Characterizing Internet Load as a Non-regular Multiplex of TCP Streams

J. Aracil and D. Morató
Dept. de Automática y Computación
Universidad Pública de Navarra
Campus Arrosadía s/n
31006 Pamplona, SPAIN
Tel: +34 948 16 97 33
Fax: +34 948 16 89 24
email: daniel.morato@unavarra.es, javier.aracil@unavarra.es

Abstract— **A commonly accepted traffic model for a large population of Internet users consists of a multiplex of Poisson-arriving heavy-tailed streams with the same constant rate (M/G/∞). We show that even though such regular model provides an accurate description of long-range dependence, the marginal distribution variance is underestimated, resulting in erroneous calculation of overflow probability in network simulations. On the other hand, we show that the traffic variability due to the marginal distribution variance can be the limiting factor for performance in the gigabit speed Next Generation Internet, rather than the long-range dependence features present in today's traffic.**

Keywords— **Internet performance analysis, heavy-tails, long-range dependence, traffic measurements.**

I. Introduction

A number of recent studies clearly show that Internet load is primarily determined by TCP connections coming from the WWW. Indeed, the traffic trace recorded from Public University of Navarra IP over ATM access link (February 2000), which is analyzed in this paper, shows that the TCP traffic percentage equals 99% in bytes transmitted, 82.8% of which (96.9% of the total number of TCP connections) are WWW connections, out of a sample of 1,029,350 TCP connections. Such high percentage is also reported in a number of measurements performed in a wide variety of academic and industrial settings [1], [2], [3].

Thus, since TCP connections determine Internet load almost completely, there is a growing interest in the development of traffic models which explain the main features of Internet traffic in terms of the underlying TCP connections multiplex. Indeed, Willinger et al. [4] show that the multiplex of on-off sources with heavy-tailed on-off periods turns out to have self-similar properties, a distinguishing feature of Internet traffic [4], [5]. Furthermore, Tsybakov and Georganas [6] show that as long as the on-period of the individual connections is heavy-tailed the resulting traffic multiplex is asymptotically second-order self-similar, even though the connection arrival process is Poisson. A heavy-tailed random variable R has a distribution tail with the form

$$P(R < r) \sim 1 - \left(\frac{K}{r}\right)^{-\alpha} \tag{1}$$

where α takes the values $1 < \alpha < 2$. The resulting random variable has finite mean but infinite variance. On the other hand, a stationary stochastic process in discrete time $X = \{X_t\} = \{X_1, X_2, \ldots\}$ is called asymptotically second-order self-similar with Hurst parameter H if for all $k \geq 1$ [6]

$$\lim_{m\to\infty} \rho^{(m)}(k) = \frac{1}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right] \tag{2}$$

being $\rho^{(m)}(k)$ the lag k autocorrelation of the aggregated process $X^{(m)}$,

$$X_t^{(m)} = \frac{1}{m}(X_{tm-m+1} + \ldots + X_{tm}), t \geq 1 \tag{3}$$

For $1/2 < H < 1$ this means that the correlation $\rho(k)$ decays to zero so slowly that,

$$\sum_k \rho(k) = \infty \tag{4}$$

The process X has long memory or *long-range dependence*. In the specific case of Internet traffic, such long-range dependence is observed in the process X_t which represents the number of bytes in fixed duration (δ ms) intervals from a sample of Internet traffic. As a consequence of the slow decay of the autocorrelation function the overflow probability in intermediate router queues increases heavily, in comparison to a process with independent increments (Poisson). In [7] an experimental queueing analysis with long-range dependent traffic is presented, which compares an original Internet traffic trace with a shuffled version, i.e. with destroyed correlations. The results show a dramatic impact in server performance due to long-range dependence.

The parameter α in (1) is related to the Hurst exponent H since $H = (3-\alpha)/2$ [6]: Values of H in the range $0.5 < H < 1$ indicate long-range dependence. Such H values correspond to values of α in the range $1 < \alpha < 2$.

