Application of the Analysis of Self-similar Teletraffic with Long-range Dependence (LRD) at the Network Layer Level

G. Millán, G. Lefranc

Ginno Millán
Universidad Católica del Norte
Escuela de Ingeniería, Larrondo #1281, Coquimbo - Chile
gmillan@ucn.cl

Gastón Lefranc*
Pontificia Universidad Católica de Valparaíso
Escuela de Ingeniería Eléctrica, Avda. Brasil #2147, Valparaíso - Chile
*Corresponding author: glefranc@ucv.cl

Abstract: In a previous paper it was proposed, and theoretically confirmed, that analysis of self-similar traffic flows with long-range dependence may be restricted to the network layer. In this paper this novel concept is applied to the study of traffic recorded in an IEEE 802.3u network environment with the aim of proving its validity as a simple and efficient tool for high speed computer network traffic flow analysis.
Keywords: Long-range-dependence, network layer, traffic models, self-similar process.

1 Introduction

It is interesting to reflect on the idea that a purely random process is no more than a theoretical concept, but it is much more interesting to do so considering that no series is yet known whose characteristics correspond exactly to those of such processes. Likewise, it is of interest to explain that a given behavioral evolutionary singularity is widely attributable to two stationary stochastic processes without considering their origins, scope and implications.

To clarify the above two assertions, [1] made an exhaustive background review that includes research proposals and their results as well as the mathematical foundations underlying all of them. However, a basic idea remains unfinished: if all the arguments given do nothing more than highlight the benefits and advantages of the parsimonious modeling of traffic flow in current high speed network environments, then why is there dissent on its use?, and even more important yet, why do all the results deal with self-similarity as ubiquitous not only on the time scales, a fact that is certainly not put in doubt, but also with respect to the set of circumstances attributable to its origin?

With the purpose of answering these questions, that same paper states and then gives the foundation for the validity of the following working hypothesis: *"It is completely feasible to restrict the evolution of a statistically self-similar process to a well defined application setting without altering its nature and its more important properties, in that way highlighting the validity of its postulates and giving greater plausibility to its physical interpretation"*, clarifying that the plausibility refers to the action of conferring an admissible, and therefore worth considering, character to one or various parameters that compose an analytical model whose interpretations are not only mathematical idealizations, and that the theoretical proof of that hypothesis is based essentially on the proposal of Ryu and Lowen [2], which consists in making a distinction between the self-similarity observed at the application level and the self-similarity observed at the network level, but with the substantial difference of not considering any particular traffic model as in the case of the authors, in which the proposal is developed to support the analysis of the results obtained from using the fractal point process (FPP) model proposed in [3], to carry

out the characterization of traffic flow in high speed networks. It should be specified that the original proposal of Ryu and Lowen consists in establishing a difference based on the OSI level at which is found the source that gives origin to the traffic flows in which self-similarity is seen, and therefore explains its origin. In this way the authors conclude that it is more adequate to refer to application level fractal traffic and network level fractal traffic instead of encompassing under fractal traffic a whole range of dissimilar behaviors that find explanations precisely in the internal processes inherent to each of those levels. It is therefore clear that this division of fractal traffic into those two subcategories addresses effectively the profound differences that exist in the design as well as in the control processes of actual high speed networks at those levels.

Concretely, the application level self-similar traffic (fractal) has its origin in a source that exhibits self-similarity over a wide range of time and frequency scales without any interactions with the network. In other words, self-similarity is inherent to the source, while network level self-similar traffic (fractal), in contrast, exhibits self-similarity over a wide range of time and frequency scales as a result of numerous interactions with the network. An example of application level self-similar traffic source is the VBR video sequence of [4], while all the applications based on TCP of [5]-[8] are examples of network level self-similar traffic sources. It should be stated that actually the behavior of an application level traffic source can be affected by the network conditions, depending on the functionalities of the low level protocols used. This effect is insignificant, however, compared to what happens with a network level traffic source, such as an FTP client, under identical conditions, where the ratio of output to input data flow depends directly and critically on the conditions of the network and can largely be considered independent of the size of the files [9]. Also, application level self-similar traffic can be managed in the context of the resource assignment admission control subjected to service quality guarantees, since it is independent of the network conditions through which it is sent.

