INTERNET TRAFFIC MODELING USING THE INDEX OF VARIABILITY

Georgios Y. Lazarou
TITL
Mississippi State University
glaz@ece.msstate.edu

Xiangdong Xia
TITL
Mississippi State University
xx4@ece.msstate.edu

Victor S. Frost
ITTC
University of Kansas
frost@eecs.ukans.edu

ABSTRACT

In this paper, we propose a novel and mathematically rigorous measure of variability, called the *index of variability* ($H_v(\tau)$), that fully and accurately captures the degree of variability of a typical network traffic process at each time scale and is analytically tractable for many popular traffic models. Using this proposed measure, we then analyzed two traffic models: the Two-State Markov Modulated Poisson Process (MMPP) and the renewal process with hyperexponential interarrival time distributions of order two (RPH2). Two-state MMPP models are popular in modeling the superposition of packet voice streams. The results show that the traffic variability can exhibit a non-monotonic behavior. In addition, the results suggest that renewal processes with interarrival times hyperexponentially distributed are suitable for modeling network traffic processes with high variability over a broad range of time scales.

KEY WORDS

Network Traffic Modeling, Variability, Burstiness

1 Introduction

Many empirical studies have shown that Internet traffic exhibits high variability[1] [1]–[4]. That is, traffic is bursty (variable) over a wide range of time scales in sharp contrast to the assumption that traffic burstiness exists only at short time scales while traffic is smooth at large time scales [3]. High variability in traffic has been shown to have a significant impact on network performance [3], [5]. The results from [5]–[8] show that knowledge of the traffic characteristics on multiple time scales helps to improve the efficiency of traffic control mechanisms. Importantly, the design and provision of quality-of-service-guarantees over the Internet requires the understanding of traffic characteristics, such as variability.

Since the publication of [3], the popular belief is that the high variability in traffic is due to the *long-range dependence*(LRD) property of the traffic processes. In general, a (weakly) stationary discrete-time real-valued stochastic process $Y = \{Y_n, n = 0, 1, 2, \ldots\}$ with mean $\mu = E[Y_n]$ and variance $\sigma^2 = E[(Y_n - \mu)^2] < \infty$ is long-range dependent if $\sum_{k=1}^{\infty} r(k) = \infty$, where $r(k)$ measures the correlation between samples of Y separated by k units of time. If $\sum_{k=1}^{\infty} r(k) < \infty$, then Y is said to exhibit *short-range dependence* (SRD).

Common traffic models with LRD are based on self-similar processes. In traffic modeling, the term self-similarity is usually used to refer to the *asymptotically second order self-similar* or *mono-fractal* processes [9]. The definition of asymptotically second order self-similarity is as follow [3]: assume that Y has an autocorrelation function of the form $r(k) \sim k^{-\beta}L(k)$ as $k \to \infty$, where $0 < \beta < 1$ and the function L is slowly varying at infinity, i.e., $\lim_{k\to\infty} \frac{L(kx)}{L(k)} = 1\ \forall x > 0$. For each $m = 1, 2, 3, \ldots$, let $Y^{(m)} = \{Y_n^{(m)}, n = 1, 2, 3, \ldots\}$ denote a new aggregated time series obtained by averaging the original series Y over non-overlapping blocks of size m, replacing each block by its sample mean. That is, for each $m = 1, 2, 3, \ldots$, $Y^{(m)}$ is given by

$$Y_n^{(m)} = \frac{Y_{nm-m+1} + \cdots + Y_{nm}}{m} \qquad n \geq 1. \qquad (1)$$

The new aggregated discrete-time stochastic process $Y^{(m)}$ is also (weakly) stationary with an autocorrelation function $r^{(m)}(k)$. Then, Y is called asymptotically second order self-similar with self-similar parameter $H = 1 - \frac{\beta}{2}$ if for all k large enough, $r^{(m)}(k) \to r(k)$ as $m \to \infty$. That is, Y is asymptotically second-order self-similar if the corresponding aggregated processes $Y^{(m)}$ become indistinguishable from Y at least with respect to their autocorrelation functions. By definition, asymptotically second order self-similarity implies LRD and vice versa [9].

The parameter H is called the *Hurst parameter*. For general self-similar processes, it measures the degree of "self-similarity". For random processes suitable for modeling network traffic, the Hurst parameter is basically a measure of the speed of decay of the tail of the autocorrelation function. And if $0.5 < H < 1$, then the process is LRD, and if $0 < H \leq 0.5$, then it is SRD. Hence, H is widely used to capture the intensity of long-range dependence of a traffic process, the closer H is to 1 the

[1] Fluctuation of traffic as a function of time.

more long-range dependent the traffic is, and vice versa [9].