Interestingly, the multiplex of heavy-tailed streams model [4], [6] assummes the same constant rate for all streams and heavy-tailedness in stream bytes and duration. We note that such constant rate does not have a physical meaning in the model construction [6]. Nevertheless, in [2] the model is adopted to explain WWW traffic self-similarity, since both size (bytes) and duration of WWW connections can be modeled with a heavy-tailed random variable (Pareto) with α values 1.15 for size and 1.2 for duration. As a model simplification, it is assummed that TCP connections have the same constant rate, so that the total traffic is a multiplex of homogeneous constant rate heavy-tailed streams. In [4] the same assumption is adopted to model LAN traffic. Furthermore, the Poisson-arriving heavy-tailed model with same constant rate per stream has been widely used for network dimensioning and simulation purposes [8] and is commonly accepted as a phenomenological model that explains long-range dependence of Internet traffic [2], [4].

In this paper we investigate the impact of non-regularity (different constant rate per stream) in the TCP streams in comparison to the regular (same constant rate per stream) case. Clearly, the rate of each individual TCP stream is not the same but varies according to the specific network settings between source and destination. By taking into account the non-homogeneity of TCP streams we observe higher variability in the marginal distribution, in comparison to the regular case. Since queueing performance depends not only on the long-range dependence properties of the incoming traffic but also on the marginal distribution variability the queuing delays predicted by the regular model are optimistic. Indeed, the introduction of different transfer modes for Internet transactions other than raw TCP, in accordance to the emerging high-speed IP switching and optical burst switching techniques, may lead to an increase of variability in the marginal distribution, while alleviating long-range dependence since connection duration decreases heavily. In such foreseeable scenario the dominant factor for network performance is the marginal distribution of the traffic multiplex, rather than long-range dependence, as we will show.

A. Network scenario and measurement tool

Our traffic traces are obtained from the ATM Permanent Virtual Circuit (PVC) that links Public University of Navarra to the core router of the Spanish academic network (*RedIris* [1]) in Madrid. The Peak Cell Rate (PCR) of the circuit is limited to 4 Mbps and the transmission rate in the optical fiber is 155 Mbps. We note that the scenario under analysis is a representative example of a number of very common network configurations. For example, the most Spanish Internet Service Providers (ISPs) hire ATM PVC links to the operators in order to provide customers with access to the Internet. On the other hand, measurements are not constrained by a predetermined set of destinations but represent a real example of a very large sample of users accesing random destinations in the Internet. Furthermore, we carefully check the utilization factor of the ATM PVC and note that never reaches 50% in intervals of 15 min. during the measurement campaign. This sanity check is performed to ensure that the results represent a general Internet case. Namely, different connections are facing different bottleneck links according to the destination, but the results are not correlated by a potential bottleneck link in the access. Finally, the wealth of data in the trace provides a strong confidence level in the obtained results. Measurements comprise one day worth of data starting Monday 14/02/2000 at midnight, making a total of 1029350 TCP connections (16375793 IP packets).

[1] http://www.rediris.es

The remainder of this paper is organized as follows: in section II we present a characterization of individual TCP streams. Section III is devoted to goodness of fit and comparison of the regular model (multiplex of same constant rate streams) and non-regular model (multiplex of different constant rate streams), followed by discussion in section IV and conclusions in section V.

II. TCP STREAMS CHARACTERIZATION

In order to construct a traffic model as a multiplex of TCP connections we need to address both connection arrival process and connection size, duration and rate. As far as the arrival process is concerned we study the connection interarrival time distribution. Then, we analyze connection duration and sizes and compare to results obtained in previous papers [1], [2]. On the other hand, we address the issue of non-regularity with an analysis of the connection rate.

