7thSASTech 2013, Iran, Bandar-Abbas. 7-8 March, 2013. Organized by Khavaran Institute of Higher Education

A comparison between self-similarity of network traffics for different protocols

M. M. Mirzaei[1], K. Mizanian[2], M. Rezaeian[3]
[1]Msc Student, Yazd University, Yazd, Iran; m.mehdi.mirzaei@stu.yazd.ac.ir
[2] Assistant Professor, Yazd University, Yazd, Iran; kiarash.mizanian@yazd.ac.ir
[3] Assistant Professor, Yazd University, Yazd, Iran; mrezaeian@yazd.ac.ir

Paper Reference Number: 07-12-6063
Name of the Presenter: M. M. Mirzaei

Abstract
Self-similarity is a phenomenon that is involved in computer network literature during last two decades and plays a significant role in modelling of network traffic. It is proved that real network traffic is self-similar and its characteristics differ from exponential traffics such as Poisson based distributions. Proposed models of network traffic are crucial to improving QoS of networks. Consequently, self-similarity must be considered in network traffic models to achieve more appropriate QoS. In this paper, we analyse two sets of real traffic which are captured from diverse regions. We examine the effect of various conditions on self-similarity of network traffic. Hurst parameter of each sets are calculated and we discuss why there are differences between Hurst parameter of them. Moreover, we categorize our traffic sets based on protocols and compute Hurst parameter of each category which itself is divided to traces of different areas. Subsequently, we compare calculated Hurst parameters within each category and discuss why there are discrepancies between them. At last, according to our observations, we explain whether infrastructure circumstances could affect self-similarity of network traffic.

Key words: Self-similarity, Network traffic modelling, Long-range dependency

1. Introduction

Understanding the nature of network traffic is vital in order to appropriately design and implement computer networks and network services. In addition, Network traffic modeling is used as the basis for the design of network applications and for capacity planning of networking systems. Therefore, it is necessary for network traffic models to be both valid and resembling the reality.

In the early 1990s, a group at Bellcore tracked an enormous series of highly detailed data to perform large time-scale analysis. The intuition for this is based on the idea that as we aggregate more and more traffic, the results will smooth to a single mean. This proves not to be the case. The results clearly showed evidence of self-similarity in real Ethernet traffics [5]. Subsequent studies by other groups [3,7,8] confirmed the discovery in a wide range of network traffics including WAN, HTTP, and so on. On the other hand, Poison-based traffic models lack this characteristic. Consequently, all previous models which are based on the poison distribution were shown to be invalid or incomplete.

Self-Similarity refers to distributions that exhibit the identical characteristics at a wide range of time scales. For example, a self-similar network trace would resemble aggregated in 10 ms scales as aggregated in 10 second scales. This is obviously not the accurate for Poisson traffics. As bin sizes increase, Poisson traffics will smooth, and eventually reach a flat line at the distribution mean. Assuredly, self-similar traffic will not; it will continue to exhibit bursts at all scales.
The degree of self-similarity of a series is denoted by the Hurst parameter. The parameter declares the speed of decay of the series' autocorrelation function. For self-similar series with long-range dependence, $1/2 < H < 1$. As $H \rightarrow 1$, the degree of both self-similarity and long-range dependency increase.
On top of everything else, we require to scrutinize the self-similarity phenomenon in order to devise an appropriate model for network traffics. To achieve this purpose and confirm previous studies in the field, in this paper, we analyze self-similarity of diverse sets of network traffics for different protocols.
In the following we review mathematical definition of self similarity. In the second section, we discuss about the traffic traces which are used in our study. Next, we mention a number of existing methods to estimate the Hurst parameter. Results and analysis of our work are discussed in the forth section. At last, we draw a conclusion from over observations and analysis.