There are several methods for estimating H from a traffic trace. One of the most widely used is the *Aggregated Variance* method: for successive values of m that are equidistant on a log scale, the sample variance of $Y^{(m)}$ is plotted versus m on a log-log plot [10], [11]. By fitting a least-square line to the points of the plot and then calculating its slope, an estimate of the Hurst parameter is obtained as $\hat{H} = 1 - \frac{slope}{2}$.

Another very popular method is based on wavelets [12]. Given a traffic trace Y_n, the Hurst parameter can be estimated as follows. For each scale j, the wavelet energy $\mu_j = \frac{1}{N_j}\sum_{k=1}^{n_j} d^2(j,k)$ is plotted versus j on a semi-log plot (i.e., $log_2(\mu_j)$ vs. j). By fitting a least-square line to the points of the curve region that *looks* linear and then computing its slope α, H is estimated as $\hat{H} = \frac{\alpha+1}{2}$.

1.1 Need for a New Measure of Variability

Commonly used measures of traffic burstiness, such as the peak-to-mean ratio, the coefficient of variation of interarrival times, the indices of dispersion for intervals and counts, and the Hurst parameter, do not capture the fluctuation of variability over different time scales.

It is claimed in [3] that the Hurst parameter is a good measure of variability, and the higher the value of H, the burstier the traffic. However, we believe that the Hurst parameter does not accurately capture the variability of network traffic over all performance relevant time scales. The popular belief from early studies [5], [13]–[15] on the impact of LRD on network performance is that high values of the Hurst parameter are associated with poor queueing performance. But, later studies [7], [16] show examples in which larger values of H are associated with better queueing performance compared to smaller values of H. In addition, the results in [8] indicate that the queueing performance depends mostly on the variability over certain time scales rather than on the value of H. Moreover, it is known [6] that different long-range dependent processes with the same value of the Hurst parameter can generate vastly different queueing behavior. Clearly, the single value Hurst parameter does not capture the fluctuation of the degree of traffic burstiness across time scales, regardless if the traffic process exhibits LRD or SRD.

For many network traffic processes, the wavelet energy-scale or variance-time plots usually do not tend to straight lines, i.e., see Fig. 1 (For information about the Auckland traffic traces, see [17].). Usually many of these processes have piecewise fractal behavior with varying Hurst parameter over some small ranges of time scales [18]. Such processes are usually referred to as multi-fractal processes [19].

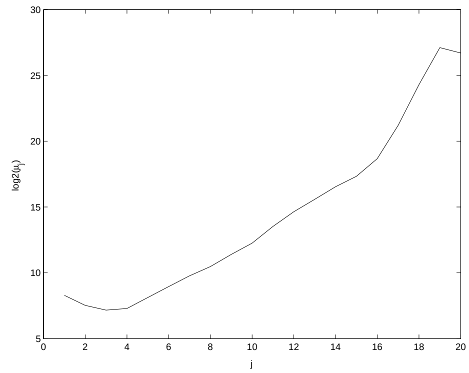

Figure 1. $log_2(\mu_j)$ versus scale j for the Auckland-IV traffic trace 20010301-310-0.

Queueing performance greatly depends on traffic irregularities at small time scales which are believed to be due to the complex dynamics of data networks [6], [20]. Multifractal analysis based on the legendre spectrum is often used to study the multiscaling behavior of traffic at small time scales [18], [21]–[23]. The process of estimating the legendre spectrum involves higher order sample moments and negative values of moments. It is known [24] that higher order sample moments are not well-behaved and negative values of moments tend to be erratic. In addition, the legendre spectrum is difficult to interpret [25].

Hence, there is a need for an intuitively appealing, coneptually simple, and mathematically rigorous measure which can capture the various scaling phenomena that are observed in data networks on both small and large scales [26].

In this paper, we propose a novel measure of variability, called the *index of variability* ($H_v(\tau)$), that fully and accurately captures the degree of variability of a typical network traffic process at each time scale and is analytically tractable for many traffic models.

The rest of this paper is organized as follows: In Section II, we define the *index of variability*. In Section III, we analytically obtain the index of variability curves for two traffic models, and we conclude in Section V.

2 Index of Variability for Packet Traffic Sequences

Let $N(t)$ denote the number of events (packet arrivals) of a stationary point process in the interval $(0, t]$. For each fixed time interval $\tau > 0$, an event count sequence $Y = \{Y_n(\tau), \tau > 0, n = 1, 2, \ldots\}$ can be constructed from each point process, where

$$Y_n(\tau) = N[n\tau] - N[(n-1)\tau] \quad (2)$$

denotes the number of events that have occurred during the n^{th} time interval of duration τ. Clearly, Y is also (weakly) stationary for all $\tau > 0$. In this study, Y represents a network traffic trace where $Y_n(\tau)$ denotes the number of packets observed from an arbitrary point in the network during the n^{th} time interval of duration τ. We refer τ as the *time scale* of the traffic trace, and it represents the length (i.e., 10ms, 1s, 10s, e.t.c.) of one sample of Y.