A. Connection arrival process

Likewise most traffic models, the regular model presented in [4], [6] assumes stationarity in the arrival process, thus restricting the scope to time intervals over which such process can be assumed to be stationary. We select hourly intervals in the morning, afternoon and evening and plot the survival function of connection interarrival time ($P(X > x)$) in log-linear scale (Fig. 1). The lines shown in Fig. 1 indicate that the arrival process is non-stationary being interarrival times best modeled by exponential random variables in hourly intervals. Since users can be assumed to be independent one another the arrival process can be assumed to be Poissonian [8].

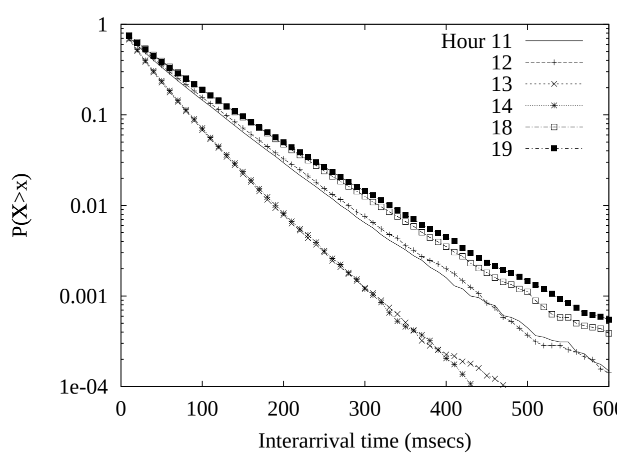

Fig. 1. Survival function of connection interarrival time

B. Connection size and duration

Fig. 2 shows survival functions of connection size (bytes) and duration (seconds) in log-log scales. We note that the survival function of the heavy-tailed random variable in (1) yields a line of slope $-\alpha$ when plotted in log-log scales. We plot the distribution tail least squared regression line in both plots and estimate values of α of 1.15 and 1.2 for duration and size respectively. Such values are in accordance with previous studies that report values of 1.2 and 1.15 [2].

Files sizes in the Internet are heavy-tailed due to the diverse nature of posted information which ranges from small text files to very large video files [2]. Regarding duration, we note an even larger variability (lower α), possibly due to the dynamics of the TCP in presence of congestion, which make connection duration grow larger if packet loss occurs. In sum, since the α values for size and duration are both in the range $1 < \alpha < 2$ the Hurst parameter H takes values in the range $0.5 < \alpha < 1$, showing strong long-range dependence features.

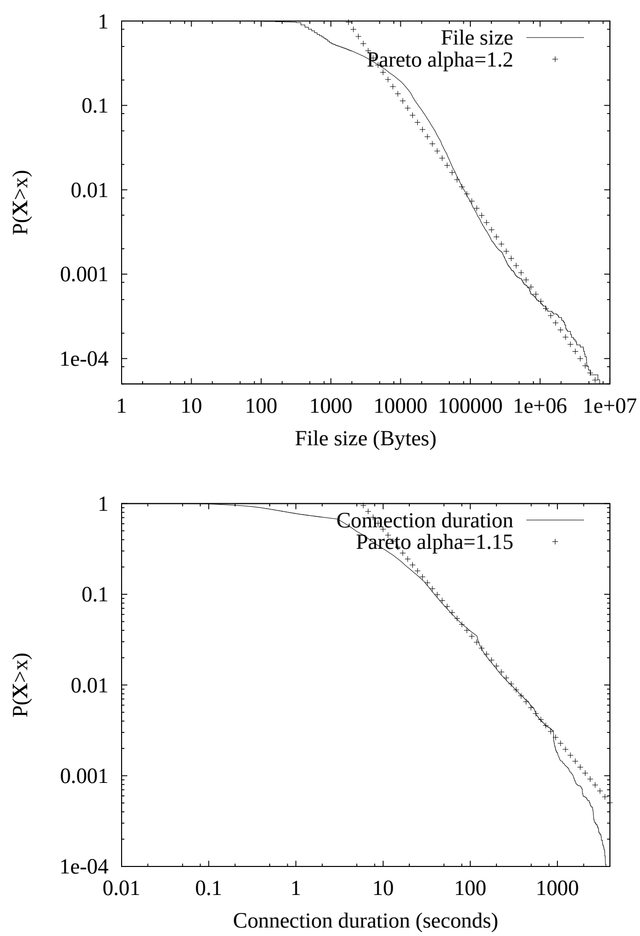

Fig. 2. Survival function of connection size (top) and duration (bottom)