1.1. Mathematical Definition of Self-Similarity

By definition, for a self-similar process $X(t)$, both $X(t)$ and $a^{-H}X(at)$ follow the same distribution for all positive a and t. It means that after scaling, and then normalizing, the distribution is unchanged. In other words, $X(t)$ is self-similar with H parameter of self-similarity if:

$$\forall a > 0\,, 0 < H < 1 \qquad X(t) \overset{\mathrm{d}}{=} a^{-H}X(at) \tag{1}$$

Since traffics are not exactly self-similar due to stochastically nature of sources behavior, the second order self-similar processes are more significant for modeling of network traffics. We are interested in this class of self-similar processes where the autocovariance function γ of X is invariant under translation. That is, $\gamma(t, s) = \gamma(t + k, s + k)$. Because of this property, $\gamma(t,s) = \gamma(t-s,0)$, and it is convenient to denote the autocovariance function with a single parameter, $\gamma(k)$.
A process is exactly second-order self-similar if for $1/2 < H < 1$

$$\gamma(k) = \frac{\sigma^2}{2}\left((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right) \quad for\ all\ k \geq 0 \tag{2}$$

A process is asymptotically second-order self-similar if

$$\lim_{m\rightarrow\infty} \gamma^{(m)}(k) = \frac{\sigma^2}{2}\left((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right) \quad for\ all\ k \geq 0 \tag{3}$$

2.Used Real Traffic Traces

In this study, we used two traffic data sets including the MAWI [6] traces, a trans-Pacific line (a 150Mbps link) and CAIDA [2] traces, a 10GigE link. These traces come from different geographical locations and expose variation in application mix and individual application characteristics.

We divided the MAWI traces into three parts which are captured in different hours per a day. The duration of all of these parts is an hour. We could not access such large amount of CIADA traces due to limitations which the owners exert. The duration of CAIDA traces to which we could access are three seconds (Table 1). However, it is appropriate for self-similarity analysis because it includes a multitude of packets.

Traces	Length	Capture Period	Date
MAWI1	1 Hour	2 Am – 3 Am	2012-03-30
MAWI2	1 Hour	11 Am – 12 Am	2012-03-30
MAWI3	1 Hour	8 Pm – 9 Pm	2012-03-30
CAIDA	3 Seconds	Unknown	2012-10-18

Table 1. The Traffic Traces

3. Research Methodology

There are a large number of estimates for the Hurst parameter. We mention five well-known methods below:

a. The R/S plot is the oldest and perhaps best known estimator for the Hurst parameter.
b. Aggregated variance looks at how the variance of a series changes as it is aggregated.
c. The Periodogram looks at the behavior of an estimate for the spectral density.
d. Wavelets are a method which can be considered as a generalization of Fourier transform.
e. Local Whittle estimator looks at the behavior of the frequency spectrum near the zero frequency.

The first two estimators are in the time domain and the last three in the frequency domain.

We selected The Aggregated Variance method to estimate the Hurst parameter of our traffic data sets due to simplicity. The Matlab codes which are established in [9] are used to calculate the parameter.

The aggregated variance technique is described in [1]. We briefly discuss this method at the following. It considers $var(X^{(m)})$ where $X_t^{(m)}$ is a time series derived from X_t by aggregating it over blocks of size m. It plots $log(var(X^{(m)})/\sigma^2)$ against $log(m)$ and approximate the slope of the plot which is equal to $-\beta$ and $H = 1 - \beta/2$. For example, in Fig. 1, it is depicted how the Matlab code approximate the slope. In this example estimated H is 0.71.

