CHARACTERISATION OF REAL GPRS TRAFFIC WITH ANALYTICAL TOOLS

Zhan-hong Lu[1], Yang Yang[2], Yong-hua Song[1] and Thomas Owens[1]

[1] School of Engineering and Design,
Brunel University, Uxbridge, UB8 3PH
London, United Kingdom

[2] Dept of Electronic & Electrical Engineering
University College London
Torrington Place, London, WC1E 7JE, UK

Zhanhong Lu's email: zhan.lu@brunel.ac.uk
Yong-hua Song's email: y.h.song@brunel.ac.uk

Yang Yang's email: y.yang@ee.ucl.ac.uk
Thomas Owens' email: Thomas.Owens@brunel.ac.uk

Keywords: GPRS, Traffic characterisation, Self-similarity, Long-range dependence, Heavy-tailed distribution

Abstract

With GPRS and UMTS networks launched, wireless multimedia services are commercially becoming the most attractive applications next to voice. Because of the nature of bursty, packet-switched schemes and multiple data rates, the traditional Erlang approach and Poisson models for characterising voice-centric services traffic are not suitable for studying wireless multimedia services traffic. Therefore, research on the characterisation of wireless multimedia services traffic is very challenging. The typical reference for the study of wireless multimedia services traffic is wired Internet services traffic. However, because of the differences in network protocol, bandwidth, and QoS requirements between wired and wireless services, their traffic characterisations may not be similar. Wired network Internet traffic shows self-similarity, long-range dependence and its file sizes exhibit heavy-tailedness. This paper reports the use of existing tools to analyse real GPRS traffic data to establish whether wireless multimedia services traffic have similar properties as wired Internet services traffic.

1 Introduction

Recent advances in wireless communications have shown a clear trend of traffic migration from voice-centric communications to data-centric communications. Web browsing based Internet services are becoming more and more popular in mobile wireless communication networks such as General Packet Radio System (GPRS) and third generation (3G) mobile communications systems. The 3G mobile systems can support high-data rate multimedia services with different Quality-of-Service (QoS) requirements. Most of these services are data-centric and bursty in nature and therefore fundamentally different from the voice-centric services of traditional telecommunication networks. Consequently, the classic Erlang approach and Poisson models for characterising voice-centric services traffic are not suitable for characterising data-centric wireless multimedia services traffic [13].

Traffic characterisation of wireless multimedia services to a large extent relies on knowledge and experience gained in the areas of wired network Internet services. Self-similarity (SS) and heavy-tailed distribution (HTD) characterise World Wide Web (WWW) traffic [5]. There is also evidence of self-similarity in Ethernet traffic [11] and in Wide Area Networks (WANs) traffic originally reported in [13] and later studied and confirmed in [19]. Furthermore, Garrett and Willinger [7] found that Variable Bit Rate (VBR) video traffic shows the property of self-similarity. However, little work has been done on wireless network traffic characterisation due to a lack of raw data. Nevertheless some research shows wireless traffic can have properties similar to that of wired network Internet traffic. Kalden and Ibrahim [10] found evidence of self-similarity in GPRS network traffic but their analysis related only to typical WWW and Wireless Access Protocol (WAP) applications traffic and their conclusions cannot easily be extended to other wireless multimedia applications traffic. This paper investigates the properties of wireless multimedia services traffic in terms of IP traffic characterisation for traffic types covering a variety of mixed applications traffic such as WWW, WAP, Email, File Transfer Protocol (FTP), Multimedia Messaging Service (MMS), and Short Message Service (SMS) traffic. The properties studied are the heavy-tailed distribution (HTD), self-similarity (SS) and long-range dependence (LRD).

The paper is organised as follows: In section 2, the concepts of HTD, SS, and LRD and the tools for their analysis are briefly described. In section 3, the results of the HTD, SS, and LRD analysis of real GPRS traffic data are presented. The conclusions together with proposals for future work are given in section 4.

2 Basic concepts and analytical tools

The concepts of HTD, SS, and LRD are described in [1, 10 and 12] where the tools for their analysis are also discussed.

2.1 Self-similarity and analytical tools

In a strict sense, self-similarity means that the statistical properties of a stochastic process do not change for all aggregation levels of the stochastic process [10]. The degree of self-similarity is estimated by the Hurst Parameter (H) [9]. If H has a large value, it means that the traffic has a strong self-similarity level.