C. Connection rate

While the regular model takes into account the traffic parameters listed so far, the variable rate of TCP connections is not captured by the model. Fig. 3 shows the survival function of TCP connections rate (namely total bytes transmitted from server to client divided by connection duration) in log-linear scale. First, we observe that rate is not constant, but differs from one connection to another. Secondly, such rate is best modeled by a Weibull random variable.

We note that the above distribution is not heavy-tailed. Should duration and size be independent random variables then the rate distribution would also have a heavy-tail, since both duration and size are indeed heavy-tailed. However, it turns out that size and duration are correlated, as we expect, and the resulting rate distribution is best fitted by a Weibull distribution.

III. Modeling the multiplex of TCP streams

In this section we compare both regular and non-regular model, which are defined as follows:

- Regular model: Internet traffic is a multiplex of Poisson-arriving heavy-tailed connections with the same constant rate. Specifically, we synthetize a multiplex with the same connection arrivals as in the trace (Poisson) and same connection duration (heavy-tailed) but we assumme that packets are transmitted at the same constant rate for all connections.
- Non-regular model: Internet traffic is a multiplex of Poisson-arriving connections with different constant rates. Now we synthesize a multiplex with the same connection arrivals as in the trace (Poisson), same connection size and duration (heavy-tailed) and same connection rate (Weibull).

In order to ensure a fair comparison we select a rate value for the regular model that yields the same load (bytes) as in the non-regular case. Then, we perform a trace-driven simulation of an infinite queue single-server system with both input

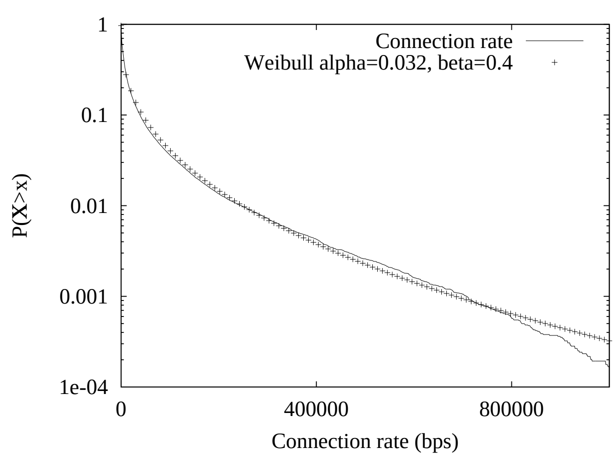

Fig. 3. Connection rate survival function

models, in order to evaluate network impact. The results are shown in Fig. 4, which presents delay versus utilization factor. A performance drop is observed for the non-regular traffic input model, which provides roughly the same performance curves obtained with the real traffic trace. We note that since the regular model is unable to capture the traffic marginal distribution burstiness it provides unrealistic results, as we explain next.

IV. Discussion

The analytical performance evaluation of an infinite-queue single-server system under self-similar input[2] is presented in [9]. The survival function of packet delay in the system is given by:

$$P(X > x) \sim exp\left(-\frac{(C-m)^{2H}}{2k(H)^2 c_v^2 m^2} x^{2-2H}\right) \quad (5)$$

where C is the link capacity, m the input traffic average rate, $c_v = \sigma/m$ the standard deviation divided by input traffic mean or marginal distribution variation coefficient [9] and $k(H) = H^H(1-H)^{(1-H)}$, being H the Hurst parameter. The previous equation shows that queueing performance depends on the following two variables, besides utilization factor: first, input traffic long-range dependence (H parameter) and, secondly, marginal distribution variability (c_v parameter). We argue that while the regular model captures long-range dependence accurately the same does not apply to the marginal distribution variation coefficient.

