A FEASIBILITY ANALYSIS OF NETWORK TRAFFIC FORECAST BASED ON FRACTAL CHARACTERISTICS

XIN WANG, BINXING FANG, XIAOCHUN YUN

Research Center of Computer Network and Information Security Technology,
Harbin Institute of Technology of China, Harbin 150001, China

Abstract: - Since fractal characteristics of network traffic were discovered by Leland in 1994, the literature is both well developed and skeptical about the value of traditional time series analysis on network data. In this paper, we investigate usability of traffic prediction based on fractal characteristics especially in the influence of assuming condition, model parameter with fractional predictors and burstiness behavior. Finally, some comments on prediction based fractal characteristics are given.

Key-words: - Traffic prediction; Fractional traffic model; Long-range dependence

1 Introduction

The ability to forecast bandwidth requirement in a next control time interval within a network is one of the fundamental requirements of network security and management, for example, to predict traffic bursts or information flows on networks before they arrive rather than having to adjust for them afterwards. Proverbially it has been used in various fields of network application, such as [1] [2] [3]. In recent years, there have been interested in research and development for traffic modeling to forecasting the upper bound of network traffic volume. Traffic prediction requires accurate traffic models which can capture the statistical characteristics of actual traffic. A number of high-quality, high-resolution measurements of Internet traffic have been carried out and analyzed. Last decade there has been a significant change in the understanding of network traffic. It has been demonstrated in numerous studies that traffic in high-speed networks exhibits presence of fractal features, viz. self-similarity (SS), long-range dependence (LRD), slowly decaying variances, heavy-tailed distributions and fractal dimensions [4], that can not be captured by previous models [5].

The quality of a forecast depends on the amount of uncertainty that accompanies the prediction, and mathematically this is measured by the variance of the prediction error. This uncertainty depends on a number of factors, including the amount of traffic history that is used to make the prediction, the prediction horizon and the nature of the traffic itself [6]. Obviously, the prediction interval has significant impact on the performance of the predictor, for real-time applications, prediction must be made rapidly from the minimum amount of stored traffic. Also, the prediction interval must be designed to include the time required to process the control information and the round-trip delay experienced before the information becomes effective. At the same time, prediction must be as accurate as possible so that bandwidth and buffer resources are not wasted, and the overall quality provided by the traffic management functions is maintained.

As the term implies, long-range dependence refers to a correlation structure that decays at a rate much slower than the exponential decrease that occur in the correlations of a short-range dependent process. The correlation structure of network traffic that accompanies long-range dependence means that the traffic exhibits sustained burstiness. From a mathematical point of view, for a stationary LRD process $X = \{X_t : t \geq 1\}$ with mean μ, variance σ^2 and autocorrelation function $r(k)$, which have $r(k) \sim k^{2H-2}$, as $k \to \infty$, means that the correlations are non-summable. In contrast, a short-range dependent process would have a correlation function that decreased according to $r_{SRD}(k) \sim a^k$, as $k \to \infty$, the Hurst parameter H is the index of self-similarity, where the invariance in the distribution of the process is defined as $X_{at} \overset{d}{=} a^H X_t$ and when the parameter H lies in the interval (0.5, 1.0) the resulting self-similarity process exhibits LRD[9]. There have been great expectations how forecast could utilize those properties, and the theoretical basis for predictors of long-range dependent traffic can be found in [2] [7] [8]. However the problem with LRD models is that they are required too many parameters, so that computationally complex. Their fitting procedure consumed a great deal of time while their parameters can not be estimated based on the real-time measurements. In this paper, we would investigate case of traffic forecast based on various fractal models. The significant conclusion is that the predictors based on fractal models fail to perform significantly forecast over the sampling of higher frequencies.

The paper is organized as follows. In section 2 diffusely employ LRD stochastic models applicable to traffic modeling and forecast is reviewed. Section 3 discusses issues of the feasibility analysis about the influence on assuming condition, model parameter with fractional predictors and burstiness behavior to traffic forecast. Finally, some conclusions are drawn in section 4.

2 Related Work about Predictor Based on Fractional Model

In this section we would summarize a few predictors based on long-range dependence models which are diffusely employed in theory and practice.