Several methods are available for estimating the Hurst parameter [1, 3, 16, and 17]. In the Local Analysis of Self-Similarity (LASS) tool [15], the Wavelet Transform approach, i.e., the Abry-Veitch method is implemented.

2.2 Long-range dependence and analytical tools

Long-range dependence reflects the persistence phenomenon in a self-similar process and is mainly defined in terms of the behaviour of its autocorrelation function. It has been recognised in many fields including hydrology, biology and semiconductor physics but is still new to telecommunications. The typical way to analyse LRD is to use wavelet transform theory, especially the multiresolution and Mallet algorithm [2]. Logscale Diagram estimator (LDestimate) code [14] is a tool designed to analyse the LRD and SS properties of scaling processes or time series which, in particular, allows the estimation of their key parameters (Hurst parameter, Alpha parameter) and the scaling exponent.

2.3 Heavy-tailed distribution and analytical tools

The heavy-tailed distribution represents power-law behaviour in the tail of the distribution of a random process. The tail index of a heavy-tail distribution can be estimated and it gives the degree of the heavy-tailedness. The Hill estimator [8] gives a method for estimating the tail index. However, it is difficult to identify the point where the power-law behaviour begins. The tool named "A" estimation (Aest) is designed to avoid this issue by using the scaling property principle to estimate the tail index of a heavy-tailed distribution [4].

3 Real GPRS traffic data and its analysis

3.1 Real GPRS traffic data

The real traffic data used for this study is taken from Vodafone's GPRS network in the Netherlands. It is a four-day (19-22 Oct 2004) measurement of IP packet size collected in each millionth of a second (μs). The total dataset comprises 82007903 samples where each sample is one packet size in bytes. For IP traffic, the burst is the aggregation of IP packet in 100ms time intervals.

The traffic data was collected at a Gn-interface between a Serving GRPS Support Node (SGSN) and a GPRS Gateway Support Node (GGSN) in the GPRS network. The measurement was of the network layer traffic that contains mixed applications traffic including WWW, WAP, FTP, Email, SMS and MMS because it is not necessary to distinguish the applications for the analysis of network layer properties.

3.2 IP packet size - SS and LRD analysis results

LASS and LDestimate are tools that estimate the Hurst parameter of a random process. In this section, the analysis of the real GPRS traffic described in section 3.1 using these tools is reported and the results are presented.

- Results obtained from the LASS tool

With respect to the analysis of the results for IP packet size, five steps analysis is output by LASS. To compare the results of LASS and LDestimate, only the LASS output at step 3 is shown in Figure 1 because it presents the estimate of Hurst parameter, H, clearly and accurately.

Figure 1 (a) (top of the figure) shows the estimate of the Hurst parameter against the window number in the case of interval Octave j is $[j_1, j_2]$ = [1, 18] and [17, 18]. Figure 1 (b) (left-bottom of the figure) is the plot of the mean of the data against the window number. Figure 1 (c) (right-bottom of the figure) is the plot of the standard deviation of the input data against the window number. In order to accurately estimate the Hurst parameter, the Octave j interval is chosen as $[j_1, j_2]$ = [17, 18], because in case of $[j_1, j_2]$ = [j, j+1], the mean of Hurst parameter H can be accurately estimated using the equation $H \approx \hat{H}_{[j,j+1]}$ [15]. Therefore, the upper curve in Figure 1 (a) shows the estimate of the Hurst parameter in different windows. The mean of H is 0.8424.

