) \approx k_1 x^{-\alpha} \qquad (1)$$

Where k_1 is a positive finite constant that does not depend on x and where the tail index α is in the interval (0,2). (For example, this property is satisfied by the well-known family of "Pareto Distribution", originally introduced for modelling the distribution of income within the population[4].

It turns out that Power Law behavior in time or space of some of their statistical descriptions often cause the corresponding traffic process to exhibit Fractal characterstics. In the present context we say that a traffic process has fractal characterstics, if there exists a relationship between certain quantities Q of the underlying process and resolution as the general form

$$Q(\tau) \approx k_2 \tau^{f(D)} \qquad (2)$$

Where k_2 is a positive finite constant that does not depend on τ. $f(.)$ is simple, often *linear*, function of D; D is a fractal dimension. To declare fractality, the above relationship is supported to hold for range of different τ-values with a value of D that is less than the embedded dimension.

In view of the general skeptism that exists in the different circles in the mathematical community concerning the needs, usefulness, and appropriateness of fractals, what can one say about fractal like scaling in measured date network traffic? To examine this question, we call a discrete time covariance stationary, zero mean *stochastic* process $X=(X_k : k \geq 1)$ *exactly* ***Self-Similar or Fractal*** with scaling parameter

$H \in [0.5,1)$ if for all levels of aggregation or "Resolution" $m \geq 1$.

$$X^{(m)} =_d m^{H-1} X \quad (3)$$

Where the equality is understood in the sense of Finite dimensional distributions, and where the aggregated process $X^{(m)}$ are defined by

$$X^{(m)}(k) = m^{-1}(X_{(m-1)k+1} + \ldots\ldots + X_{km}) \quad (4)$$

When assessing the validity of describing a process using a ***self-similar model***, one must be very careful not to mistake actual non-Stationarity for highly variable but stationary fractal behavior. The two can appear very similar, both to the eye and to a number of statistical tests. However this concern can be addressed by making good use of the very large size of Network traffic traces.

The switch from Poisson to Fractal thinking in Network traffic research has had a major impact on the understanding of actual network traffic to the point where we now know why aggregate Internet Traffic exhibits fractal-scaling behavior over different time scales. A measure of the success of this shift in thinking is that the corresponding mathematical arguments are at the same time rigorous and simple, are in full agreement with the Networking researchers' intuition, and can be explained readily to a non-networking expert.

Mathematics of Self-Similarity

Self-Similarity and Fractals, notions pioneered by Beniot Mandelbrot, [3] describe the phenomenon where a certain property of an object-for example, a natural image, the convergent sub domain of certain dynamical system, a time series (the mathematical object of our interest)-is preserved with respect to scaling in space and /or time. If an object is self-similar or fractal, its parts, when magnified, resemble, - in a suitable sense - the shape of the whole.

Stochastic Self-Similarity admits the infusion of non-determinism as necessitated by measured traffic traces but, nonetheless, is a property that can be illustrated visually. Fig. 2(a) shows a traffic trace, where throughput in Bytes plotted against time where time granularity is 100s. That is, a single data point is the aggregated traffic volume over a 100 second interval, Fig. 2(b) is the same traffic series whose first 1000 second interval is "blown up" by a factor of ten. Thus, the truncated time series has a time granularity of 10s. The remaining two plots (Figs. 2(c) and (d)) zoom in further in the initial segment by rescaling successively by factor of 10, for 1s and 10ms granularity respectively [4].

Unlike deterministic fractals, the objects corresponding to Fig. 2 do not posses resemblance of their parts with the whole at finer details. Here, we assume that the measure of "resemblance" is the shape of the graph with the magnitude suitably normalized. Indeed, for measured traffic traces, it would be too much to expect to observe exact, deterministic self-similarity given the stochastic nature of many network events (e.g. Source arrival behavior) that collectively influence actual network traffic.

