Ehternet traffic self-similarity analysis using wavelet transform

Michal Kocisky, Radoslav Vargic, Ivan Kotuliak

Dept. of Telecommunications
Slovak University of Technology
812 18 Bratislava, Slovakia
Email: {kocisky, vargic, ikotul}@ktl.elf.stuba.sk

Abstract

Self-similarity is well known statistical phenomenon observed in various situations. In communication technology, the impact of self-similarity on loss was shown in the networks.
In this article, we focus on the use of Wavelet transform for characterizing measured Ethernet traffic. Our research is concentrated on self-similarity presence and we show that this phenomenon can be found in measured data.

1. Introduction

It is now widely accepted that network traffic exhibits long-range dependence (LRD)[1]. However, the practical implications of this discovery are not yet completely understood. It is known that traditional Poisson models cannot capture the behavior of LRD traffic. Often LRD data is also self-similar (SS), indicating that it possesses similar statistical properties on multiple time scales. Self-similarity and long-range dependence are closely related phenomena. The former relates to the fact that under suitable scaling, the statistics of the traffic (for instance mean, variance, or marginal distribution) are the same at any time scale. The latter relates to correlations in the data which, though decreasing over wider ranges, never become insignificant.

Discrete wavelet transform (DWT) is suitable tool, while it lets us to decompose traffic to different time scales and study it. Self similarity can be characterized by Hurst parameter (H).

2. Hurst parameter estimation

The main parameter describing self-similarity is the *Hurst* parameter H .
The closely related LRD phenomena is characterized by an autocorrelation function $\rho(k) \sim k^{-\beta} L_1(k)$, *as* $k \to \infty$, where $0 < \beta < 1$ and L_1 is slowly varying at infinity. Moreover, it is easy to see that $\Sigma_k \ \rho(k) = \infty$. This LRD non-summability means, that while high-lag correlations are all individually small, their cumulative effect is important. On the other hand the short-range dependent processes (SRD) are characterized by an exponential decay of the autocorrelations, i.e., $\rho(k) \sim \delta k$, *as* $k \to \infty$ $(0 < \delta < 1)$, resulting in a summable autocorrelation function $0 < \Sigma_k \ \rho(k) < \infty$.
LRD is also commonly defined by the slow, power-law decrease in the autocovariance function:

$r(k) \sim c_r |k|^{-(1-\alpha)}$, $k \to \infty$, $\alpha \in (0,1)$ (1)

The main parameter of LRD is the dimensionless scaling exponent α. It describes how behavior on different scales is related:

$\alpha = 0$ short-range dependent

$\alpha = (0,1)$ long-range dependent

$\alpha >$ self-similar

It is common practice to describe LRD through the *Hurst* parameter

$$H = (\alpha +1) / 2 . \quad (2)$$

The related parameter c_r is quantitative parameters which give a measure of the magnitude of LRD induced effects. The parameters may be estimated jointly using the Abry-Veitch wavelet based estimator [4], or separately by a number of other techniques

3. Network traffic consideration

Network traffic is one of the crucial questions of performance analysis. Consequently, a lot of investigation is done in this field. Our main interest is on the LRD (Long Rage Dependence) traffic, having self similar behavior [1,2].
We have analyzed measured data from the scope of presence scaling processes. For that purpose we used Abry-Veitch wavelet based estimator, which generates the Logscale Diagram [5]. Logscale Diagram (LD) is a wavelet analysis framework for scaling processes and time series which allows the estimation of key parameter, the scaling exponent α. We have investigated the time series which we get from measured data by splitting the number of transferred bytes into the 2 and 20 seconds intervals. We have analyzed the whole traffic and single protocols TCP and UDP.

4. Model understudy and result analysis

In this section, we provide firstly studied model followed by obtained results. Ethernet traffic traces deployed in our study have been captured on University network [6].