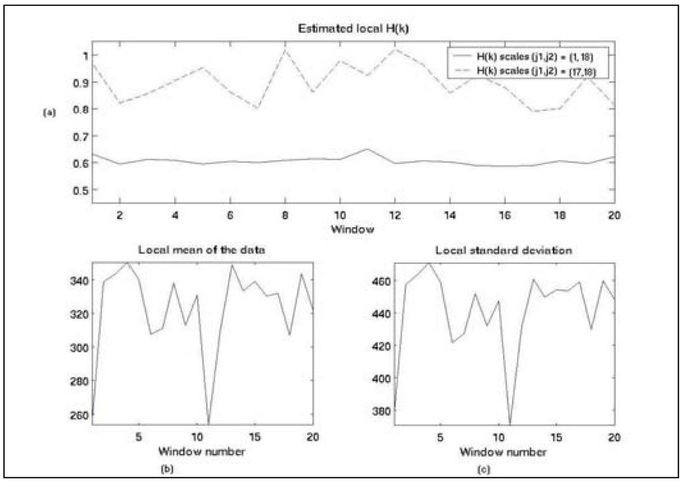

Figure 1: IP packet size - step 3 output of LASS tool

According to the definition of SS and the estimate of H, the conclusion can be drawn that the IP packet size distribution exhibits the property of SS. Since the estimate of H (0.8424) falls in the range of [0.5, 1], it also suggests the LRD property [6].

- Results obtained from the LDestimate tool

The dataset described in section 3.1 was input into the LDestimate tool. According to the principle of the LDestimate tool [14], the horizontal-axis of Figure 2 represents Octave j and the vertical-axis is the y_j which can be obtained from the wavelet transform coefficients $d_{j,k}$ using the equations below

$$y_j = \log_2(\mu_j) \tag{3.2.1}$$

$$\mu_j = \frac{1}{n_j}\sum_{k=1}^{n_j}\left|d_{j,k}\right|^2 \tag{3.2.2}$$

where n_j is the number of coefficients available at Octave j.

The slope of the regression line plotted in Figure 2 is the estimate of the Hurst parameter.

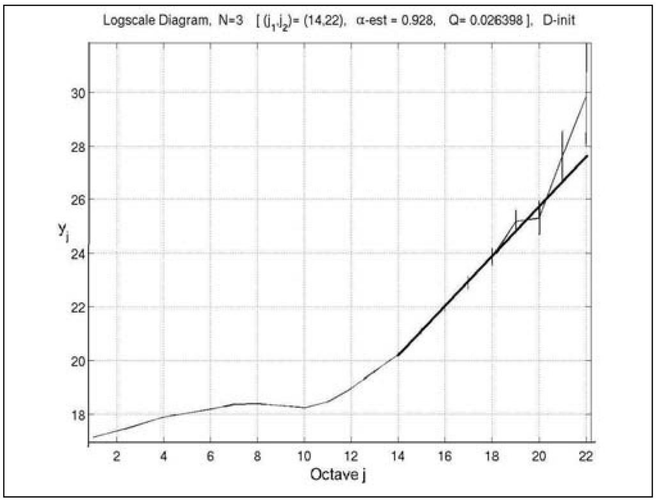

Figure 2: IP packet size - logscale diagram plot of the output of the LDestimate Tool

The plot in Figure 2 illustrates that the available Octave j is 22 and the first step regression can be obtained from the Octave j interval of $[j_1, j_2] = [1, 22]$, but the goodness of fit is zero. Based on the principle of LDestimate, a new suitable Octave j interval $[j_1, j_2]$ should be fixed to reach the maximum goodness of fit. The algorithm used to fix the new choice of j_1 is described in [18]. The bold diamond point in Figure 3 identifies the suitable value of j_1 as 14. Therefore, the regression is carried out again using $[j_1, j_2] = [14, 22]$, shown by the bold line in Figure 2. The more accurate estimate results are summarized in the Table 1. The mean estimate of the Hurst parameter is 0.964 and of Alpha is 0.928.

Octaves	Available	Selected	Goodness of Fit
	1-22	14-22	0.02640
Scaling Parameters	Alpha (LRD)	Hurst (LRD)	
	0.928	0.964	

Table 1: Estimates of the Hurst parameter and the Alpha parameter of IP packet size

From the results obtained from the LDestimate tool, it is quite clear that the IP packet size distribution shows the property of SS. The estimate of H (0.964) also suggests that the IP packet size process is characterised by strong LRD.