The expected number of events that have occurred during the interval $(0, t]$ is always: $E[N(t)] = \frac{t}{E[X]} = \lambda t$ where $E[X]$ is the expected interarrival time and λ is the mean event (packet) arrival rate. The index of dispersion for counts (IDC) is defined as: $IDC(t) \equiv \frac{Var[N(t)]}{E[N(t)]} = \frac{Var[N(t)]}{\lambda t}$. The IDC was defined such that it provides some comparison with the Poisson process, for which $IDC(t) = 1 \; \forall t$. Note that since the point process is stationary, IDC has the same value over any interval of length t. Hence, t can be viewed as the time scale τ of the traffic process Y defined in (2). From now on we will be using t to denote generality and τ to denote time scales, i.e., the time length of each sample of the packet-count sequence Y.

An important feature of IDC is that it is mathematically equivalent to the Aggregated Variance method for estimating the Hurst parameter H of a self-similar process. For a self-similar process, plotting $log(IDC(m\tau))$ against $log(m)$ results in an asymptotic straight line with slope $2H - 1$. When Y is a long-range dependent process, the slowly decaying variance property of LRD processes [3] with parameter $0 < \beta < 1$ is equivalent to an IDC curve[2] with an asymptotic straight line with slope $1 - \beta$, implying $0 < slope < 1$. When the IDC curve converges to an asymptotic straight line with $slope = 0$ for some $\tau < \infty$, then Y is a short-range dependent process. Based on the above property of IDC, we define the following new measure of variability:

Definition 1 *For a general stationary traffic process Y as defined by (2) whose $IDC(\tau)$ is continuous and differentiable over $(0, \infty)$, we call*

$$H_v(\tau) \equiv \frac{\frac{d(log(IDC(\tau)))}{d(log(\tau))} + 1}{2} \tag{3}$$

the index of variability of Y for the time scale τ, where $\frac{d(log(IDC(\tau)))}{d(log(\tau))}$ is the local slope of the IDC curve at each τ when plotted in log-log coordinates.

Note that the index of variability is so defined in order that for a long-range dependent (asymptotically or second-order self-similar) process $H_v(\tau) = H \in (0.5, 1)$ for all $\tau \geq \tau_o > 0$. The value of τ_o depends on the particular process. If the process is exactly self-similar then $H_v(\tau) = H \in (0.5, 1)$ for all $\tau > 0$. That is, if $log(IDC(\tau))$ is linear with respect to $log(\tau)$), then $H_v(\tau)$ reduces to H. The Index of Variability can be thought of as the Hurst parameter defined at each time scale.

[2] In log-log coordinates.

In general[3], the process Y exhibits significant variability for those time scales τ such that $0.5 < H_v(\tau) < 1$. When $\frac{d(log(IDC(\tau)))}{d(log(\tau))} \rightarrow 1$, then $H_v(\tau) \rightarrow 1$ implying very high variability. A plot of $H_v(\tau)$ versus τ would depict the behavior of the traffic process Y in terms of variability (burstiness) at each time scale τ $(= 10ms, \; 100ms, \; 1s, \; \ldots)$.

Expanding the local slope of the IDC curve at each time scale, we get

$$\begin{aligned} \frac{d(log(IDC(\tau)))}{d(log(\tau))} &= \frac{\tau}{IDC(\tau)} \frac{d(IDC(\tau))}{d\tau} \qquad (4) \\ &= \frac{\tau}{Var[N(\tau)]} \frac{d(Var[N(\tau)])}{d\tau} - 1. \end{aligned}$$

Using the above in (3), we obtain a more convenient form of the Index of Variability:

$$\begin{aligned} H_v(\tau) &= 0.5\tau \left(\frac{\frac{dVar[N(\tau)]}{d\tau}}{Var[N(\tau)]} \right) \\ &= \frac{1}{2} \left\{ 1 + \tau \left(\frac{\frac{d(IDC(\tau))}{d\tau}}{IDC(\tau)} \right) \right\} \qquad (5) \end{aligned}$$