2.1 Fractional Gaussian Noise Prediction

Assuming network traffic is fractional Brownian traffic, unfortunately, that a fBm process defined does not have stationary increments, although fractional Brownian motion is useful for theoretical analysis, but it's not utilized with facility in practice. A later refinement of this model is the fractional Brownian motion process introduced by Mandelbrot and Van Ness [11], this process has stationary increments $Z_H(t)$ called *Fractional Gaussian Noise*, and again H is the *Hurst parameter*.

$$X_H(t) = Z_H(t+1) - Z_H(t), n \in Z^+$$

Norros [8] derived equations for short term prediction of fractional Gaussian noise traffic. In their work predictor is an integral over the observed part of the process in the form

$$\hat{X}_{h,T} = \int_{-T}^{0} gT(h,t)dX_t$$

Where $\hat{X}_{h,T}$ is the predicted value on the basis of observing $X(t-T,s)$ for $s \in [t-T,t]$ and $gT(h,t)$ is an appropriate weight function and is given by

$$gT(h,-t) = \frac{\sin(\pi(H-\frac{1}{2}))}{\pi} t^{0.5-H}(T-t)^{0.5-H} \times \int_0^h \frac{\sigma^{H-0.5}(\sigma+T^{H-0.5})}{\sigma+t} d\sigma$$

for $T < \infty, t \in (o,T)$ and for $T \to \infty, t > 0$

$$g_\infty(h,-t) = \frac{\sin(\pi(H-\frac{1}{2}))}{\pi} t^{0.5-H} \int_0^a \frac{\sigma^{H-\frac{1}{2}}}{\sigma+t} d\sigma$$

$$= \frac{\sin(\pi(H-\frac{1}{2}))}{\pi} \left(\frac{1}{H-\frac{1}{2}}\left(\frac{t}{h}\right)^{0.5-H} - B_{\frac{h}{(t+h)}}\left(H-\frac{1}{2}, \frac{3}{2}-H\right)\right)$$

Where $B(\cdot,\cdot)$ is the incomplete beta function. The function $gT(h,\cdot)$ is a solution of the integral equation

$$(2H-1)\int_0^T gT(h,-t)|t-s|^{2H-2}\, dt = (h+s)^{2H-1} - s^{2H-1}, s \in (0,T)$$

and has the scaling property

$$gT(h,t) = gT/h(1,\frac{t}{h})$$

As an application, literature [7] obtain an expression for the variance of the predictor $E[X_h \mid X_s, s \in (-T,0)]$:

Corollary 1 For $h > 0, T \in (0,\infty]$ we have

$$D^2(E[X_h \mid X_s, s \in (-T,0)]) = D^2(X_h)H\int_0^{T/h} gT/h(1+s)^{2H-1} - s^{2H-1})ds$$

For $T \to \infty$, the result can be stated in terms of the gamma function:

$$D^2(E[X_h \mid X_s, s \le 0]) = D^2(X_h)\left(1 - \frac{\sin(\pi(H-\frac{1}{2}))}{\pi(H-\frac{1}{2})} \frac{\Gamma(\frac{3}{2}-H)^2}{\Gamma(2-2H)}\right)$$

2.2 Fractional ARIMA Model Prediction

Recent investigations on the long-range dependence of computer network traffic have showed the complicated correlation structures of this traffic. To better model the LRD process, FARIMA was proposed for modeling and predicting the LRD process since it can better describe the slow decrease of its autocorrelation function [12][13]. The FARIMA *(p, d, q)* process proposed by Hosking in 1980 is an extension to ARIMA (*p, d, q*), is defined as[14]

$$\Phi(B)\nabla^d X_t = \theta(B)\varepsilon_t \qquad (1)$$

Where *d* is the indicator for the strength of LRD and assumes the value between 0 and 1/2. ε_t is a Gaussian white noise, and

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \cdots \Phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots \theta_q B^q$$

are polynomials of degree *p* and *q*, respectively, in the backward shift operator *B*. The operator $\nabla^d = (1-B)^d$ can be expressed by the binomial expansion

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k$$

$$\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} \qquad (2)$$

Where $\Gamma(x)$ denotes the gamma function; note for all positive integers, only the first *d+1* terms are non-aero in Eq. (2).