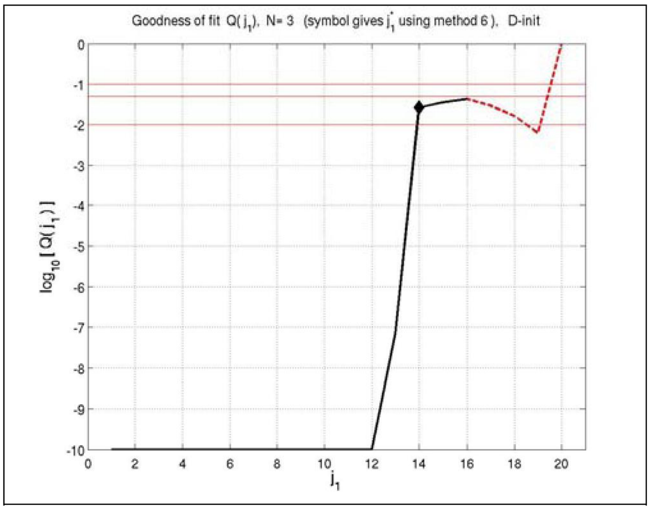

Figure 3: IP packet size – plot of the goodness of fit output by the LDestimate Tool

3.3 IP packet size - HTD analysis results

The principle of the Aest tool was introduced in [4]. In summary, X_i is a random variable whose distribution is S_α, a strictly α-stable distribution with $0 < \alpha < 2$. $X_i^{(m)}$ is the time-average aggregation of process X_i using the equation:

$$X_i^{(m)} = \sum_{j=(i-1)m+1}^{im} X_j \tag{3.3.1}$$

In Figure 4, for two adjacent aggregations $X_i^{(m_1)}$ and $X_i^{(m_2)}$ (m_1 and m_2 is the aggregation block), the horizontal distance δ and vertical distances τ are measured. Using the following equation, the heavy-tail index α can be obtained.

$$\alpha = \delta/\tau = \left(\ln(m_2/m_1)\right)/\delta \tag{3.3.2}$$

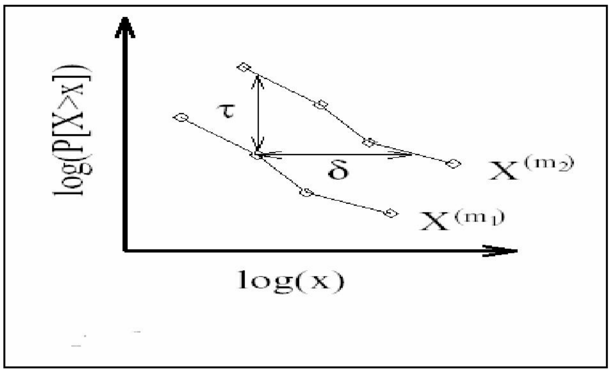

Figure 4: Principle of the scaling property in Aest tool [18]

In this analysis, a three-day measurement of packet size was input into Aest tool. The aggregation level selected is the default value 10 and m_1/m_2 equals 2 for simplicity. The bold points in Figure 5 identify the points that fall in the heavy-tailed section of the distribution. The estimate of the heavy-tailed index for IP packet size is 1.700169. The subtracted mean of the packet size is 322.645603 bytes. Therefore, it is concluded that the IP packet size distribution is heavy-tailed based on the estimate of the heavy-tailed index provided by the Aest tool using real GPRS traffic, the reason for this is the subject of further study.

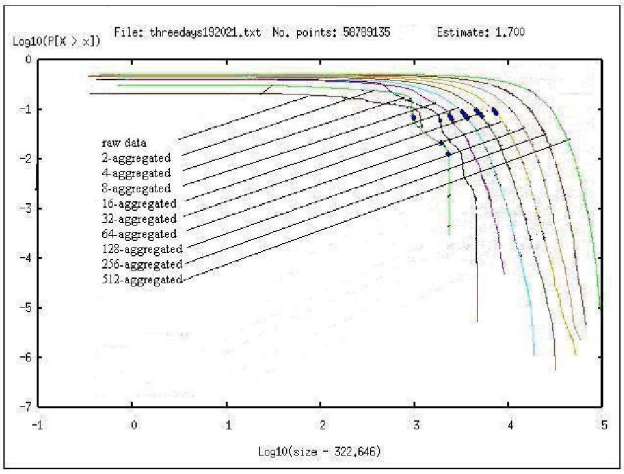

Figure 5: Estimate of heavy-tailed index by the Aest tool – for a three-day measurement of IP packet size

3.4 IP traffic - SS and LRD analysis results

For IP traffic, the burst is the aggregation of IP packet size data over a 100ms time interval. In this section, the analysis results output by the LASS and LDestimate tools are illustrated.

- <u>Results Obtained from the LASS Tool</u>

The measurements of IP packet size were aggregated in 100ms time intervals and input into the LASS tool.