In addition, setting $\tau = mT$, where $T > 0$ and $m = 1, 2, \cdots$, and using the relation $Var[Y^{(m)}] = \frac{Var[N(mT)]}{m^2}$, we can express the index of variability function in terms of $Var[Y^{(m)}]$ versus m:

$$H_v(mT) = 0.5m \frac{\frac{dVar[Y^{(m)}]}{dm}}{Var[Y^{(m)}]} + 1. \tag{6}$$

Suppose now Y is an aggregate sequence of packet counts resulting from the superposition of M independent packet-traffic sources, not necessarily identical. Then $N(t) = N_1(t) + \cdots + N_M(t)$, where $N_i(t)$ denotes the number of packet arrivals in the interval $(0, t]$ from the i^{th} traffic source. Assuming again stationarity, we have

$$IDC(t) = \frac{\sum_{i=1}^{M} Var[N_i(t)]}{\sum_{i=1}^{M} \lambda_i t} = \sum_{i=1}^{M} \left(\frac{IDC_i(t)}{\Lambda_i} \right) \tag{7}$$

where λ_i is the mean packet arrival rate from the i^{th} source, and $\Lambda_i = \frac{\sum_{j=1}^{M} \lambda_j}{\lambda_i}$. In addition, $\frac{log(IDC(t))}{log(t)} = \frac{log(\sum_{i=1}^{M} Var[N_i(t)])}{log(t)} - \frac{log(\sum_{i=1}^{M} \lambda_i t)}{log(t)}$, and upon taking the derivative in respect to log(t) we get the index of variability for the aggregate traffic stream to be

$$H_v(\tau) = 0.5\tau \left(\frac{\sum_{i=1}^{M} \frac{dVar[N_i(\tau)]}{d\tau}}{\sum_{i=1}^{M} Var[N_i(\tau)]} \right)$$

[3] The generality here is confined for those processes that are suitable in modeling network packet traffic.

$$= \quad \frac{1}{2}\left\{1+\tau\left(\frac{\sum_{i=1}^{M}\frac{d(IDC_i(\tau))}{d\tau}\left(\frac{1}{\Lambda_i}\right)}{\sum_{i=1}^{M}\left(\frac{IDC_i(\tau)}{\Lambda_i}\right)}\right)\right\} \quad (8)$$

As we can observe from (8), the variances or the indices of dispersion for counts of the M independent point-processes completely characterize the variability function of the aggregate packet-count sequence Y. If $\lim_{\tau\to\infty} IDC(\tau) = \lim_{\tau\to\infty}\left(\sum_{i=1}^{M}\left(\frac{IDC_i(\tau)}{\Lambda_i}\right)\right) = c < \infty$, then obviously, $\lim_{\tau\to\infty} H_v(\tau) = 0.5$. In case that all M underlying point processes of making up Y are also identical, then (8) reduces to (5). If all M underlying point processes are Poisson, then $\frac{d(IDC_i(\tau))}{d\tau} = 0$ for all τ and i and hence $H_v(\tau) = 0.5$ for all τ.

3 Analysis of Traffic Models

In this section, we obtain the index of variability functions for two traffic models: the Two-State Markov Modulated Poisson Process (MMPP) and the renewal process with hyperexponential interarrival time distributions of order two (RPH2). MMPP models are very popular traditional model which yield SRD traffic processes. Two-state MMPP models are popular in modeling the superposition of packet voice streams [27]. As shown below, such models can be used to capture the high variability of traffic over a range of small time scales.

The work in [28] shows that long-tail distributions can be approximated by hyperexponentional distributions. Thus, we believe that renewal processes with hyperexponential interarrival time distributions are suitable for capturing the high variability of traffic over any range of (short or long) time scales. In addition, a major advantage of these models is their relative ease of analytically obtaining queueing performance predictions.

3.1 Two-State Markov Modulated Poisson Process (MMPP)

Here we consider that the underlying point process of Y is an MMPP with two-state Markov chain where the mean sojourn times in state 1 and 2 are α^{-1} and β^{-1}, respectively. When the chain is in state i ($i = 1, 2$) the point process is Poisson with rate λ_i. Letting $\rho = \alpha + \beta$ and $\upsilon = \lambda_1\beta + \lambda_2\alpha$, we have from [27] that $E[N(t)] = \frac{\upsilon t}{\rho}$ and $IDC(t) = 1 + \rho A - A\left(\frac{1-e^{-\rho t}}{t}\right)$, where $A = \frac{2\alpha\beta(\lambda_1-\lambda_2)^2}{\rho^3\upsilon}$. It is easy to see that $\lim_{t\to\infty} IDC(t) = 1 + \rho A$. Upon taking the derivative of IDC(t) we obtain the index of variability of Y as

$$H_v(\tau) = 0.5\left\{1 + \frac{A\left[1-(1+\rho\tau)\,e^{-\rho\tau}\right]}{(1+\rho A)\tau - A\left(1-e^{-\rho\tau}\right)}\right\}.$$

Numerical Results: Assume $\alpha^{-1} = \beta^{-1} = 100$ seconds, $\lambda_1 = 4$ packets/second and λ_2 to vary from 1 to 1000 packets/second. Plots of $H_v(\tau)$, are shown in Fig. 2. Notice that when $\lambda_2 = \lambda_1$, we have a pure Poisson process and therefore zero variability. But as the difference between λ_1 and λ_2 increases, so does the index of variability. Based on the two plots of Fig. 2, we observe that the index of variability increases with λ_2 up to its maximum value, and any further increase in λ_2 does not have any affect on variability. We also observe that the index of variability increases with τ up to its maximum value and then decays with an exponential rate. But most importantly, we observe that for values of λ_2 not very close to λ_1, the packet-count process Y has substantial variability over a wide range of time scales that spans about 200 seconds.