This paper presents an experimental application that validates the hypothesis of [1]. Considering a real network scenario implemented under IEEE standard 802.3u, traffic capture experiments are performed and then analyzed, restricting the results according to the precepts presented above as well as in [1]. This paper is therefore centered on showing the validity of the analysis of restricted self-similar traffic at the network layer level as a simple and efficient tool to study the behavior of traffic flow in present day high speed computer networks.

2   Traffic Measurements

2.1   Description of the network environment

Figure 1 is a diagram of the topology of the implemented experimental network scenario. It is a LAN IEEE 802.3u environment that has the following main operational characteristics:

- Ten workstations that uninterruptedly request an on-demand video service from the video server equipment provided for it. Both the client equipment as well as the server make use of the VLC Media Player application for that purpose. Continuous reproduction is achieved by predefining a video program in the server equipment. The ten stations keep an XML file with the page index of the web server. The state of the page index updating is consulted randomly by each of the stations with the purpose of always having the latest version of the file.

  The network monitoring equipment (tagged "Sniffer" in Figure 1) makes use of the Ethereal application to perform the traffic packet capture.

- The Internet access functions for the purpose of updating both the operating systems and the antivirus applications are enabled and automated in all the network's equipment.

2.2 Work Methodology

The procedure to carry out the experiment consists of the following steps:

- Program the algorithms to estimate the values and study the behavior of the Hurst parameter, of the variance-time (V-T), rescaled adjusted range or the R/S statistic, and spectral density analyses. All the programs were developed over MATLAB because of their availability. Verify the correct operation of the programmed algorithms. To perform this operation, use is made of the traffic sample normalized series of [5], BC-pAug89.TL, available for downloading in [10], and the values obtained are then compared with those reported in the literature by the authors.

  Capture traffic from the experimental network shown in Figure 1. It should be pointed out that the duration of each of the traffic capture periods is governed only by a criterion of availability of storage capacity in the Sniffer equipment, trying to capture the largest possible number of samples to face a possible decision scenario based on a figure of merit coming from the bias versus variance relation.

- Using the capabilities of the Ethereal application the filtering of the captured packets is carried out in such a way as to create time series of data that contain the length of the packets and the arrival times of each of them. These series are then stored in flat text files.

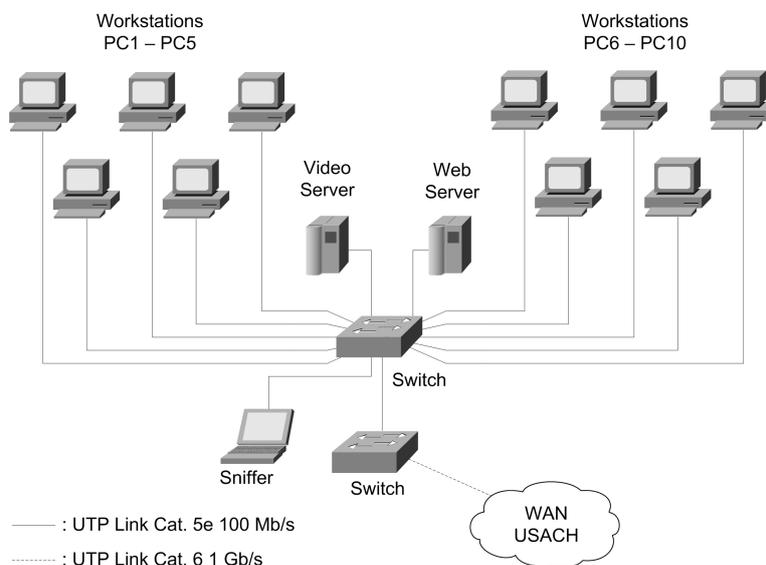- Apply the V-T, R/S, and periodgram analyses over the data series previously specified.