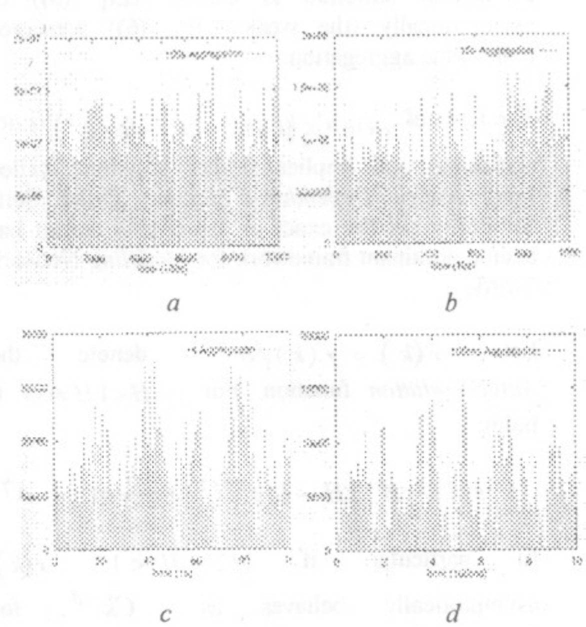


Figure 2 Stochastic Self-Similarity-in the "Burstiness preservation sense"-across time scales *(a)*100s *(b)* 10s *(c)* 1s *(d)* 100ms

Second order statistics are statistical properties that capture *burstiness* or *variability*, and the *autocorrelation* function is a yardstick with respect to which scale invariance can be fruitfully defined. The shape of the autocorrelation function-above and beyond its preservation across rescaled time series- plays an important role. In particular, correlation, as a function of time lag, is assumed to decrease polynomially as opposed to exponentially. The existence of non-trivial correlation "*at a* *distance"* is referred to as *Long Range Dependence.*

A stochastic process or time series X(t) is exactly Second Order Self-Similar with Hurst Parameter $H, (1/2 < H < 1)$ if

$$\gamma(k) = \sigma^2/_2 \left((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right) \quad (5)$$

for all $k \geq 1$.

X(t) is *asymptotically* Second Order Self-Similar if

$$\lim_{m\to\infty} \gamma^{(m)}(k) = \sigma^2/_2 \left((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right) \quad (6)$$

It can be checked that Eq. (5) implies $\gamma(k) = \gamma^{(m)}(k)$ for all $m \geq 1$. Thus second order self similarity captures the property that the correlation structure is exactly (Eq. (5)) or asymptotically (the weaker Eq. (6)) preserved under time aggregation.

The form of $\gamma(k) = \sigma^2/_2 \left((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right)$ is not accidental and implies further structure; called *Long Range Dependence* Second Order Self-Similarity (in the exact or asymptotic sense) has been a dominant framework for *Modeling* Network Traffic.

Let $r(k) = \gamma(k)/\sigma^2$ denote the *Autocorrelation* function. For $0 < H < 1, H \neq 0.5$ it holds

$$r(k) \sim H(2H-1)k^{2H-2},\ k \to \infty \quad (7)$$

In particular, if $1/2 < H < 1$, $r(k)$ asymptotically behaves as $Ck^{-\beta}$, for $0 < \beta < 1$, where $C>0$ is a constant, and $\beta = 2 - 2H$. Then it implies:

$$\sum_{k=-\infty}^{\infty} r(k) = \infty \quad (8)$$

That is, the autocorrelation function decays slowly (that is, *Hyperbolically*), which is the essential property that causes it to be not summable. When $r(k)$ decays hyperbolically, we call the corresponding stationary process *X(t) Long range Dependent. X(t)* is *Short Range Dependent* if the Autocorrelation function is summable.

In the case of asymptotic second order self-similarity however, by the restriction $0.5 < H < 1$ in the definition, Self-Similarity implies Long Range Dependence, and vise versa[4]. It is for this reason and the fact that asymptotic second-order Self-Similar Processes are employed as "Canonical" traffic models, that we use Self-Similarity and Long-Range dependence *interchangeably* when the context does not lead to confusion.

ESTIMATION OF LONG-MEMORY

Many methods for estimating the *self-similarity parameter H, called Hurst Parameter, or the intensity of Long-range dependence* in a time series are available. They are typically validated by appealing to self-similarity or to an asymptotic analysis where one *supposes* that the *sample size* of the time series *converges to infinity.* For the analysis of Internet Traffic and Packet Round Trip Delay time series data, we use two *Heuristic* and one *Periodogram* based methods. Out of several Heuristic methods to estimate the Long memory parameter *H,* the R/S statistic, which was first proposed by Hurst (1951) in a Hydrological context, and the variance plot are discussed. The Periodogram-based approach used in the analysis is Whittle Estimator.