To a FARIMA *(p ,d, q)* process $\{X_t : t = 0,1,2,\ldots\}$, assumptions of causality and invertibility allow to write[15] $X_t = \sum_{i=0}^{\infty} \psi_i a_{t-i}$ and $a_t = \sum_{i=0}^{\infty} \pi_i X_{t-i}$

Where

$$\sum_{i=0}^{\infty} \psi_i B^i = \theta(B)\Phi^{-1}(B)\Delta^{-d}$$

Let $\hat{X}_t(h)$ denote the *h*-step forecast made at origin *t* of X_{t+h} at some future time $t+h$. Then using theorem 5.5.1 of [15] on the Eq. (1), we obtain its *h*-step forecast:

$$\hat{X}_t(h) = -\sum_{i=1}^{\infty} \pi_i^{(h)} \hat{X}_{t+h-i}$$

2.3 Generalized Autoregressive Moving Average Model Prediction

The GARMA model of a process $\{X_t\}$ is defined as

$$\phi(B)(1 - 2\eta B + B^2)^d (X_t - \mu) = \theta(B)\varepsilon_t,$$
$$|d| \le 1/2, |\eta| \le 1$$

Where $\{\varepsilon_t\} \sim iid(0, \sigma_\varepsilon^2)$, B is the lag operator, μ is the mean of the process, and $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B_i$, and $\theta(B) = 1 - \sum_{i=1}^{q} \theta_i B_i$, then the *h*-step-ahead least-squares predictor $\hat{X}_t(h)$, for $h \ge 1$, is given by the following expression:

$$\hat{X}_t(h) = \xi_h(B) X_t$$

Where:

$$\xi_h(B) = \sum_{i \ge o} \xi_i(d, v, h) B^i \qquad (3)$$

The weights $\left(\xi_j(d, v, h)\right)_{i \in Z}$ in expression () are given by:

$$\xi_j(d, v, h) = -\sum_{i=0}^{h-1} \psi_i(d, v) \pi_{h+j-i}(d, v) (j \ge 0)$$

3 A Feasibility Analysis of Fractional Predictors

In above section, we reviewed mainly traffic predictors based on fractal characteristics. Those predictors are all linear predictor based on LRD process and Gaussian with stationary increments. In the rest segment of this paper, we will discuss the influence of assuming condition, model parameter with fractional predictors and burstiness behavior on fractional predictors.

3.1 Assuming Condition

Assuming that network traffic is Gaussian is more problematic, Norrors [8] indicated when a model is build on second-order properties alone, a Gaussian process is the simplest choice. However the predictor based linear expression presence unpredictable part, where stationary is assumed, The concept of describing the overall network behavior as a series of subsection stationary intervals seems equally applicable to our traffic data, a few authors have identified various examples of non-stationary [16][17]. Indeed, LRD estimators can be fooled by non-stationary behavior such as level shifts, linear and polynomial trends [21]. These forecast models seems to fly in the face of certain obvious and characteristic features of real trace, such as the fact that it arrives in discrete bundles and that there is often a non-zero probability of zero traffic in a time interval of significant length. For above reason, the authors in [12] suggest dividing up the time-series into non-intersectant segments and separately calculating the value of the Hurst exponent for each segment. We applied the same procedure to our OC1 traces by partitioning one of the 100 minute traces to form 20 disjoint time-series. The results of our analysis are shown in Fig. 1, which shows that the Hurst exponent value varies significantly over time, oscillating between 0.5-0.85. Similar observations hold for the case of more larger scales. This indicates

that our current Internet traces are non-stationary, at least for time scales on the order of approximately one hour.
Though among the overwhelming number of studies documenting long-range dependence in Internet traffic, a phenomenon has been observed but not been cited, if there is long-range dependence, the values of the correlations between windows do not go to zero over long periods of time. For the power spectrum views, if a process has long-range dependence, the power spectrum will go to infinity as the frequency goes to zero.
Another implication is that a low order autoregressive (AR) model will not remove the low frequency bulge in the spectrum. Literature [22] use a standard spectral analysis method to compute the sample spectrum and use the AR fit on the entire distribution.
Figure 2 is the spectrum of the mean for each window of log transformed originator bytes for all transactions. The horizontal axe of both spectra is the frequency in Hertz. Figure 3 is the spectrum of residuals of the AR(5) fit. If

