Queuing Performance of Long-Range Dependent Traffic Regulated by Token-Bucket Policers

Stefano Bregni, *Senior Member, IEEE,* Roberto Cioffi, Paolo Giacomazzi

Politecnico di Milano, Dept. of Electronics and Information, Piazza Leonardo Da Vinci 32, 20133 Milano, ITALY
Tel.: +39-02-2399.3503 – Fax: +39-02-2399.3413 – E-mail: {bregni,giacomaz}@elet.polimi.it

***Abstract* — Long-range dependence (LRD) is a widely verified property of Internet traffic, which severely affects network performance yielding longer queuing delays. Token-bucket policers are commonly proposed to enforce the statistical profile of input traffic, but they can hardly reduce LRD and thus can be ineffective to protect service level agreements against LRD increase in input traffic. In this paper, we investigate by simulation the queuing performance in a downstream scheduler of LRD traffic regulated by token-bucket policers. We compare the delay distributions of regulated and unregulated LRD flows. We demonstrate that policers reduce significantly the downstream queuing delay, although they do not alter much the $1/f^{\alpha}$ spectrum of regulated traffic. We observe also that the policed traffic does not obey a plain fractional Gaussian traffic model: first, it is not Gaussian anymore; moreover, also its third-order two-lags covariance is altered by the policer. Finally, we point out that policers can reduce noticeably the negative impact of traffic LRD increase on queuing delay, although they can diminish actual LRD only slightly.**

***Index Terms* — Communication system traffic, fractional noise, Internet, long-range dependence, queuing analysis, traffic control (communication).**

I. Introduction

Self-Similarity (SS) and Long-Range Dependence (LRD) are widely verified properties of Internet traffic [1][2]. In a SS random process, a dilated portion of a realization, by the scaling Hurst parameter H, has same statistical characterization as the whole. On the other hand, LRD is a long-memory property, usually equated to an asymptotic power-law decrease of the power spectral density (PSD) as $\sim f^{-\alpha}$ (for $f \rightarrow 0$) or, equivalently, of the autocovariance function. Under some common hypotheses [1], the integral of a LRD process is SS with H related to α (e.g., fractional Brownian motion, integral of fractional Gaussian noise).

It is well recognized [3]–[6] that traffic LRD yields longer queuing delay in network buffers. In the case of fractional Gaussian traffic, for example, the delay tail is Weibull-distributed [3][4], i.e. the delay D exceeds a given threshold d with asymptotic probability

$$P(D > d) \approx e^{-\beta d^{1-\alpha}} \quad \text{for } d \rightarrow \infty \tag{1}$$

where $\beta > 0$ is function of α and other network parameters. Thus, the delay tail depends significantly on α: remarkably, when $\alpha \rightarrow 1$, the Weibull distribution flattens and average and variance of the queuing delay even tend to infinite.

LRD is an almost ubiquitous property of network traffic that is very hard to remove or control by Token-Bucket (TB) traffic regulators [7]–[12]. In previous papers [13][14], we demonstrated by simulation that TB policers and shapers can hardly reduce traffic LRD (i.e., its α parameter) and that statistical bounds of network delay may not be met against increase of LRD of offered traffic. This leads to suspect that providers might be unable of guaranteeing statistical delay bounds, if customers offer LRD traffic exhibiting large swings of α. We provided empirical evidence of this claim in [13][14].

In this paper, we investigate further the queuing performance in a downstream scheduler of LRD traffic regulated by token-bucket policers. We compare the delay tail distributions of regulated and unregulated LRD flows, in order to determine whether they differ and present different probability of violating delay thresholds, even with similar values of α.

At least to our knowledge, this aspect has been rather overlooked in literature till now. Cited papers [7]–[12] focused mainly on the limited effect of token-bucket policers on LRD of traffic, not emphasizing its behaviour in following queues.

A notable exception is the paper [8], which studied the queuing behaviour of LRD traffic regulated by a leaky-bucket policer. In this work, a leaky bucket followed by a scheduler with large finite buffer was studied analytically, characterizing the distribution of the time until buffer overflow and proving that the regulator does make the system overflow less often, whereas long-range dependence still makes its presence felt.