In order to verify our hypothesis we obtain the values of Hurst parameter H and variation coefficient c_v for regular, non-regular model and the real traffic trace. Both H and c_v parameters are obtained in hourly intervals, in which traffic may be assummed to be stationary. The coefficient of variation c_v can be obtained in a straightforward manner dividing standard deviation by mean. The Hurst parameter H is obtained through regression of variance-aggregation level plots. The results are displayed in table I and II.

We observe that the Hurst parameter takes similar values in all cases, while the constant rate model provides lower variation coefficient (ten times less than the real variation coefficient in some cases). Thus, the regular model provides an optimistic modeling of marginal distribution variability. For example, Fig. 5 shows the marginal distribution (survival function in log-linear scales) for both regular and non-regular models and real traffic trace in a two hours interval. We note that indeed the non-regular model distribution has a heavier tail than the regular counterpart, thus providing a better approximation of the real curve.

[2] Fractional Gaussian Noise

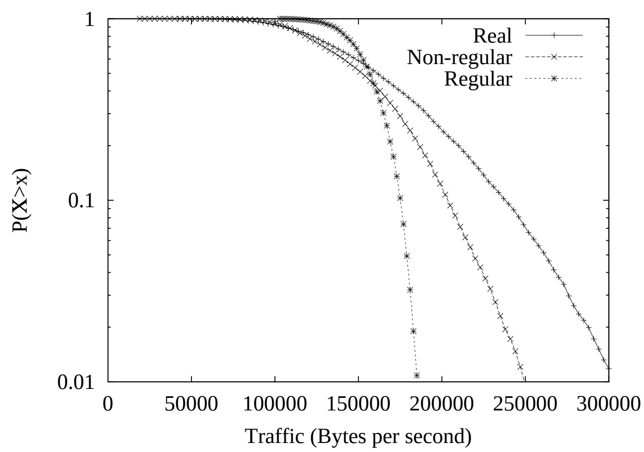

Fig. 5. Survival function of the marginal from both traffic models

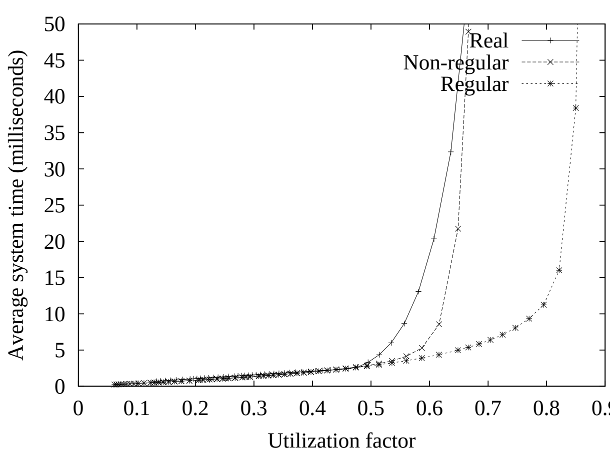

Fig. 4. Performance evaluation of single-server system with real, variable rate and constant rate input model

TABLE I
H ESTIMATION VERSUS TIME

Hour	Real	Non-reg	Regular
8	0.97	0.98	0.98
9	0.87	0.78	0.95
10	0.78	0.81	0.99
11	0.85	0.90	0.95
12	0.85	0.88	0.97
13	0.81	0.87	0.97
14	0.91	0.95	0.99
15	0.92	0.95	0.97
16	0.88	0.80	0.96
17	0.80	0.87	0.99
18	0.86	0.87	0.91
19	0.88	0.94	0.95
20	0.83	0.86	0.99
21	0.94	0.97	0.99