Figure 1: Experimental network connections

2.3 Simulations

Table 1 shows the detail of the captured Ethernet frames. In that respect the following aspects must be considered:

- The temporal resolution of the arrival times that are recorded by the Sniffer equipment is set in microseconds. This resolution is delivered as default measurement by the Ethereal application. The existence of time fluctuations that are not considered as those due to the latency of the circuits of the equipment's network card and those due to the code

processing time by the equipment is suggested. However, and in spite of the great impact
that both can get to have on the finally recorded times, the option is taken to consider
them anomalies belonging to the data processing systems, and they are therefore part of
their behaviors.

The individual captures present a time resolution of 6 μs, which is the average recorded
value.

- The data sets Trace-1 to Trace-4 contain the data of the time series representative of each
  capture process. In this respect, each series is composed of an ordered list of pairs of data:
  the arrival time of the packet, recorded according to the above considerations in floating
  point format with six positions, and the size of the captured Ethernet packet, which records
  the length of Ethernet data. The recorded value does not include the following fields:
  preamble, address of origin, destination address, length, and CRC or verification sequence.
  It must be recalled that the Ethernet protocol forces the frames to have a size with a
  minimum of 64 bytes and a maximum of 1518 bytes, so the recorded values are within that
  interval, with the 1518 bytes value as that recorded mostly in all the capture processes.

- 99.9% of the Ethernet PDUs are encapsulated in IP datagrams.

- With the purpose of testing the algorithms programmed for each method, whose math-
  ematical expressions are given by (1)-(3), the analysis sequence begins using the series
  BC-pAug89.TL, from Figure 2, with the V-T, R/S and periodgram analyses, from left to
  right, respectively. Then, and in the same order, the analyses for each of the four data sets
  are shown.

Table 1: Qualitative description of the traffic capture sets

Measurement Period		Data Set	Number of Packets
November 2010	Total: 32 h		2118505
Start of Trace:	First Period:	Trace-1	918896
Nov. 29, 06:00 am	(06:00 am - 12:00 am)		
End of Trace:	Second Period:	Trace-2	1199609
Nov. 30, 08:00 pm	(08:00 am - 08:00 pm)		
December 2010	Total: 38 h		6096937
Start of Trace:	First Period:	Trace-3	1338789
Dec. 1, 08:00 am	(08:00 am - 08:00 pm)		
End of Trace:	Second Period:	Trace-4	4758148
Dec. 2, 10:00 pm	(12:00 am - 10:00 pm)		

$$\mathrm{Var}[X^{(m)}] \sim m^{-\beta}, \ 0 < \beta < 1, \ H = 1 - \beta/2 \tag{1}$$

$$\frac{R(n)}{S(n)} = \frac{1}{S(n)}[\max(0, W_1, ..., W_n) - \min(0, W_1, ..., W_n)], \quad W_k = \sum_{i=1}^{k} X_i - k\bar{X}(n) \tag{2}$$

$$f(\lambda) \sim \lambda^{1-2H}, \quad \text{when } \lambda \to \infty \tag{3}$$

3 Discussion of Results

Table 2 summarizes the results obtained. It shows that the value of H for the BC-pAug89.TL
data series using variance-time and R/S analyses is correct with respect to the value determined

by the authors, $H = 0.9$ using the R/S graphic method. It is also verified that all the values of H found for each of the experimental series (Trace-1, Trace-2, Trace-3, and Trace-4) are within the interval of interest $1/2 < H < 1$, which certainly implies an asymptotic behavior of the self-correlation function given by $r(k) \sim H(2H-1)k^{2H-2}$, when $k \to \infty$ [11], which ensures that these series present a hyperbolic type drop in the tails of their distributions, thereby reinforcing a behavior different from a typically exponential one.