Actually, these theoretical results are asymptotic, describing the system behaviour when the scheduler finite buffer size tends to infinite, lacking however a quantitative analysis of the reduction of the buffer overflow probability due to traffic regulation. In our study, on the other hand, we evaluate the distribution of the delay in a scheduler with infinite buffer, following an empirical approach based on simulations. This way, we estimate delay bounds and associated violation probabilities, which are very important from a practical standpoint.

II. Self-Similarity and Long-Range Dependence

A random process $X(t)$ (e.g., cumulative packet arrivals in time interval $[0, t]$), is said to be *self-similar* (SS), with scaling parameter of self-similarity or Hurst parameter $H>0$, $H \in \Re$, if

$$X(t) =_d a^{-H} X(at) \tag{2}$$

for any $a>0$, where $=_d$ denotes equality for all finite-dimensional distributions [1]. In other terms, the statistical description of $X(t)$ does not change by *scaling* its amplitude by a^{-H} and its time by a. Most SS processes are not stationary.

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.

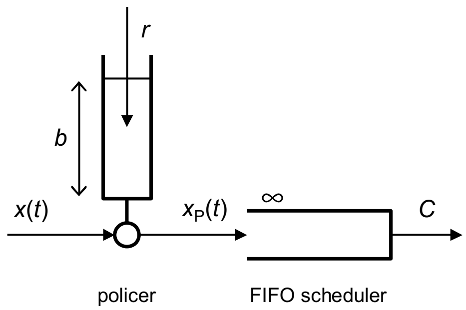

Fig. 1: Traffic policer followed by a FIFO scheduler.

The class of SS processes is usually restricted to that of *self-similar processes with stationary increments* (SSSI), which are "integral" of some stationary process. For example, consider the δ-increment process of $X(t)$, defined as $Y_\delta(t) = X(t) - X(t-\delta)$ (e.g., packet arrivals in the last δ time units). For a SSSI process $X(t)$, $Y_\delta(t)$ is stationary and $0 < H < 1$ [1].

Long-range dependence (LRD) of a process is defined by an asymptotic power-law decrease of its autocovariance and PSD [1]. Let $Y(t)$ be a 2nd-order stationary random process. $Y(t)$ exhibits LRD if its autocovariance follows asymptotically

$$R_Y(\tau) \sim c_1 |\tau|^{\alpha-1} \quad \text{for } \tau \to +\infty,\ 0 < \alpha < 1 \tag{3}$$

or, equivalently, its two-sided PSD follows asymptotically

$$S_Y(f) \sim c_2 |f|^{-\alpha} \quad \text{for } f \to 0,\ 0 < \alpha < 1 \tag{4}$$

In general, a random process with non-integer power-law PSD is also known as fractional (not necessarily Gaussian) noise. SSSI processes $X(t)$ with $1/2 < H < 1$ have LRD increments $Y(t)$, with [1]

$$\alpha = 2H - 1 \tag{5}$$

III. Source Traffic Regulation for Guaranteeing QoS

In Internet models of Quality of Service (QoS), the customer contracts with the Internet Service Provider (ISP) for the transport of flows under a *Service Level Agreement* (SLA), specifying the QoS requirements that the ISP must meet. In this paper, we focus on statistical delay bounds [15], commonly defined as the maximum fraction of packets $P_{\max}$ exceeding a given end-to-end delay limit $d_{\max}$.

The contract between customer and ISP includes a *Traffic Conditioning Agreement* (TCA), which describes the statistical profile of traffic allowed to enter the network in order to guarantee the SLA. The ISP allocates resources based on TCA parameters, which usually include [16]: average rate [byte/s], burst size [byte], peak rate [byte/s], minimum policed unit [byte] and maximum packet length [byte].

To enforce the TCA, a common solution is using traffic regulators based on the token-bucket scheme. If the source traffic complies with the TCA (*in-profile* traffic), it is transferred unaltered by the regulator. Otherwise, if the traffic violates the TCA (*out-of-profile* traffic), it is dropped (*policer*) or delayed in an internal buffer (*shaper*), until it is possible to inject it into the network complying with the TCA.