TABLE II
c_v ESTIMATION VERSUS TIME

Hour	Real	Non-reg	Regular
8	1.13	1.00	0.59
9	0.19	0.15	0.07
10	0.12	0.06	0.15
11	0.10	0.05	0.01
12	0.08	0.03	0.01
13	0.07	0.03	0.01
14	0.07	0.03	0.05
15	0.06	0.05	0.02
16	0.07	0.05	0.02
17	0.09	0.04	0.01
18	0.09	0.04	0.003
19	0.09	0.04	0.01
20	0.13	0.06	0.05
21	0.21	0.18	0.27

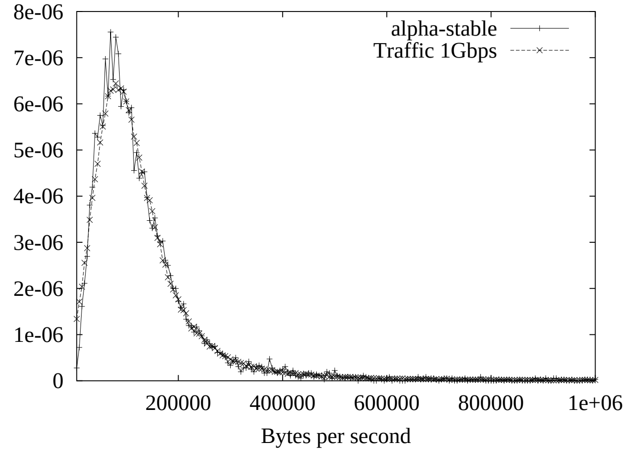

Fig. 6. Link traffic probability density function for a 1Gbps access link

A. Evolution of Internet traffic

The results presented in the last section show that Internet performance is not only determined by long-range dependence but also by marginal distribution variability. Considering the case of an Internet link which concentrates a large population of users we note that the connection arrival process can be considered Poissonian, due to inherent random demand of such users population. On the other hand, connection bytes can be modeled with a heavy-tailed random variable (infinite variance) due to wide variety of information objects in the Internet, from text files to video clips [2]. However, the long-range dependence features of Internet traffic are due to the multiplex of traffic streams whose *service times* are heavy-tailed. In fact, Fig. 2 shows that TCP connection duration is heavy-tailed. This is due to the heavy-tailed nature of file sizes, together with the limited bandwidth of the Internet links, which makes transmission time non-negligible. Furthermore, the dynamics of the TCP protocol also contribute to an increase in connection duration.

What if there is a phenomenal increase in network bandwidth? We are currently witnessing the development of optical technologies, which are offering data rates of 1.6 Tbps in a single fiber using DWDM technology. In order to take advantage of such technologies the TCP is clearly limited, due to the enormous bandwidth-delay product in the optical segment. Indeed, such optical technologies will be offering burst or coarse packet switching transfer modes [10], which will enable the transfer of files swiftly across the network in a single optical pulse. In such case, we note that the traffic marginal distribution variability is the limiting factor in queueing performance and not the long-range dependence features.

For example, we consider the case in which the link capacity is 1 Gbps (point-to-point Gigabit Ethernet) instead of the 4 Mbps of the IP over ATM link from which our trace was recorded. We implement a synthesized trace with the same connection arrivals as in the real case and also same file sizes, which are transmitted at the line speed of 1 Gbps. The marginal probability density function is plotted in Fig. 6, along with the goodness-of-fit curve to an α-stable random process [11].

Such α-stable random processes have higher variability in the marginal distribution in comparison to a Gaussian process (infinite variance in the marginal distribution). Since the input traffic now shows independent increments, due to the poissonian nature of connection arrivals and the near-infinite bandwidth, connection service times tend to be very small. Thus, the traffic marginal distribution variability is extremely large because of the heavy-tailed nature of file sizes in the Internet, an inherent characteristic of the current and future Internet. As a result, we note that the harmful Internet traffic burstiness is due to marginal distribution variability rather than long-range dependence in a hypothesized future high-speed file switching scenario.