Clearly, the central limit theorem reinforces the previous condition by considering that the self-covariance function of these processes in the interval $1/2 < H < 1$, depends on the value of the H in the approximate form $\gamma(k) \sim Ck^{2H-2}$, when $k \to \infty$, with C < 0 [12].

The presence of long-range time dependence is seen from $r(k) = r^m(k)$, $\forall k \geq 1$, $\forall m \geq 1$ [13], a fact that is evidenced from the linear type behavior of the relation subjacent between different levels of aggregation with respect to the variance, which becomes evident from the V-T graph analysis of the Figures 3, 4, 5, and 6.
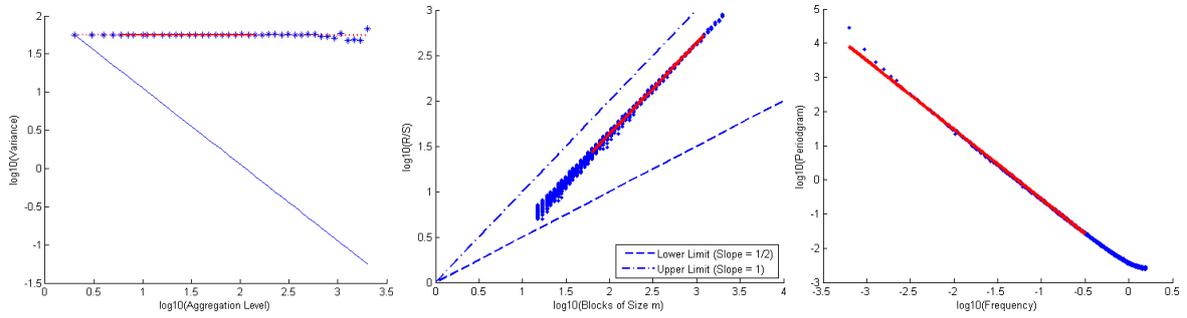

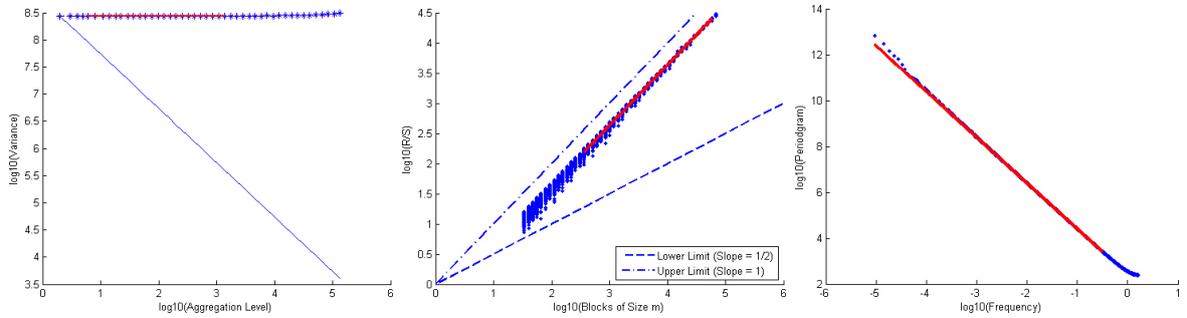Figure 2: V-T, R/S, and periodgram for BC-pAug89.TL trace. $H = 0.9000$, $H = 0.8991$, and $H = 0.7681$