R/S Statistic

Suppose we want to calculate the capacity of a reservoir such that it is *ideal* for the time span between t and $t+k$. To simplify matters, assume that *time is discrete* and that there are *no storage losses.* By ideal capacity we mean that we want to achieve the following: *the outflow is uniform, that at time* $t+k$ *the reservoir is as full as at time* t, *and that the reservoir never overflows.*

Let X_i denote the inflow at time i and $Y_j = \sum_{i}^{j} X_i$ is the *cumulative* inflow up to time j. Then the ideal capacity can be shown to be equal to

$$R(t,k) = \max_{0\leq i\leq k}\left[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)\right] - \min_{0\leq i\leq k}\left[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)\right] \quad (9)$$

$R(t,k)$ is called the *adjusted range*. In order to study the properties that are independent of the scale, $R(t,k)$ is standardized by

$$S(t,k) = \sqrt{k^{-1}\sum_{i=t+1}^{t+k}(X_i - \bar{X}_{t,k})^2} \quad (10)$$

$$\text{where } \bar{X}_{t,k} = k^{-1}\sum_{i=t+1}^{t+k} X_i$$

Then, the ratio $\frac{R}{S} = \frac{R(t,k)}{s(t,k)}$ (11)

is called the *Rescaled Adjusted Range or R/S statistic.* Hurst plotted the logarithm of *R/S* against several values of k. He observed that, for large values of *k*, $\log R/S$ was scattered around a *straight line* with a slope that exceeds ½. In probabilistic terminology this means that for large *k*,

$$\log E[R/S] \approx a + H \log k, \quad (12)$$

with $H > 1/2$

This empirical finding was in *contradiction* to results for Markov Processes, Mixing Processes, and other stochastic processes that where usually considered at that time.

For any stationary process with Short-range dependence, *R/S* should behave asymptotically like constant times $k^{1/2}$. Therefore, for large values of *k*, $\log R/S$ should be randomly scattered around a straight line with slope ½.

With $Q(t,k) = \frac{R(t,k)}{S(t,k)}$, the R/S method can be summarized as follows: [7]

1. Calculate Q for all possible (or for a sufficient number of different) values of t and k.
2. Plot $\log Q$ against $\log k$.
3. Draw a straight line $\log Q = a + b\log k$ that corresponds to the "ultimate" behaviour of the data. The coefficients a and b can be estimated, for instance by *least squares* [8]. Set the value of H to be the estimated slope b of the fitted straight line.

Variance Plots

One of the striking properties of Long-memory processes is that the variance of the sample mean converges *slower* to zero than n^{-1}, where *n* is the sample size.

$$Var(\bar{X}_n) \approx cn^{2H-2} \quad (13)$$

where c>0.

Equation (13) suggests the following method for estimating *H*: [7]

1. Let k be an integer. For different Integers k in the range $2 \le k \le n/2$, and a sufficient number (say m_k) of sub series of length k, calculate the sample means $\bar{X}_1(k), \bar{X}_2(k), \ldots\ldots, \bar{X}_{m_k}(k)$ and the overall mean
$$\bar{X}(k) = m_k^{-1}\sum_{j=1}^{m_k}\bar{X}_j(k) \quad (14)$$
2. For each k, calculate the sample variance of the sample means
$\bar{X}_j(k),\ j = 1,2,\ldots\ldots, m_k$
$$S^2(k) = (m_k - 1)^{-1}\sum_{k=1}^{m_k}(\bar{X}_j(k) - \bar{X}(k))^2 \quad (15)$$
3. Plot $\log S^2(k)$ against $\log k$

For large values of *k*, the points in this plot are expected to be scattered around a *straight* line *with negative slope 2H-2.* The straight line is fitted with *least squares* method.

In the case of Short-range dependence or Independence, the ultimate slope is $2 \times 1/2 - 2 = -1$. Thus, the slope is steeper (more negative) for short memory processes.

Periodogram Method

The Periodogram of the time series:

$$I(\lambda) = \frac{1}{2\pi N}\left|\sum_{j=1}^{N} X_j e^{ij\lambda}\right|^2 \quad (16)$$

where λ is a frequency, *N* is the number of terms in the series, and X_j is the data.