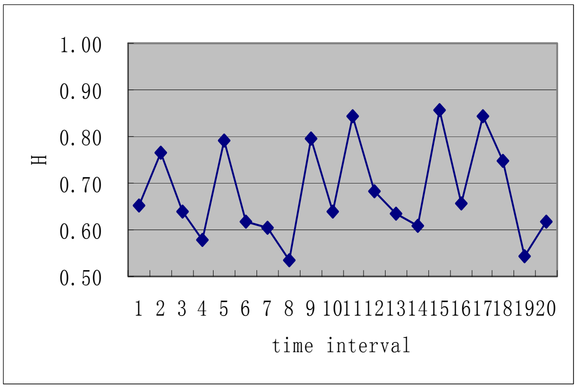

Fig.1. Hurst exponent variety with time

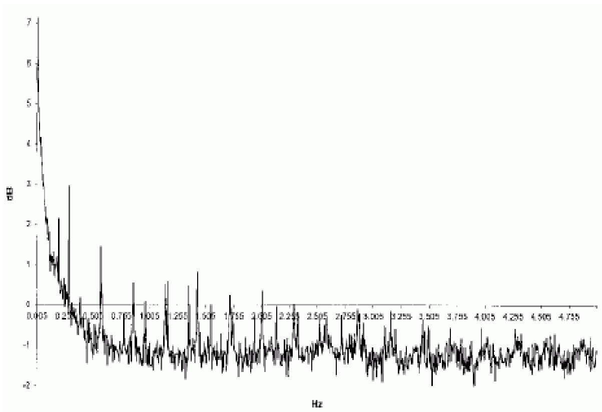

Fig.2. Spectrum of log of mean originator bytes, REJ = No, standard error = 0.21

there were long-range dependence, the spectra would have a much larger peak at the lowest frequency band. In the case, there is no evidence for long-range dependence.

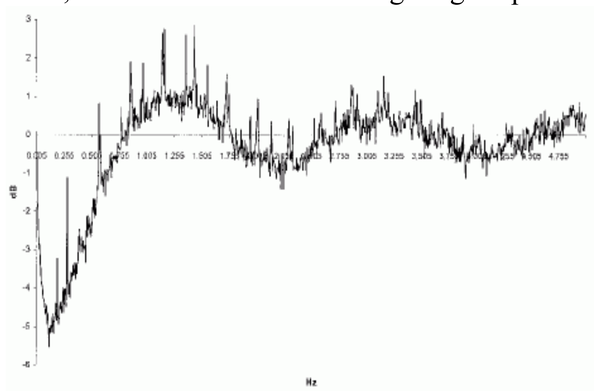

Fig.3. Spectrum of residuals of an AR(5) to log of mean originator bytes, REJ = No, standard error = 0.21

3.2 Model Parameter Estimate

Real-time traffic prediction must take place on the small time scales implied by the high bandwidth of modern telecommunication network. The ever increasing volume of data that can be collected over a given time interval brings huge storage and processing problems. The above mentions of predictor parameter show in table 1. The strength of long-range dependence measured by *d* is the same as that by H upon $d = H - 1/2$. Then we can estimate parameter *d* from this relationship. In this paper we merely discuss their common *Hurst* parameter without weight coefficient, gamma function take into account the paper's space limit.

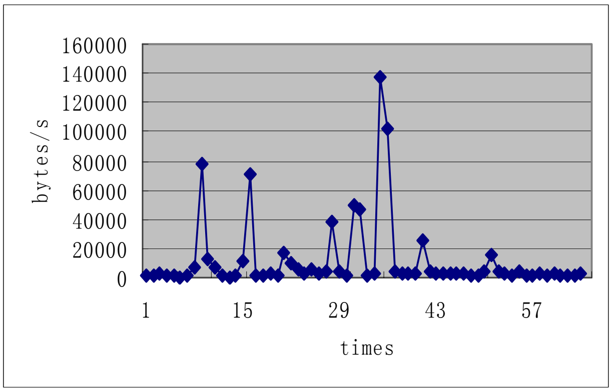

Fig. 4. A real trace of a minute

There are several ways to calculate H. Those include methods like the variance-time analysis, R-S analysis, periodogram-based analysis, Whittles's estimator and so on. Up to now, no single method is accepted as definitive the Hurst parameter H [18]. Especially, the later two methods are known to have some limitations [19], though they drive extensive adoption. We studied the