Figure 6 (a) (top of the figure) shows the estimate of H against the window number for $[j_1, j_2]$ = [1, 13] and $[j_1, j_2]$ = [12, 13]. Figure 6 (b) (left-bottom) and 6 (c) (right-bottom) are the mean value and the standard deviation of IP traffic volume against the window number. According to the principle in [15], the Octave j interval is chosen as $[j_1, j_2]$ = [12, 13], then the mean estimate of H is 0.9312 in case. .

The estimate of the Hurst parameter of IP traffic output by the LASS tool results in the conclusion that the IP traffic exhibits the SS and LRD properties.

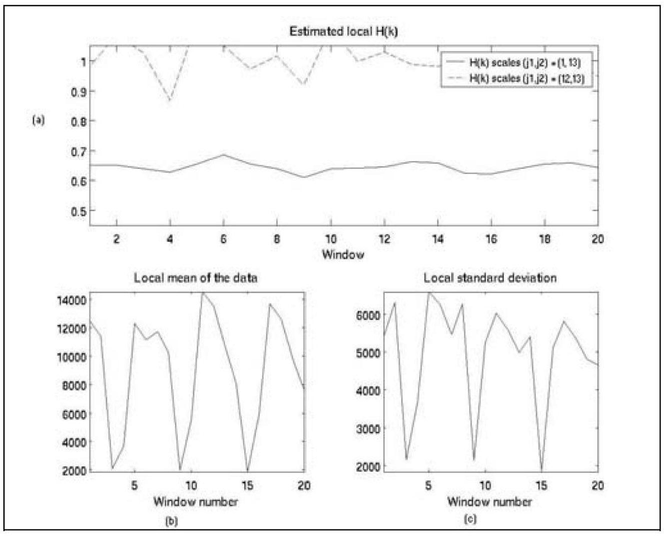

Figure 6: IP traffic- step 3 output of the LASS tool

- <u>Results obtained from the LDestimate tool</u>

The same IP traffic data was analysed using the LDestimate tool. According to the input data sample, the available Octave j range is from 1 to 18 (see Table 2). According to the principle in [18], the Octave j_1 is chosen as 11 to achieve the maximum of goodness of fit. In Figure 7, the bold line shows the regression from $[j_1, j_2]$ = [11, 18] and the final estimates are listed in the Table 2.

Octaves	Available	Selected	Goodness of Fit
	1-18	11-18	0.00006
Scaling Parameters	Alpha (LRD)	Hurst (LRD)	
	1.261	1.130	

Table 2: Estimate of the Hurst parameter and the Alpha parameter of IP traffic

From the estimates given in Table 2 it is clear that the IP traffic burst exhibits the SS property. Since the IP traffic burst is the aggregation of IP packet size, it confirms that IP packet size exhibits SS according to the definition of SS from aggregation point of view.

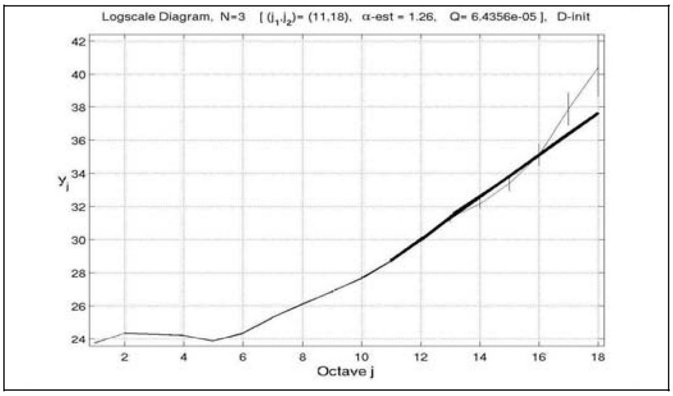

Figure 7: IP traffic - logscale diagram plot of LDestimate tool

4 Conclusions

The IP packets of mixed applications traffic including WWW, WAP, FTP, Email, SMS and MMS traffic were collected at a Gn-interface of a GPRS network. The tools LASS, LDestimate and Aest were applied to analyse and characterise the resulting real traffic trace, specifically according to its HTD, SS and LRD.