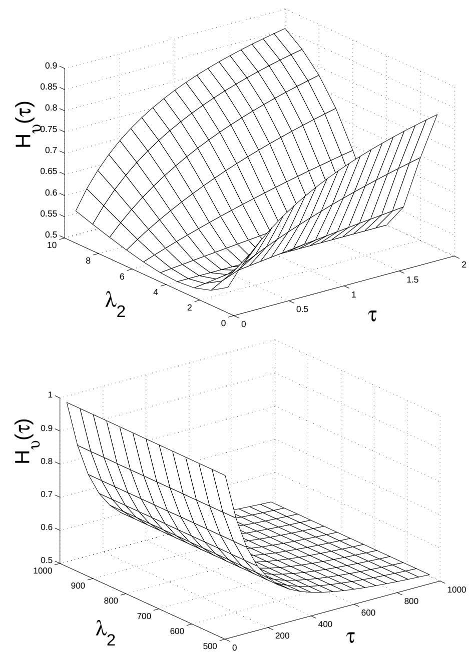

Figure 2. Index of Variability for The Two-State MMPP: $\alpha^{-1} = \beta^{-1} = 100$ Seconds, $\lambda_1 = 4$ Packets/Second.

3.2 Renewal Processes with Interarrival Times Hyperexponentially Distributed

We assume here that the underlying point processes of Y are stationary renewal processes with interarrival times hyperexponentially distributed. We call this model as the *hyperexponential model*. A hyperexponential distribution of order K, $(= 1, 2, 3, \ldots)$, is the weighted sum of K exponential distributions:

$$F_K(x) = Pr[X \le x] = \sum_{i=1}^{K} w_i(1 - e^{-\alpha_i x}) \qquad (9)$$

where $w_i > 0$ are the weights satisfying $\sum_{i=1}^{K} w_i = 1$, and $\alpha_i > 0$ are the rates of the exponential distributions

[33]. It was shown in [29] that if $w_i = w^i$ and $\alpha_i = \frac{\mu}{\eta^i}$ for $0 < w < 1$, $\eta > 1$, and $\mu > 0$, then the tail of the hyperexponential distribution gets longer and longer with K. The advantages of the hyperexponential distributions over heavy-tailed distributions like Pareto are two-fold: their Laplace transform exists, therefore they can be utilized in analytic models, and they have finite variance for all K. The hyperexponential distribution defined in Equation 9 is used in the simulation study in [30] for different values of K up to 64. In that study, a Poisson process is merged with a renewal process for which the interarrival times have this special hyperexponential distribution which was developed in [29]. For $K = 64$, $\alpha = -\frac{log_{10}(w)}{log_{10}(\eta)} = 1.4$, $\lambda_{Hyperexponential} = \frac{1}{E[X]} = 1$, and $\lambda_{Poisson} = [0.1, 0.7]$, the autocorrelation function appears linear for more than 8 orders of magnitude when plotted in log-log coordinates, suggesting LRD behavior over a very broad range of time scales. Here we analytically obtain the variability index of the traffic process Y only for the case of $K = 2$ for any values of w_i and α_i, $(i = 1, 2)$. We leave the cases of $K > 2$ for future work.

Since the underlying point processes of making up Y are independent from each other, it is sufficient to obtain the variance of the aggregate counting process $N(t)$, and thus the index of variability and autocorrelation function, by considering only a single renewal process with hyperexponential interarrival time distribution of order two.

Letting $a = \alpha_1$ and $b = \alpha_2$, the pdf of the interarrival times is then

$$f_2(x) = w_1 a e^{-ax} + w_2 b e^{-bx}. \quad (10)$$

The mean packet arrival rate is $\lambda = \frac{ab}{aw_2 + bw_1}$, and the squared coefficient of variation of the interarrival times is $\mathcal{C}^2(X) = 2\left[\frac{a^2 w_2 + b^2 w_1}{(aw_2 + bw_1)^2}\right] - 1$. In addition, $\lim_{w_2 \to 0} \mathcal{C}^2(X) = 1$ and $\lim_{b \to 0} \mathcal{C}^2(X) = \frac{2}{w_2} - 1$. As shown in Fig. 3, for constant values of a and b, $\mathcal{C}^2(X)$ increases exponentially up to its maximum value and then decreases to one very abruptly. The maximum value depends on the value of b, and as shown by these two plots, it can get extremely high.