Because $I(\lambda)$ is an estimator of the Spectral density, a series with Long-Range dependence should have a Periodogram which is proportional to $|\lambda|^{1-2H}$ close to the origin. Therefore, a regression of the logarithm of the Periodogram on the Logarithm of the frequency λ should give a coefficient of $1-2H$. This provides an approximation to the parameter *H*.

The Whittle estimator is based on the Periodogram. It involves the function:

$$Q(\eta) = \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda;\eta)} d\lambda \qquad (17)$$

where $I(\lambda)$ is the Periodogram and $f(\lambda;\eta)$ is the spectral density at frequency λ, and where η denotes *the Vector of Unknown Parameters.* The Whittle Estimator is the value of η which *Minimizes* the function $Q(\eta)$.

EXPERIMENTATION AND RESULTS

This section presents the measurements and experimentations we follow to prove the *Self-Similarity* of Internet Packet Traffic and Internet Packet Round Trip Delay Processes. The statistical properties observed on the collected Internet traffic and Packet Delay data are discussed.

LRD in Internet Traffic

The collected traffic data is a count of the number of packets or bytes traversing a selected port per a *10ms* time unit. The collection is done using Artiza Packet Trapping Software for trapping Internet packets at an Internet Service Provider Gateway (Ethiopian Telecommunications Corporation (ETC)) and *AG Group Etherpeek version 3.5* for trapping packets at the Internet Gateway in a company Local Area Network (*Icon Networks*). These software tools only *count* the traffic *Intensity (number of Packets or Bytes)* and have no facility for revealing the content of the packets.

Scale Invariance and Autocorrelation

The minimum resolution of the packet trapping software is 10ms. The number of bytes of the packets collected for each 10ms time interval is sequenced as a *time series.* Then these time series are analyzed to investigate the statistical properties.

Figure 3*(a)* and *(b)* show the Internet Traffic intensity in Packets per time unit versus index of the sequence. Fig. 3*(a)* represents the original time series (10ms aggregation) and Fig. 4*(b)* represents a four level aggregation of the original time series (40ms).

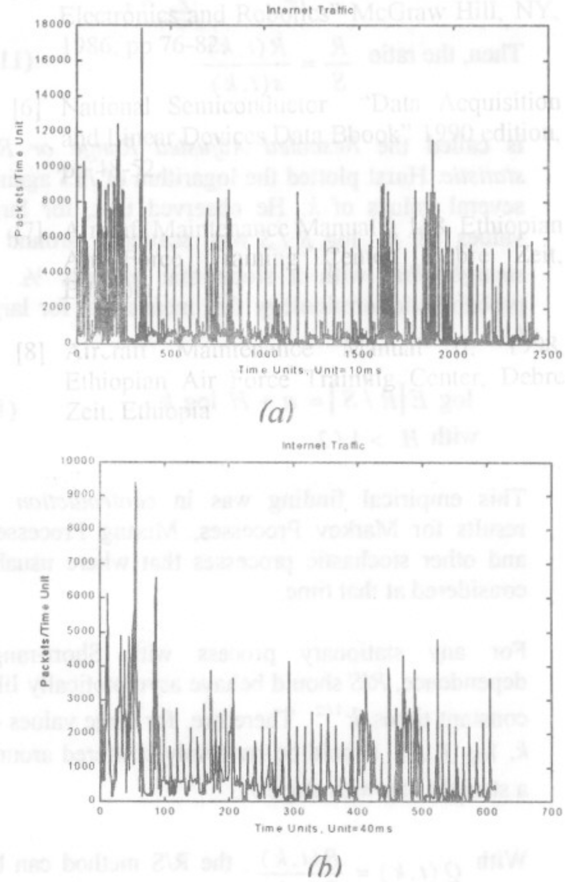
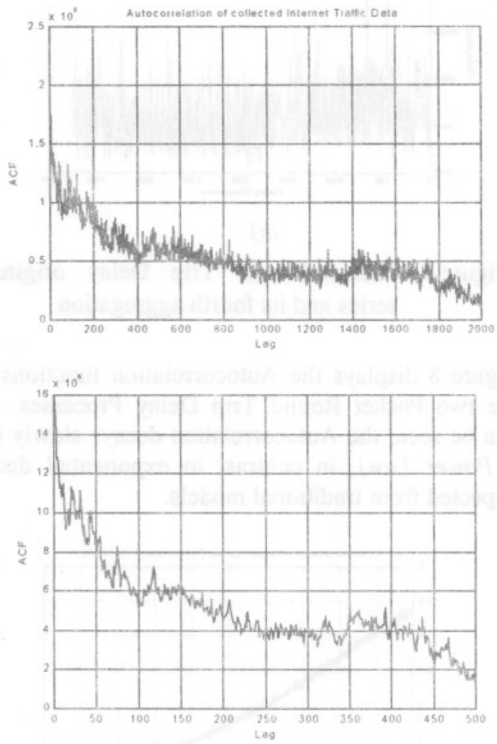