Based on the results provided by the tools LASS and LDestimate, it is concluded that both the IP packet size and the IP traffic burst exhibit the property of self-similarity especially long-range dependence. Specifically, the self-similarity property of the traffic burst shows that IP packet size is a self-similar process according to the definition of self-similarity. The results output by the Aest tool indicate that the IP packet size distribution also exhibits heavy-tailedness. In future work, the causes of the identified properties in wireless multimedia services traffic will be investigated. The differences in the nature of wired and wireless multimedia services traffic will be investigated in depth.

Acknowledgements

The authors thank Vodafone Group R&D- the Netherlands for providing live GPRS traffic data. The authors also thank Bart Sanders and Jeroen Wieland from Vodafone Group R&D- the Netherlands for their supervision and great support with obtaining the traffic data and with its analysis.

References

[1] P. Abry and D. Veitch. "Wavelet Analysis of Long-Range Dependent Traffic", *IEEE Transactions on Information Theory*, Vol 44, No.1, pp.2-15, (1998).

[2] P. Abry, P. Flandrin, M. S. Taqqu and D. Veitch. "Wavelets for the analysis, estimation and synthesis of scaling data", published as one chapter of book "Self-Similar Network Traffic and Performance Evaluation" 1st edition, John Wiley & Sons 2000.

[3] J. Beran. "Statistics for Long-Memory Process", Chapman and Hall, 1994.

[4] M. E. Crovella and M. S. Taqqu. "Estimating the Heavy Tail Index from Scaling Properties", *Methodology and Computing in Applied Probability*, Vol.1, No.1, pp.1-21, (1999).

[5] M. E. Crovella and A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes", *IEEE/ACM Transactions on Networking*, Vol.5, No.6, pp.835-846, (1997).

[6] R. G. Clegg, "A Practical Guide to Measuring the Hurst Parameter," *Proceedings of 21st UK Performance Engineering Workshop, School of Computing Science Technical Report Series*, CS-TR-916, University of Newcastle, ISSN 1368-1060 (2005).

[7] M. W. Garrett and W. Willinger. "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic", *ACM SIGCOMM*, London, Sep 1994.

[8] B. M. Hill. "A Simple General Approach to Inference about the Tail of a Distribution", *The Annals of Statistics*, Vol.3, pp 1163-1174, (1975).

[9] H.E. Hurst, R. Black and Y. Simaika. "Long-term storage: An Experimental study", London: Constable, 1965.

[10] R. Kalden and S. Ibrahim. "Searching for Self-Similarity in GPRS", *Proceedings of the 5th International Workshop on Passive and Active Network Measurement (PAM 2004),* Antibes Juan-les-Pins, France 19-20 Apr, 2004, pp.83-92.

[11] W. E. Leland and M. S. Taqqu, "On the Self-Similar Nature of Ethernet Traffic", *IEEE/ACM Transactions on Networking*, Vol.2, No.1, pp.11-15, (1994).

[12] A. Popescu. "Traffic Self-Similarity", *IEEE ICT 2001*, Bucharest, Romania, Jun 2001.

[13] V. Paxson and S. Floyd. "Wide-Area Traffic: The Failure of Poisson Modeling", *IEEE/ACM Transactions on Networking*, Vol.3 ,No.3, pp.226-244, (1995)

[14] M. Roughan, D. Veitch and P. Abry. "Real-Time Estimation of the Parameters of Long-Range Dependence", *IEEE/ACM Transactions on Networking*, 8 (4), pp. 467--478, (2000).

[15] S. Stoev, M. S. Taqqu and C. Park. "LASS: A Tool for the Local Analysis of Self-Similarity", *Statistical and Applied Mathematical Sciences Institute*, SAMSI Report 2004-7, 26 Mar 2004.

[16] W. Stallings. "High-Speed Networks, TCP/IP and ATM Design Principles", Prentice-Hall, Inc., 1998.

[17] M. S. Taqqu and V. Teverovsk. "On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series", preprint Boston University, USA, 1996.

[18] D. Veitch, M. Taqqu and P. Abry. "On the Automatic Selection of the Onset of Scaling", Fractals 11(4), 377-390, (2003).

[19] W. Willinger, M. S. Taqqu, R. Sherman and D.V. Wilson. "Self-similarity Through High-variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level", *IEEE/ACM Transactions on Networking*, 5(1): 71-86, Feb 1997.