It can be shown that [31]

$$Var[N(t)] = \frac{2\lambda[(aw_1 + bw_2)^2 - (a^2 w_1 + b^2 w_2)]}{(aw_2 + bw_1)^3} \left(1 - e^{-[aw_2 + bw_1]t}\right) + \lambda \mathcal{C}^2(X) t, \quad (11)$$

and

$$IDC(t) = \frac{2[(aw_1 + bw_2)^2 - (a^2 w_1 + b^2 w_2)]}{(aw_2 + bw_1)^3} \left(\frac{1 - e^{-[aw_2 + bw_1]t}}{t}\right) + \mathcal{C}^2(X). \quad (12)$$

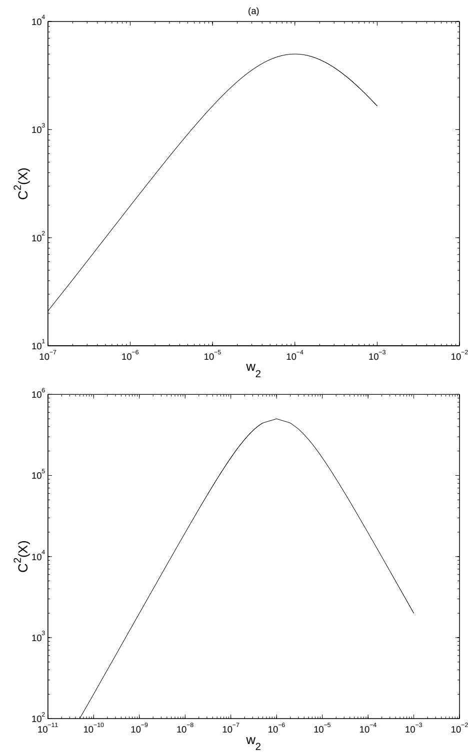

Figure 3. Squared Coefficient of Variation of the Interarrival Time vs. w_2 for the Case of Hyperexponential Distribution of Order Two: $a = 100$, (a) $b = 0.01$, and (b) $b = 0.0001$.

Observe that $\lim_{t \to \infty} IDC(t) = \mathcal{C}^2(X)$. Equation (11) or (12) can then be used in (5) to obtain the index of variability. It is obvious to see that $\lim_{\tau \to \infty} H_v(\tau) = 0.5$.

Numerical Results: Let $a = 100$. Table 1 lists the values of the mean packet rate (λ) and the squared coefficient of variation of the interarrival times ($\mathcal{C}^2(X)$) for $b = 0.01$ and $b = 0.0001$ for different values of w_2. Note that $w_1 + w_2 = 1$. Interesting, the maximum value of $\mathcal{C}^2(X)$ occurs when $\lambda = \frac{a}{2}$. Also, Fig. 4 indicates that at this value of λ the process attains the widest range of time scales of high variability, and in this range the index of variability reaches its maximum value (curve (i), maximum$H_v = 0.9988$). Observe that this widest range of time scales of high variability most likely covers all time scales that impact network performance evaluation [6]. In this example and for $\lambda = \frac{a}{2}$ packets/s, the range of time scales that the packet-count sequence Y exhibits high variability spans 7 order of magnitude. If this traffic process Y is observed only for these time scales, then Y will be viewed as long-range dependent (LRD) process.

In addition, Fig. 4 shows that the maximum value of variability as well as the range of time scales of substantial variability become smaller as $\lambda \to a$. Let

$$\tau_{on} = inf\{range\,of\,time\,scales\,of\,substantial\,variability\},$$

Table 1. Values of Mean Packet Rate (λ) and Squared Coefficient of Variation of Interarrival Times ($\mathcal{C}^2(X)$) for the Numerical Example of the Case of Hyperexponential Distribution of Order Two: $a = 100$.

	λ	(packets/sec)	$\mathcal{C}^2(X)$	
w_2	$b = 0.01$	$b = 0.0001$	$b = 0.01$	$b = 0.0001$
10^{-3}	9.1000	0.0999	1.6522x10^3	1.9950x10^3
10^{-4}	50.0000	0.9901	5.0000x10^3	1.9605x10^4
10^{-5}	90.9000	9.0909	1.6536x10^3	1.6529x10^5
10^{-6}	99.0000	50.0000	197.0202	5.0000x10^5
10^{-7}	99.9000	90.9091	20.9561	1.6529x10^5
10^{-8}	99.9900	99.0099	2.9992	1.9607x10^4
10^{-9}	99.9990	99.9001	1.2000	1.9970x10^3
10^{-10}	99.9999	99.9900	1.0200	200.9596
10^{-11}	100.0000	99.9990	1.0020	20.9996
10^{-12}	100.0000	99.9999	1.0002	2.9999
10^{-13}	100.0000	100.0000	1.0000	1.2001