Table 1, Fractional Predictor Model Parameters

Model	Model Parameters
FGN	H, μ, σ^2
FARIMA	p, d, q
GARMA	d, v

small-time scaling behaviors of Internet traffic. Literature [20] observed the Hurst parameters at small time scales (1ms - 100ms) are fairly close to 0.5, hence for these traces the traffic fluctuations at these time scales are nearly uncorrelated. In consideration of the time expense of the other factor, so our choice minute is divided in to compute the segment. In the last few years, the Internet traffic keeps on to increase 80% annually. Because the application sends a large number of small packets with high frequency, produces high variance at certain time scale, Hurst parameter estimates for the packet-counts is easy to be hoaxed, so we adopt the H estimates for the byte counts. Figure 4 shows a real trace, comparison with various ways to calculate H in table 2. Using packet traces collected from OC1 links on an ISP pertained to China Unicom at 05/19/2003. Dismayingly, methods for the estimation of the parameter from data have obviously diversity and have suffered from poor statistical performance, or high computational complexity inappropriate for large data sets or real-time use. There exist several estimating methodologies in table 2, but they can give misleading and conflicting estimates ($H > 1$ indicated non-stationary). Hereinbefore, experimentation have indicated H fluctuated with time variety, that must be computed by interval, the problem of parameter estimate will be the obstacle of bigness that predictors are applications.

3.3 Burstiness Behavior

Most real-world traffic is self-similar and bursty (e.g. Ethernet, web, video and disk traffic). A comprehensive overview of the area can be found in [4]. Modeling of bursty time sequences has recently received considerable attention in a flood of literatures. Traditional Poisson models and Markov models have fundamental problem: they greatly underestimate burstiness. Since bursts over many or all time scales, while a Poisson or Markov process, which display burstiness over much shorter time scales. As a result, traditional models tend to yield overly optimistic performance prediction. Last five years, research works have therefore focused on LRD models capable of capturing the burstiness property of traffic processes. Whereas Internet is always changing, you do not have a lot of time to understand it. Statistics according to the LBNL, the traffic increase with 80% of every year, sustained for at least ten years. By almost any measure, data generation capabilities exceed and are growing faster than data analysis capabilities. Gigabyte-sized data sets are common, terabyte-sized data sets exist, and petabyte-sized data sets are on the way. For great variation in Internet traffic, we are desperate for parsimonious models (few parameters) handles figure 5 shown complicated circumstances is viable, isn't? So far, there is no evidence for burstiness traffic prediction.

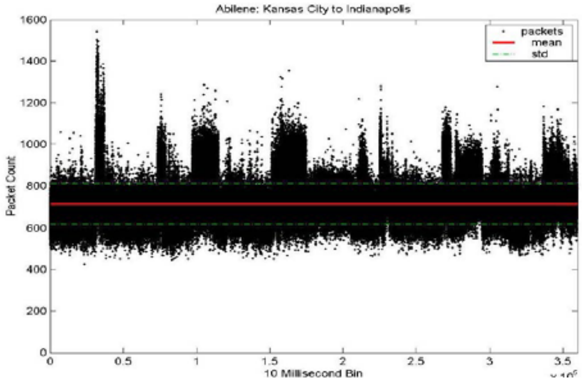

Figure. 5. A real trace with burstiness

4 Conclusion

In this paper we have explored the issue of exploiting network traffic fractional predictors. What differs here from standard network analysis is the focus on traffic forecast based on fractal characteristics. Combined assuming condition, model parameter estimate and traffic burstiness behavior to forecast model influence proceeding thorough study. The main results can be summarized as follows. Our experiment results support that the tradeoff between a large maximum prediction interval and a small prediction error reveals inconsonant. The evidence of experiment about assuming condition may mean that prediction of network traffic based fractional properties can not be accomplished with current techniques. In fact, Internet is always in the last decade at the variety that keep on, we do not have a lot of time to understand it. For instance, a single peer-to-peer file sharing application that suddenly became popular between 2002 and 2003 had a strong influence on the *Hurst* parameter for packet counts in the aggregate traffic and caused the differences in *H* between the two years. Some consistently reported Internet traffic invariants (fractal characteristics) that we are applying to forecast in the analysis of network and Internet traffic will need to be

Table 2, Comparison with various ways to calculate *H*

Methods	Aggregate Variance	R/S	Periodogram	Absolute Moments	Variance of Residuals	Abry-Veitch Estimator	Whittle Estimator
H	0.157	0.360	1.690	1.979	1.086	1.277	0.717

developed and be considered afresh. Many of the last decade's questions "What are the basic characteristics and properties of Internet traffic?" have returned.