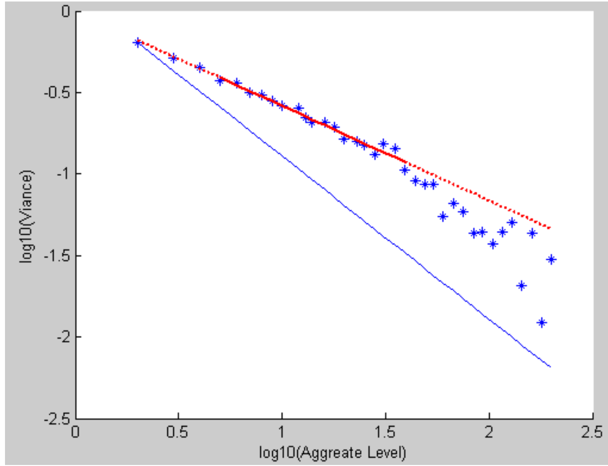

Fig 1. The Aggregated Variance Technique

4. Results and Analysis

In this study, we attempt to view self-similarity from various aspects and observe impact of different circumstances on the Hurst parameter. First of all, we examined influences of particular time on the parameter. For this purpose, we divide our traffic sets to three time periods. Second of all, we were interested to perceive how the Hurst parameter changes with protocols and whether we could relate a particular value of the Hurst parameter to a particular protocol. Lately, we investigate how the Hurst parameter alters due to geographical conditions.

To begin with, we calculated the Hurst parameter for all three parts of the MAWI traffics. As it is shown in Fig. 2, differences are not considerable enough and they are almost the same. However, it could not result in independency of the Hurst parameter and Time due to the fact that we require more observations to conclude. Moreover, the networks administrators might exert a policy which leads to this condition.

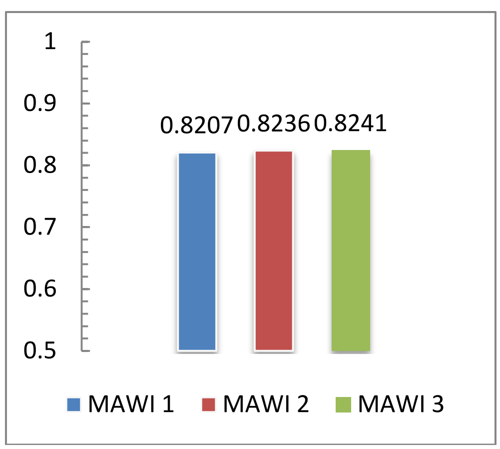

Fig 2. Estimated Hurst Parameter

It is sensible that different protocol traffics could have various degree of self-similarity. We consider, in our investigation, five well-known protocols including TCP, UDP, ICMP, HTTP, and DNS. As it is shown in Fig. 3 the protocols differ in value of the Hurst parameter. It is self-evident that TCP and HTTP have a remarkable role in the self-similarity characteristic of networks traffics. Furthermore, it is obvious that protocols which are connectionless demonstrate lower degree of self-similarity. In Fact, The impact of network dynamics on self-similar processes can cause noticeable changes in the network traffic behavior [4].

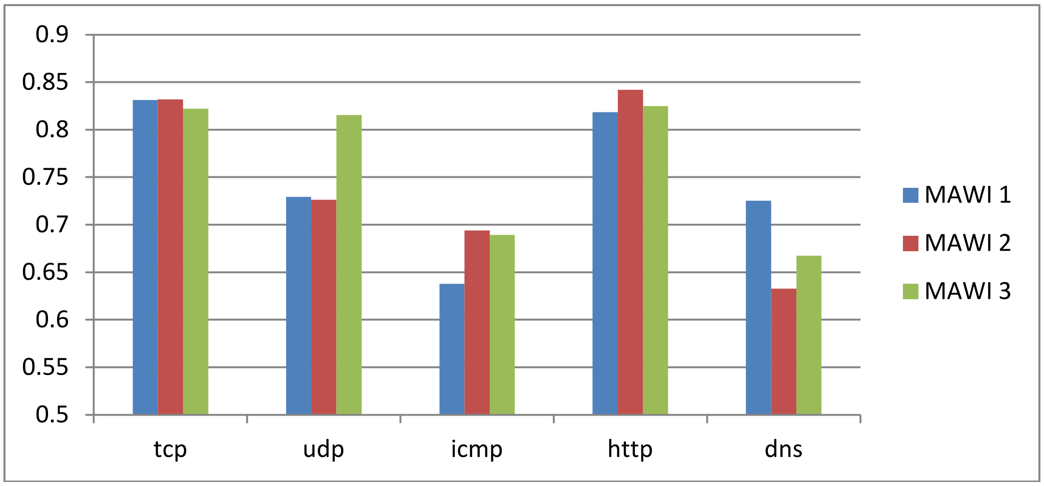

Fig 3. Estimated The Hurst Parameter for different protocol

At last, we draw a comparison between two data sets. As it is shown in Fig. 4, although they are highly different in their time periods, we could observe that they demonstrate nearly same degree of self-similarity. It might suggest that geographical conditions have no influences on the Hurst parameter of network traffics; however, it is not sufficient and more traffic sets from diverse areas must be analyzed.