Figure 3: V-T, R/S, and periodgram for Trace-1. $H = 0.8999$, $H = 0.8986$, and $H = 0.7680$
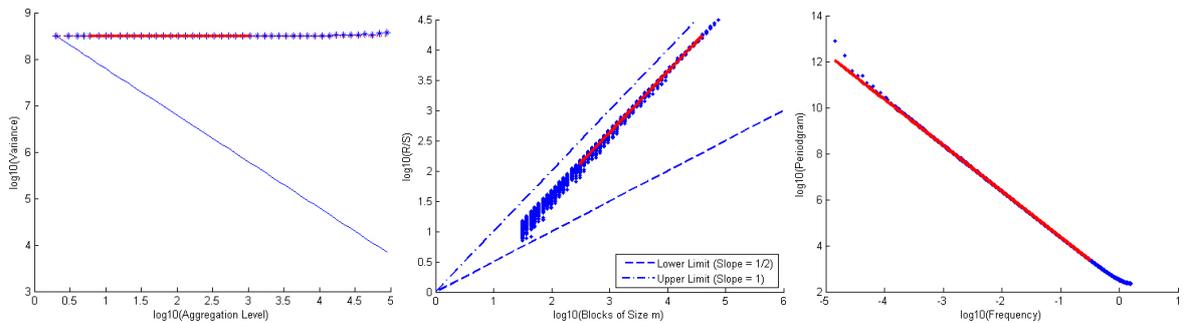

Figure 4: V-T, R/S, and periodgram for Trace-2. $H = 0.8999$, $H = 0.8982$, and $H = 0.7678$
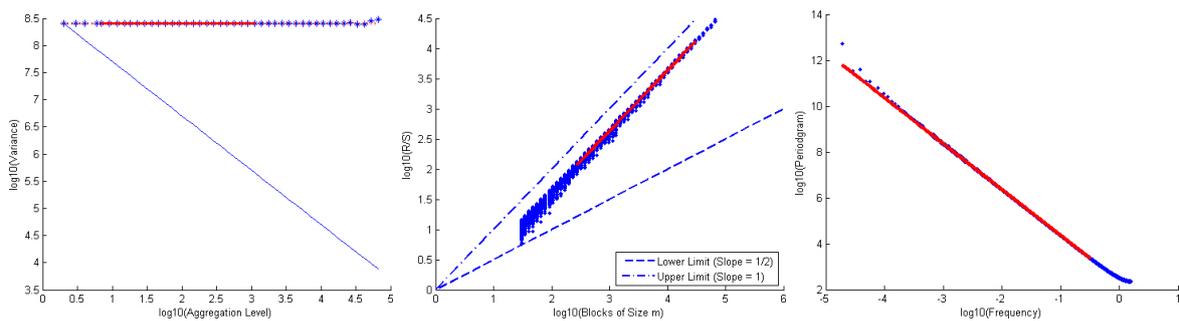
Figure 5: V-T, R/S, and periodgram for Trace-3. $H = 0.8999$, $H = 0.8985$, and $H = 0.7677$
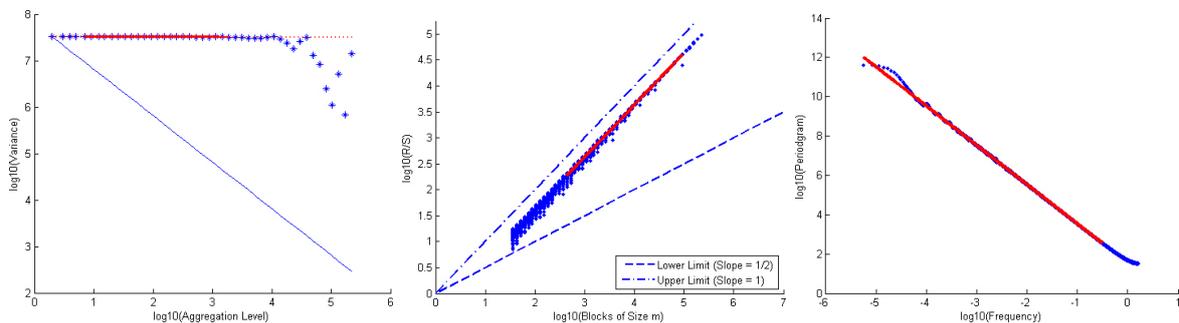