Figure 3 Internet Traffic in (a) 10ms and (b) 40ms time aggregation

As can be seen from Fig. 3, the *burstiness* of the traffic never *smoothed* even though different levels of aggregation are used. This suggests that there is no natural length of a burst in Internet traffic; the bursts remain in all levels of aggregation.

Figures 4*(a) and (b)* depict the Autocorrelation of the original time series (10ms) for the first 2000 lags and its aggregated (50ms) time series for the first 500 lags, respectively.

As can be seen from Fig. 4, the Autocorrelation decays *slower* than the exponential rate.



(b)

Figure 4 Autocorrelation of (a) the original time series (b)its 50ms aggregated time series

Hurst Parameter Estimation

The *intensity of Self-Similarity and Long-Range Dependence* can be catched through the help of the *Hurst Parameter.* The following section presents the results of the Hurst Parameter estimation techniques applied to the collected Internet Traffic Data. (*Recall that values of H in the range (0.5 < H <1) imply Self-Similarity or Long-Range dependence).*

R/S **analysis**

The result obtained after analyzing the collected Internet Traffic data using the *Rescaled Range (R/S)* analysis shows the self-similar behavior of the Internet traffic. The value of *H* (Hurst parameter) obtained was ***0.90126***.

Figure 5 depicts the log-log plot of R/S statistic. It shows an asymptotic slope that is distinctly different from 0.5 and is estimated using least squares to be ***0.90126***.

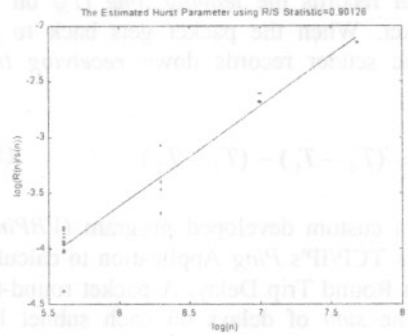


Figure 5 R/S Analysis of Internet Traffic Trace

VARIANCE PLOT

The result obtained using the Variance Plot method proves that the data has long memory or it is long-range dependence. The estimated Hurst parameter equals ***0.80612***. Figure 6 asserts this.

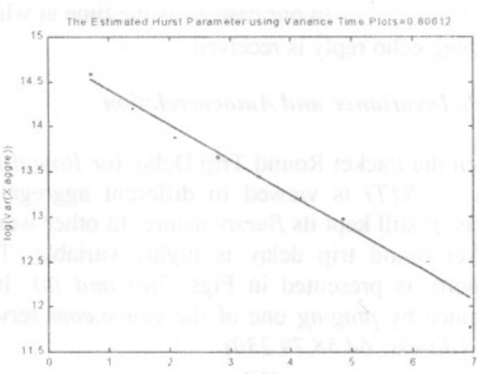


Figure 6 Variance plot of Internet Traffic Trace

The Whittle Estimator

The result obtained using the Whittle Estimator method proves that the data has long memory or it is long-range dependence. The estimated Hurst parameter is between ***0.96 and 0.97***.

LRD in Packet Round Trip Delays

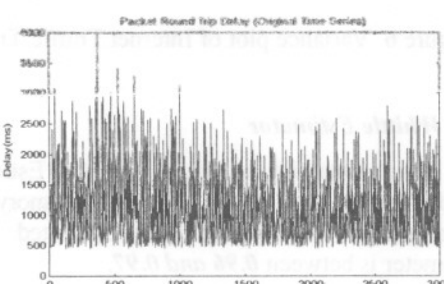
The calculation of packet delay needs *four* time stamps, namely, T_1, T_2, T_3 *and* T_4. When a computer sends out a packet, it records the *leaving* *time* (T_1) on the out going packet. When the packet gets to the peer, the peer records the *arrival time* (T_2) on the packet too; then the peer passes back the packet and records the *leaving time* (T_3) on the back packet. When the packet gets back to the sender, the sender records down *receiving time* (T_4). [9]