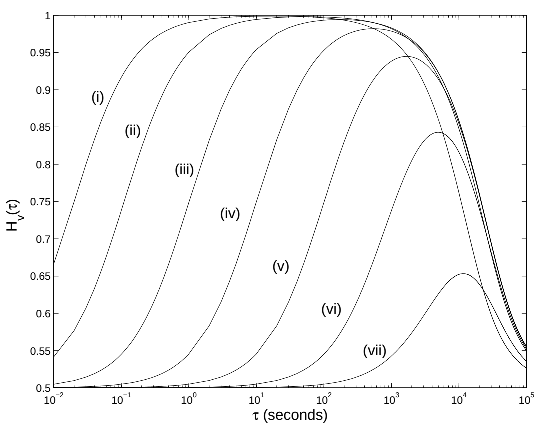

Figure 4. Index of Variability vs. Time Scale for the Case of Hyperexponential Distribution of Order Two: $a = 100$, $b = 0.0001$, $w_2 = $ (i) 10^{-6} (ii) 10^{-7} (iii) 10^{-8} (iv) 10^{-9} (v) 10^{-10} (vi) 10^{-11} (vii) 10^{-12}.

and

$$\tau_{off} = sup\{range\,of\,time\,scales\,of\,substantial\,variability\}.$$

As we can see from these curves, τ_{on} gets bigger as λ approaches a. Although it is not completely shown in Fig. 4, it is not difficult to see that τ_{off} becomes smaller as $\lambda \rightarrow b$. Notice that for all $\tau \ni [\tau_{on}, \tau_{off}]$ the process looks like Poisson.

Clearly, the above results show that the hyperexponential distribution can be used to model the interarrival distribution of highly bursty packet traffic. Specifically, the results indicate that this traffic model can capture traffic burstiness over a broad range of time scales. A major advantage over models that uses heavy-tailed distributions is that with this model the exact degree of variability at each time scale can be measured.

4 Conclusion

All commonly used measures of traffic burstiness do not capture the fluctuation of variability over different time scales. Therefore, we developed a novel and mathematically rigorous measure of variability, called the *index of variability* ($H_v(\tau)$), which can capture the various scaling phenomena that are observed in data networks on both small and large scales [26].

Using this proposed measure, we then analyzed two traffic models: the Two-State Markov Modulated Poisson Process (MMPP) and the renewal process with hyperexponential interarrival time distributions of order two (RPH2).

The results show that conventional traffic models can capture the high variability observed in network traffic over a considerable range of time scales. We show that a synthetic packet traffic process generated by using the two-state MMPP can exhibit a substantial variability over a wide range of time scales that spans 200 seconds. In addition, the results show that the index of variability can fully capture the multifractal behavior of traffic processes, especially at small time scales. The results also suggest that renewal processes with interarrival times hyperexponentially distributed are suitable for modeling highly bursty network traffic processes.

We are currently working in developing (a) a procedure for estimating the index of variability from empirically measured traffic traces and (b) a method of fitting analytically obtained index of variability curves from the hyperexponential model to the curves estimated from traffic traces.

References

[1] D. Duffy, A. McIntosh, M. Rosenstein, D. Wilson, Statistical Analysis of CCSN/SS7 Traffic Data from Working CCS Subnetworks, *IEEE JSAC*, 12(3), 1994.

[2] H. J. Fowler, W. E. Leland, Local Area Network Traffic Characteristics, with Implication for Network Congestion Management, *IEEE JSAC*, 9(7), 1991.

[3] W. Leland, M. Taqqu, W. Willinger, D. Wilson, On the Self-Similar Nature of Ethernet Traffic (Extended Version), *IEEE/ACM Trans. on Net.*, 2(1), Feb. 1994.

[4] V. Paxson, S. Floyd, Wide-Area Traffic: The Failure of Poisson Modeling, *IEEE/ACM Trans. on Net.*, 3(3), June 1995.

[5] A. Erramilli, O. Narayan, W. Willinger, Experimental Queueing Analysis with Long-Range Dependent Packet Traffic, *IEEE/ACM Trans. on Net.*, 4(2), April 1996.

[6] M. Grossgluser and J.-C. Bolot, On the Relevance of Long-Range Dependence in Network Traffic, *IEEE/ACM Trans. on Net.*, 7(5), Oct. 1999.

[7] K. R. Krishnan, A. L. Neidhardt, A. Erramilli, Scaling Analysis in Traffic Management of Self-Similar Processes, Proc. 15th ITC, Washington, D.C., USA, June 1997, 1087-1096.

[8] A. L. Neidhardt, J. L. Wang, The Concept of Relevant Time Scales and Its Application to Queueing Analysis of Self-Similar Traffic (or Is Hurst Naughty or Nice?), Bellcore, Red Bank, New Jersey, USA, Sept. 12, 1997.