Figure 6: V-T, R/S, and periodgram for Trace-4. $H = 0.9000$, $H = 0.8994$, and $H = 0.7681$

Table 2: Results obtained for H after applying them to different methods of analysis

Method	Value of H for the Data series Measurement Period				
	BC-pAug89.TL	Trace-1	Trace-2	Trace-3	Trace-4
V-T	0.9009	0.8999	0.8999	0.8999	0.9007
R/S	0.8969	0.8996	0.8982	0.8985	0.9004
Periodgram	0.6603	0.6680	0.6678	0.6677	0.6681

The self-similar behavior of the different levels of aggregation, as well as its pronounced long-range dependence, are made evident from the linearity shown by the graphs of the size m blocks with respect to the R/S factor. See R/S graph analysis of the Figures 3, 4, 5, and 6.

The self-similar behavior of the different data series is also shown by the last row of Table 2, which considers the spectral estimation based on the periodgram of the series. The noticeable differences between the values of H for each of the series lie in the absence of confidence intervals. The analyses based on the R/S and variance-time graphs show simplified limiting cases that do not consider fine interactions between aggregate components when their differences are small, i.e., when considering, for example, two levels of immediately contiguous aggregation. In this respect, the frequency analysis does consider them, because the spectral handling is, most of the time, sufficiently exact to not omit the bias generated between the compromise of the representation of the singularities of each aggregation level and their general values in terms of the variance. Therefore, that is the reason to consider the aggregation of frequencies in relation to the periodgram, instead of the block sizes with respect to the variance. However, for the correct use of this method each result must go together with confidence intervals, which in general it is recommended to obtain from Whittle's Maximum Likelihood Estimator (MLE), but this goes beyond the objectives stated for this research, beside the fact that interest is focused on getting

a general view of the presence or absence of self-similarity characteristics, and not of their exact value, so the values obtained through the variance-time and R/S analyses as upper limits can therefore be considered with minimum error in the final interpretation of the results.

It is important to note that, even though no traffic model has been developed and it has only been assumed that the different arrivals are related to an On/Off model, as evidenced by the use of the variance-time and R/S statistics in terms of representing aggregations, parsimony is a provable characteristic based on the number of parameters needed to establish self-similarity and long-range dependence.

Finally, the proposal of restricting the results derived from representing the traffic flows as second order self-similar time series at the network level when necessary, certainly results not only feasible based on the set of arguments presented, but it is also useful for the correct understanding of and search for the origin of the self-similarity (fractality) and the long-range dependence that it may eventually exhibit. In that respect, it is undeniable that it is more adequate to interpret the presence of these singularities as the product of the internal processes subjacent in each OSI level in particular, instead of referring to them as intrinsic characteristics of the data flows because of the indetermination to which it leads. Therefore, the possibility of restricting a statistically self-similar process to a well defined application environment without altering its nature and its most important properties is true, thereby highlighting the validity of its postulates and adding greater plausibility to its physical interpretation, which is derived precisely from not forcing a reality to fit all the parameters of a given model, but on the contrary, it is taking care directly of a characteristic of the behavior of traffic flows that is exposed through its observation and mathematical interpretation.

4   Conclusions

A new point of view has been presented with respect to the systematic classification of self-similar processes with long-range dependence, whose purpose, applied to the parsimonious modeling of high speed computer networks, is to propose a common working framework for the interpretation of the results derived from the representation of traffic flows in the form of self-similar second order stationary time series.

Supported by the use of a set of four high time resolution traffic samples, it is shown that the behavior of traffic flow in a high speed computer network is self-similar and exhibits long-range dependencies over a wide range of time scales, which is evidenced when the dependence that exists among the different aggregation levels with respect to their second order statistics analyzed, and it is confirmed that regardless of the level considered, the trend reflects linearity. This linearity implies a behavior that does not depend on the time scales under consideration, but a behavior inherent to the data flows.