$$T = (T_4 - T_1) - (T_3 - T_2) \qquad (16)$$

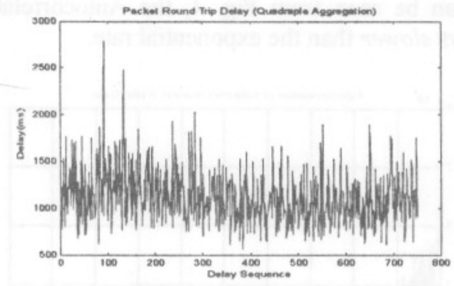
We used a custom developed program (*VBPing),* which uses TCP/IP's *Ping* Application to calculate the Packet Round Trip Delay. A packet round-trip delay is the *sum* of delays on each subnet link traversed by the packet. Each link (or hop) in turn consists of *four* components, including *Processing* delay, *Queueing* delay, *Transmission* delay and *Propagation* delay. So a packet Round Trip delay *process,* $T(t)$, is a *random variable* at time t. $T(t)$ describes the *packet round trip delay (Stochastic) process.* $T(t)$ can be studied through the characteristics of a time series T_i. It is obtained by *discretizing* $T(t)$ with *t*. Simply, T_i is a *sample process* of $T(t)$ at $t = t_i$ where $i = 1,2,3,........$ In our case t_i is the time at which the *ping* echo reply is received.

Scale Invariance and Autocorrelation

When the Packet Round Trip Delay (*or Round trip time or RTT)* is viewed in different aggregation levels, it still kept its *Bursty* nature. In other words, packet round trip delay is highly variable. This property is presented in Figs. 7*(a) and (b)*. It is obtained by *pinging* one of the *yahoo.com* servers (*IP Address: 64.58.79.230*).



(a)



(b)

Figure 7 Packet Round Trip Delay original series and its fourth aggregation

Figure 8 displays the Autocorrelation functions of the two Packet Round Trip Delay Processes. As can be seen, the Autocorrelation decays slowly (*as a Power Law)*, in contrast to exponential decay expected from traditional models.

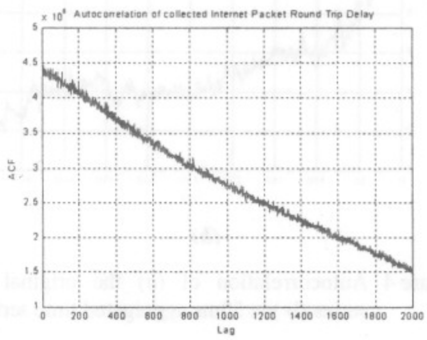
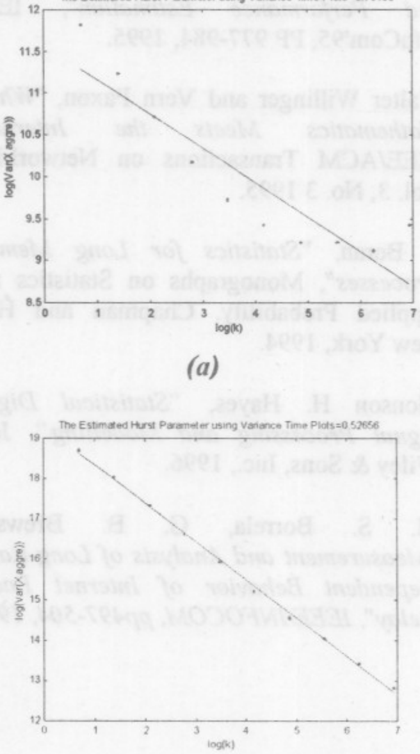


Figure 8 Autocorrelation of the Packet Round Trip Delay Process

To investigate the *effect of Network Load* on the LRD of Packet Round Trip Delays, we ping ETC's DNS server (*IP Address:* 196.27.22.43) during high traffic time (10-12AM) and negligible traffic (3AM-4AM). This hour demarcation is based on the local population Internet browsing trend, since the server is located in Addis Ababa. When the Hurst parameter is estimated for these different time series, there is a significant change. Results show a Hurst parameter near to *1* during busy times or *High Network load* and near *0.5* during negligible traffic times.