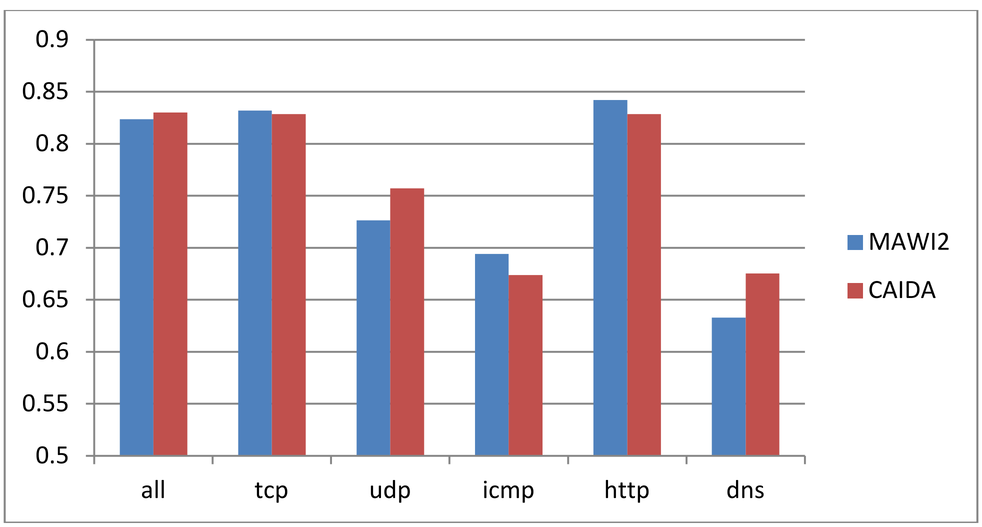

Fig 4. A Comparison Between the Hurst parameter of MAWI2 and CAIDA Datasets

5. Conclusions

In this research, we studied effects of diverse conditions on the Hurst parameter of real traffic traces. We find out that protocols' behavior could impact on the degree of self-similarity, and networks dynamics play an undeniable role in this subject.

For future works, we suggest to profoundly examine these circumstances such as geographical condition and collect large enough amount of traces in order to reach a valid conclusion.

References

[1] Beran, J. (1994). *Statistics For Long-Memory Processes.* Chapman and Hall

[2] CAIDA. *http://www.caida.org/research/traffic-analysis/*

[3] Crovella, M.E., & Bestavros, A. (1997). *Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes.* In IEEE/ACM Transactions on Networking, Vol 5, Number 6, pages 835-846

[4] Erramilli, A., & Roughan, M., & Veitch, D., & Willinger, W. (2002). *Self-Similar Traffic and Network Dynamics.* In Proc. of the IEEE., Vol 90, no. 5.

[5] Leland, W.E., & Taqqu, M.S., & Willinger, V, & Wilson, D.V. (1994). *on the self-similar nature of network traffic (extended version).* IEEE/ACM Transaction on Networking, Vol.2, pp. 1-15

[6] MAWI working group traffic archive. http://mawi.wide.ad.jp/mawi/

[7] Paxson, V., & Floyd, S. (1995). *Wide-Area Traffic: The Failure of Poisson Modeling.* IEEE/ACM Transaction on Networking, 3(2), pp. 226-244

[8] Vishwanath, K. V., & and Vahdat, A. (2009). *Realistic and responsive network traffic generation.* IEEE/ACM Transactions on Networking 17(3):712-725

[9] *http://www.mathworks.com/matlabcentral/fileexchange/19148-hurst-parameter-estimate*