The mathematical models of current high speed computer networks must consider self-similarity as well as the long-range dependencies. Although it is true that the literature often proposes the equivalence of both concepts starting from an indistinct treatment, it happens that these concepts are actually independent, without one necessarily being the consequence of the other, and more profoundly, one not implying the existence of the other; these last two considerations are regardless of the direction in which they are taken.

To consider a restriction in the interpretation of the results derived from representing the traffic flows as self-similar stationary time series in relation to the OSI observation level results not only feasible, but it must be considered as a useful practice aimed at accuracy in determining the origins of self-similarity and long-term dependence, and for a better understanding of their implications, concretely when dealing with high speed computer networks where the protocol dependencies of the data flows require an adequate interpretation for their practical use in network

engineering. In this respect, it is doubtlessly a better practice to interpret these singularities as a product of the internal processes that underlie each OSI level in particular, than to interpret as the indetermination caused by referring to them as intrinsic characteristics of flows without an origin.

Bibliography

[1] Millán, G.; Kaschel, H.; Lefranc, G. (2010); Discussion of the Analysis of Self-similar Teletraffic with Long-range Dependence (LRD) at the Network Layer Level, *International Journal of Computers Communications & Control* , ISSN 1841-9836, 5(5): 799-812.

[2] Ryu, B.K.; Lowen, S.B. (1997); Point Process Approaches for Modeling and Analysis for Self-Similar Traffic. Part II: Applications, *Proc. 5th International Conference on Telecommunications Systems, Modeling and Analysis*, Nashville, TN.

[3] Ryu, B.K.; Lowen, S.B. (1996); Point Process Approaches for Modeling and Analysis for Self-Similar Traffic. Part I: Model Construction, *Proc. IEEE INFOCOM '96*, San Francisco.

[4] Garrett, M.W; Willinger, W. (1994); Analysis, Modeling and Generation of Self-Similar VBR Video Traffic, *Computer Communication Review*, 24(4):269-280.

[5] Leland, W.E.; Taqqu, M.S.; Willinger, W.; Wilson, D.V. (1994); On the Self-Similar Nature of Ethernet Traffic (Extended Version), *IEEE/ACM Trans. Netw.*, 2(1): 1-15, Feb. 1994.

[6] Paxson, V.; Floyd, S. (1995); Wide-Area Traffic: The Failure of Poisson Modeling, *IEEE/ACM Trans. Netw.*, 3(1):226-244.

[7] Ryu, B.K. (1996); Implications of Self-Similarity for Providing End-to-End QoS Euarantees in High-Speed Networks: A Framework of Application Level Traffic Modeling, *Lectures Notes in Computer Science (Proceedings of International Zurich Seminar on Digital Communications (IZS '96))*, Plattner, B. (ed.), Zurich, Switzerland: Springer-Verlag, (1044):65-79.

[8] Willinger, W.; Taqqu, M.S.; Sherman, R.; Wilson, D.V. (2007); Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level, *IEEE/ACM Trans. Netw.*, 5(1):71-86.

[9] Aracil, J.; Edell, R.; Varaiya, P. (1997); A Phenomenological Approach to Internet Traffic Self-Similarity, *35th Annual Allerton Conference on Communication, Control and Computing*, Urbana-Champaign, IL.

[10] http://ita.ee.lbl.gov/html/contrib/BC.html.

[11] Park, K.; Willinger, W. (2000), Self-Similar Network Traffic: An Overview, *Self-Similar Network Traffic and Performance Evaluation*, Park, K.; Willinger, W. (eds.), New York: John Wiley & Sons.

[12] Sheluhin, O.I.; Smolski, S.M.; Osin, A.V. (2007); *Self-Similar Process in Telecommunications*, Chichester, UK: Wiley.

[13] Stallings, W. (2004); *Redes e Internet de Alta Velocidad. Rendimiento y Calidad de Servicio*, 2nd ed., Madrid, España: Pearson - Prentice Hall.