[9] K. Park, W. Willinger, *Self-Similar Network Traffic And Performance Evaluation*, (New York: John Wiley & Sons Inc., 2000).

[10] J. Beran, *Statistics for Long-Memory Processes. Monographs on Statistics and Applied Probability* (New York: Chapman and Hall, 1994).

[11] M. S. Taqqu, V. Teverovsky, W. Willinger, Estimators for long-range dependence: an emperical study, *Fractals*, 3(4), 1995.

[12] D. Veitch, P. Abry, A Wavelet Based Joint Estimator for the Parameters of LRD, Special issue on Multiscale Statistical Signal Analysis and its Applications, *IEEE Trans. Info. Th.*, 45(3), April 1999.

[13] I. Norros, A storage model with self-similar input, *Queueing Systems* 16, 1994, 387-396.

[14] W. E. Leland, D. V. Wilson, High Time-Resolution Measurement and Analysis of LAN Traffic: Implication for LAN Interconnection, *INFOCOM*, 3, 1991, 11d.3.1.

[15] A. Baiocchi, N. Blefari Melazzi, M. Listanti, A. Roveri, R. Winkler, Modeling Isssues on an ATM Multiplexer Within a Bursty Traffic Environment, *INFOCOM*, 1, 1991.

[16] K. R. Krishnan, A New Class of Performance Results for a Fractional Brownian Traffic Model, *Queueing Systems*, 22, 1996, 277-285.

[17] NLANR Network Traffic Packet Header (Long) Traces, [Online] Available: http://pma.nlanr.net/Traces/long/

[18] Q. Li, D. Mills, Investigating the Scaling Behavior, Crossover and Anti-persistence of Internet Packet Delay Dynamics, *Globecom* 3, 1999, 1843-1852.

[19] V.J. Ribeiro, R.H. Reidi, R.G. Baraniuk, Wavelets and Multifractals for Network Traffic Modeling and Inference, *ICASSP*, 2001.

[20] K. Kant, On aggregate Traffic Generation with Multifractal Properties, *Globecom* 2, 1999, 801-804.

[21] R.H. Riedi, M.S. Crouse, V.J. Ribeiro, and R.G. Baraniuk, A Multifractal Wavelet Model with Application to Network Traffic, *IEEE Trans. on Information Theory* 45(3), April 1999.

[22] J. Levy Vehel, B. Sikdar, A Multiplicative Multifractal for TCP Traffic, *Proceedings of Sixth IEEE Symposium on Computers and Communications*, 2001, 714-719.

[23] A.C. Gilbert, W. Willinger, A. Feldmann, Scaling Analysis of Conservative Cascades, with Applications to Network Traffic, *IEEE Trans. on Information Theory*, 45(3), April 1999.

[24] M.S. Taqque, Vadim Teverovsky, W. Willinger, Is Network Traffic Self-Similar or Multifractal? *Fractals*, 5, 63-73.

[25] A. Horvath, M. Telek, A Markovian Point Process Exhibiting Multifractal Behaviour and its Application to Traffic Modeling, *Fourth International Conference Matrix Analytic Methods in Stochastic Models*, Adelaide, Australia, 14-18 July.

[26] W. Willinger, V. Paxson, R.H. Riedi, M. Taqqu, Long-Range Dependence and Data Network Traffic, *Long-range Dependence: Theory and Applications*, P. Doukhan, G. Oppenheim, M.S. Taqqu, eds., Birkhauser, 2002.

[27] H. Heffes, D.M. Lucantoni, A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance, *IEEE Journal on Selected Areas in Communications*, SAC-4(6), Sept. 1986, 856-868.

[28] A. Feldman, W. Whitt, Fitting Mixtures of Exponentials to Long-Tail Distributions to Analyze Network Performance Models, AT&T Laboratory – Research, December 1996.

[29] M. Greiner, M. Jobmann, L. Lipsky, The Importance of Power-Tail Distributions for Telecommunications Traffic Modeling, *Technical Report*, Institute für Informatik, Technische Universität München, München, Germany, 1995.

[30] P. M. Fiorini, Modeling Telecommunication Systems with Self-Similar Data Traffic, Ph.D. Dissertation, University of Connecticut, Sept. 9, 1998.

[31] G. Y. Lazarou, On the Variability of Internet Traffic, Ph.D. Dissertation, Electrical Engineering and Computer Science, University of Kansas, 2000.

[32] D. R. Cox, Renewal Theory, (London: Methuen: New York: J. Wiley & Sons Inc., 1962).

[33] Riccardo Gusella, Characterizing the Variability of Arrival Processes with Indexes of Dispersion, *IEEE JSAC*, 9(2), Feb. 1991.