In order to verify this hypothesis we perform a trace-driven simulation of a single-server infinite queue system with the original α-stable trace (original trace at 1 Gbps), shuffled α-stable trace and original low-speed trace (4 Mbps), whose results are presented in Fig. 7. We first note that the α-stable trace provides a dramatic performance decrease, in comparison to the low-speed counterpart. On the other hand, both shuffled and original α-stable trace provide nearly the same results. Since shuffling destroys any possible correlation in the trace we note that indeed it is the large marginal distribution variability and not long-range dependence the limiting factor for performance.

V. CONCLUSIONS

In this paper we have shown that Internet traffic presents higher marginal distribution variability than that predicted by

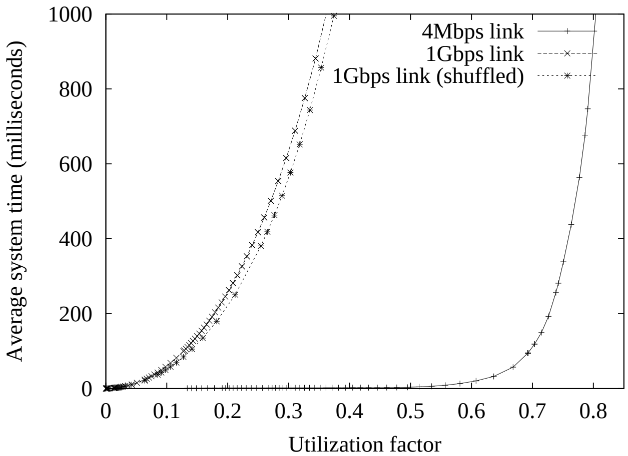

Fig. 7. Performance evaluation of single-server system with low-speed and α-stable traffic (original and shuffled)

a regular model of constant rate streams [4], [6]. Such traffic variability affects queueing performance heavily and, ultimately, may have a stronger influence on Internet performance than long-range dependence, as the current trend towards switched high-speed connections for IP flows in the backbone is on the rise in the networking scenario.

REFERENCES

[1] G. J. Miller K. Thompson and R. Wilder. Wide-area Internet traffic patterns and characteristics. *IEEE Network*, pages 10–23, November/December 1997.

[2] M. E. Crovella and A. Bestavros. Self-similarity in World Wide Web traffic: evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6):835–846, December 1997.

[3] D. Morató J. Aracil and M. Izal. Analysis of Internet services in IP over ATM networks. *IEEE Communications Magazine*, 37(12):92–97, December 1999.

[4] R. Sherman W. Willinger, M. S. Taqqu 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), February 1997.

[5] W. Willinger M. S. Taqqu and R. Sherman. Proof of a fundamental result in self-similar traffic modeling. *ACM SIGCOMM Compuer Communication Review*, 1997.

[6] B. Tsybakov and N. D. Georganas. On self-similar traffic in ATM queues: definitions, overflow probability bound and cell delay distribution. *IEEE/ACM Transactions on Networking*, 5(3):397–409, June 1997.

[7] O. Narayan A. Erramilli and W. Willinger. Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Transactions on Networking*, 4(2):209–223, April 1996.

[8] M. Nabe, M. Murata, and H. Miyahara. Analysis and modeling of World Wide Web traffic for capacity dimensioning for Internet access lines. In *SPIE Video, Voice and Data Communications Conference*, Dallas, TX, November 1997.

[9] I. Norros. On the use of Fractional Brownian Motion in the theory of connectionless networks. *IEEE Journal on Selected Areas in Communications*, 13(6):953–962, August 1995.

[10] C. Qiao M. Yoo. New optical burst-switching protocol for supporting quality of service. In *SPIE All-Optical Networking: Architecture, Control and Manage ment*, volume 3531, pages 396–405, Boston, MA, November 1998.

[11] G. Samorodnitsky and M. S. Taqqu. *Stable non-gaussian random processes.* Chapman & Hall, 1994.