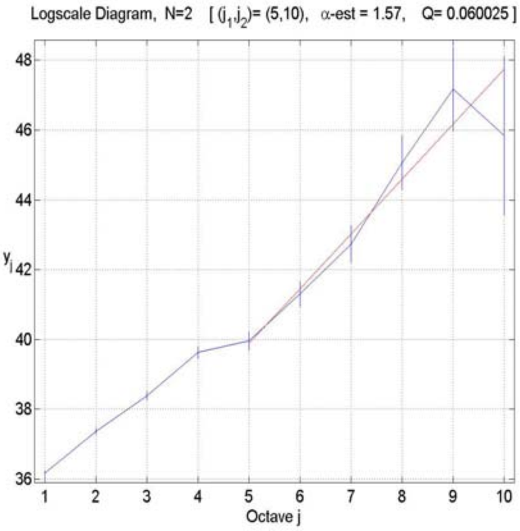

Figure 1. LD for traffic divided into 2s intervals

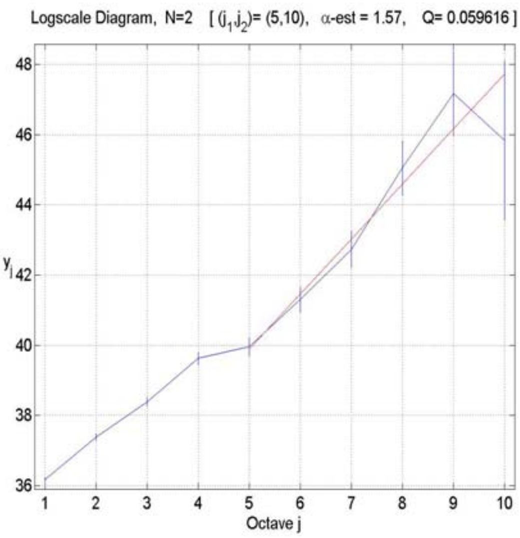

Figure 2. LD for TCP divided into 2s intervals

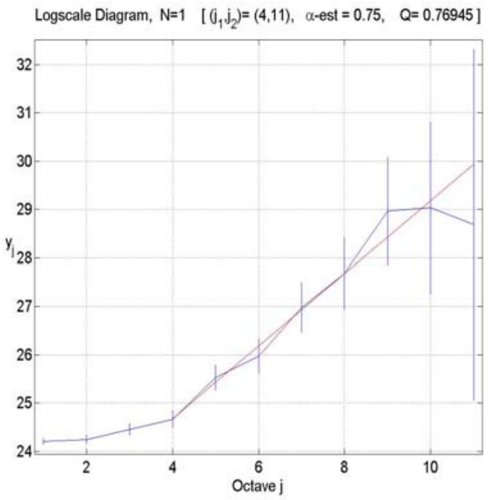

Figure 3. LD for UDP divided into 2s intervals

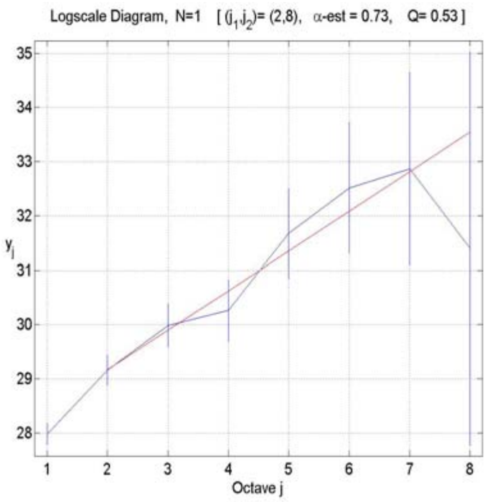

Figure 4. LD for UDP bytes divided into 20s intervals

5. Conclusion

Traffic profile can have crucial impact on the network performance. In thiis article, we have focused on the SS properties of the data traffic.

It is common to describe LRD through the *Hurst* parameter H = (α +1) / 2. *Hurst* parameter h for SS is never the same as for LRD and we obtain it from simple transformation h = (α -1) / 2.

The analysis have been made on data captured on the LAN network and backbone network. We have shown that the profile changes with traffic type, but also with aggregation.

However, we have shown that Ethernet traffic has self similar nature and it is necessary to take this fact into account.

Further work concerns more detiled data traffic analysis as well as preparation of traffic generator.

6. Appendix and acknowledgments

This research was supported by the Grant Agency of the Ministry of Education of the Slovak Republic (VEGA) under No. 1/1048/04, and by COST 279.

Table 1. Chart for data measured on backbone of LAN network

	α	H for LRD	h for SS
TCP bytes divided into 2s intervals	1.57	————-	0.285
TCP bytes divided into 20s intervals	1.8	————-	0.4
UDP bytes divided into 2s intervals	0.75	0.875	————-
UDP bytes divided into 20s intervals	0.73	0.865	————-
whole traffic bytes divided into 2s intervals	1.57	————-	0.285
whole traffic bytes divided into 20s intervals	1.8	————-	0.4

Table 2. Chart for data measured on segment of LAN network

	α	H for LRD	h for SS
TCP bytes divided into 2s intervals	2,2	————-	0,6
TCP bytes divided into 40ms intervals	2,24	————-	0,62
UDP bytes divided into 2s intervals	0,246	0,623	————-
UDP bytes divided into 40ms intervals	0,097	0,549	————-
whole traffic bytes divided into 2s intervals	2,15	————-	0,575
whole traffic bytes divided into 40ms intervals	2,13	————-	0,565

References

[1] H. E. Biaze, T. Atmaca and G. Hebuterne: Impact of shaping on network performance with self-similar traffic. In Proceedings of ICT'2000 May 2000

[2] V. Paxton and S. Floyd: Wide-area traffic: The failure of poisson modelling. ACM SIGCOM pp 257--268. Aug. 1994

[3] Vetterli,M., Kovacevic,J.: Wavelets and Subband Coding, Prentice Hall, 1995.

[4] Patrice Abry, Darryl N. Veitch – “A Wavelet Based Joint Estimator of the Parameters of Long-Range Dependences

[5] Cox, D.R., “Long-Range Depedence: A Review”, in Statistics: An Appraisal, H.A.

[6] Benkovic R., Lasz, J., Kotuliak.: “Measurement of traffic profile in Ethernet networks”, RTT 2004,