References

[1] A. Adas, "Using adaptive linear prediction to support real-time VBR video under RCBR network service model," *IEEE/ACM Transactions on Networking, vol. 6, no. 5, pp. 635-644, Oct. 1998.*

[2] Y. Shu et al., "Prediction-Based Admission Control Using FARIMA Models," in Proc. *IEEE ICC'2000, 2000*

[3] M. Goyal, R. Guerin, and R. Rajan, "Predicting TCP throughput from non-invasive data," 2001. http://www.seas.upenn.edu/˜guerin/.

[4] Willinger, W., Taqqu M. S., Erramilli, A., ed. *A bibliographical guide to self-similar trace and performance modeling for modern high-speed networks. Stochastic Networks: Theory and Applications, ed. F.P. Kelly, Zachary, S. and Ziedins, I. (1996),* Clarendon Press, Oxford University Press: Oxford. 339-366.

[5] V. Frost and B. Melamed. Traffic modeling for telecommunications networks. *IEEE Communications Magazine, 32:70–81, March 1994.*

[6] S. Östring and H.Sirisena, "The Influence of Long-range Dependence on Traffic Prediction", *Proceedings of ICC'01*, Helsinki, June 2001.

[7] G. Gripenberg and I. Norros, "On the prediction of fractional Brownian motion," *Journal of Applied Probability, vol. 33, pp. 400–410, 1996.*

[8] I. Norros, "On the use of fractional Brownian motion in the theory of connectionless networks," *IEEE Journal on Selected Areas in Communications, vol. 13, no. 6, pp. 953–962, 1995.*

[9] J. Beran, *Statistics for Long-memory Processes*, Chapman & Hall, New York, 1994.

[10] B. B. Mandelbrot. "*A Fast Fractional Gaussian Noise Generator*", Water Resource Researches 7, 1971, pp 543-553.

[11] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Rev., vol. 10, Oct. 1968.*

[12] W.E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson, On the self-similar nature of ethernet traffic (extended version), *IEEE/ACM Transactions on Networking 2 (1) (1994) 1–15*.

[13] V. Paxson, S. Floyd, Wide area traffic: the failure of Poisson modeling, *IEEE/ACM Transactions on Networking 3 (3) (1995) 226–244.*

[14] J. Hosking." *Modeling persistence in hydrological time series using fractional differencing"*. Water Resources Research, 20:1898–1908, 1984.

[15] Joanna Duarte, Francklin Rivas Echeverrva, "Time series forecasting using ARIMA, Neural Networks and Neo Fuzzy Neurons" in Proc. *WSEAS NNA-FSFS-EC 2002.*

[16] Y. Zhang and N. Duffield and V. Paxson and S. Shenker. On The Constancy of Internet Path Properties. *In Proceedings of ACM SIGCOMM Internet Measurement Workshop, 2001.*

[17] K. Sriram and W. Whitt. Characterizing Superposition Arrival Processes in Packet Multiplexors for Voice and Data. *In IEEE J. Select Areas Commun., volume 4, pages 833–846, 1986*.

[18] Long-Range dependence:Now you see it, now you don't! T. Karagiannis, M. Faloutsos, R.H. Riedi *IEEE Global Internet, 2002*

[19] M. S. Borella, G. B. Brewster, Measurement and analysis of Long Range Dependent Behavior of Inter Packet Delay, *IEEE/Info, VOL. 2, pp. 497-504, 1998.*

[20] V. Riberio, Z.-L. Zhang, S. Moon, and C. Diot, "*Small-time scaling behaviors of internet backbone traffic: An empirical study*," Sprint ATL Technical Report, July 2002.

[21] S. Molnar and T. D. Dang. Pitfalls in Long Range Dependence Testing and Estimation. *In GLOBECOM, 2000.*

[22] Melvin J. Hinich, Robert E. Molyneux, Predicting Information Flows in Network Traffic. *JASIST 54(2): 161-168 (2003)*

[23] Francesco Palmieri, Ugo Fiore, Exploring regularities and self-similarity in Internet traffic, WSEAS MMACTEE 2001, Vouliagmeni, Athens, Greece, December 29-31, 2001