Figures 9 displays the Hurst parameter estimation results using Variance-Time plot. As can be seen from results displayed in Fig. 9 *(a) and (b)* there is a variation in the intensity of self-similarity, as measured by the Hurst parameter, for the *busy* and *less busy* periods.



(a)

(b)

Figure 9 Hurst Parameter estimates of the Packet Round Trip Delay of the ETC's DNS Server (a) During busy hours, H=0.79403 (b) During less busy hours, H=0.52656

CONCLUSIONS AND RECOMMENDATIONS

In this paper, Study of Internet Traffic and Packet Round Trip Delays are covered. With proper statistical and mathematical tools and computer programs, the following main findings are taken out as results.

- *Internet traffic* and *Packet Round Trip Delay processes,* as seen as a time series, are *statistically Self-Similar.* The autocorrelation function doesn't decay exponentially, a slow decay is observed. Hence the data sets have *Long Range Dependence* (LRD).

- The degree of *Self-Similarity*, as measured by the Hurst Parameter, of Packet Round Trip delay process is dependent on the Network Load. As the utilization increases the Hurst parameter will tend to be close to 1.

Important implication of the long-range dependence and scale invariant "burstiness" of the Internet traffic is drastically different from both conventional telephone traffic and from stochastic models for packet traffic usually considered in the literature (Example is the *Poisson* Model). Traffic models for voice traffic, developed over the years to serve the telephone network, did not apply as might have been hoped because voice traffic does not give rise to the same traffic characteristics as Internet data traffic, which is *burstier*.

With the help of these findings one can extend this work in the following areas:

- *Physical modelling of self-similarity:* Developing physical models that can explicate traffic characterstics in terms of elementary, verifiable system properties and network mechanics.

- *Traffic control for self-similar traffic*: It can be exploited on two fronts: One as extension of performance analysis in the resource provisioning context and the other from the multiple time scale traffic control perspectives where correlation structure at large time scales actively exploited to improve network performance.

- *Congestion control for self-similar network traffic*: This can be facilitated by using the long-term correlation structure present in long-range traffic for congestion control purposes.

- *Wavelet analysis of self-similar or scaling phenomenon*: Due to their ability to localize a given signal or time series both in time and scale (or frequency), wavelets provide a powerful and refined technique for detecting and quantifying scaling behavior in measured traffic.

- *Self-similarity and Network performance:* How a Queueing system perform when Packet arrivals taken from Self-Similar processes. Also effect of self-similarity on network Transport protocols, such as TCP throughput performance can be studied.

ACKNOWLEDGMENTS

We would like to thank colleagues and friends at Ethiopian Telecommunication Corporation and Icon Networks for their support during packet trapping.

REFERENCES

[1] Taye Abdulkadir, "Internet Self-Similarity, Modelling and Performance Evaluation", a thesis submitted for a partial fulfillment of M.Sc. in Electrical Engineering, Addis Ababa University, January 2003.

[2] W. E. Leland, M. S. Taqqu, and D. V. Wilson, *"On the self similarity nature of Ethernet Traffic"(Extended Version)*", IEEE/ACM Transactions on Networking, 1994.

[3] B. B. Mandelbrot,"*The Fractal Geometry of Nature*", Freeman, New York, 1983.

[4] "*Self Similar Network Traffic and Performance Evaluation*", Edited by K. Park and W. Willinger, John Wiley and Sons, 2000.

[5] R. Addie, M. Zukerman, and T. Neame, "*Fractal Traffic, Measurement, Modelling and Performance Estimation*", IEEE InfoCom'95, PP 977-984, 1995.

[6] Walter Willinger and Vern Paxon, "*Where Mathematics Meets the Internet*", IEEE/ACM Transactions on Networking, Vol. 3, No. 3 1995.

[7] J. Beran, "*Statistics for Long Memory Processes*", Monographs on Statistics and Applied Probability, Chapman and Hall, New York, 1994.

[8] Monson H. Hayes, *"Statistical Digital Signal Processing and Modelling"*, John Wiley & Sons, Inc., 1996.

[9] M. S. Borrela, G. B. Brewster, *"Measurement and Analysis of Long-Range Dependent Behavior of Internet Packet Delay", IEEE/INFOCOM, pp497-504, 1997.*