IV. Models of the System and LRD Traffic

We adopted a fluid traffic model [4], where traffic units are bits. As shown in Fig. 1, a TB policer is followed by a First-In-First-Out (FIFO) scheduler with infinite buffer and link capacity C [bit/s]. The TB policer has a counter of credits (tokens) with maximum value b [bit] (*token bucket size*). The credit counter is increased every $1/r$ s, where r is the *token rate*. One bit of offered traffic is allowed through the policer if the counter is positive; then, the counter is decremented. Otherwise, if the counter is equal to zero, the bit is dropped.

The r parameter controls the average rate of the through traffic, as the policer cannot output more than r bit/s on the average. The b parameter, instead, controls the length of output traffic bursts. If the token counter is full (i.e., it holds b tokens), the policer can output a burst of b bits at maximum rate. After that, it must stop to wait further tokens.

As far as the LRD traffic model is concerned, in this paper we focus on fractional Gaussian traffic (fGt), being this model commonly adopted in literature. Our procedure of traffic synthesis, detailed in [13], generates LRD pseudorandom series $\{x_k\}$ of fractional Gaussian traffic $\mathrm{fGt_R}(\alpha, m_x, \sigma_x^2)$ with length N, PSD $\propto 1/f^\alpha$ for assigned values of α with $0 \le \alpha < 1$, normally-distributed samples, mean m_x and variance σ_x^2, rectified to replace negative samples with zero. The sequence $\{x_k\}$ represents the incremental data count [bit/s] input at each time unit into the policer under study (i.e., the input traffic rate).

Also other models and methods have been proposed in literature to generate pseudo-random LRD traffic. To be safe, we repeated most simulations synthesizing multifractal log-normal input traffic instead. In all cases, we obtained substantially same results as those presented in this paper.

V. Simulation Results

We generated $\mathrm{fGt_R}$ sequences $\{x_k\}$ made of $N = 2^{23}$ samples. We set the time unit $\tau_0 = 1$ ms, the mean $m_x = 2279$ bit per time unit (i.e., 2.279 Mbit/s) and the deviation $\sigma_x = 773.9$ bit per time unit (i.e., 773.9 kbit/s), as in [4]. We varied α liberally in range $0 \le \alpha < 1$.

As in Fig. 1, the traffic $x(t)$ was fed into the policer. The regulated traffic $x_P(t)$ was then analyzed, characterizing it for various values of the token rate r and size b. In particular, we estimated the value of its LRD parameter α_P and we studied the distribution of its queuing delay in the scheduler buffer.

For estimating accurately α, we used the Modified Allan Variance (MAVAR) [17]. MAVAR is a well-known time-domain quantity, conceived in 1981 for frequency stability characterization of precision oscillators [18]–[21] by modifying the definition of the original Allan Variance. MAVAR was proven to feature superior spectral sensitivity and accuracy in LRD parameter estimation, coupled with excellent robustness against data nonstationarity (e.g., drift and steps) [17].

Fig. 2 plots the *loci* of the $(r/m_x,\ b/(\sigma_x\tau_0))$ pairs (contour lines), for which the same α_P was estimated (grand average

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.

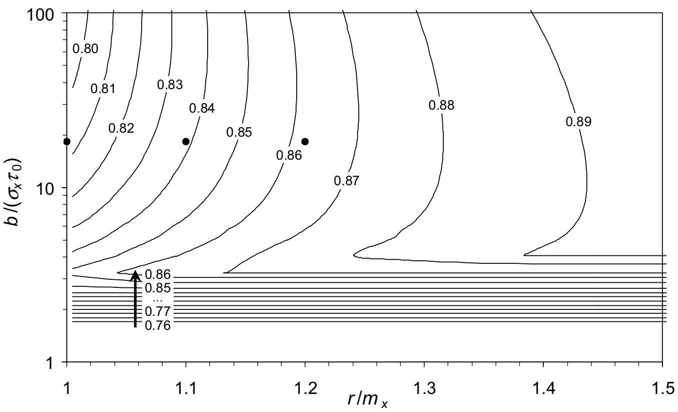

Fig. 2: Contour lines of $\alpha_P(r/m_x, b/(\sigma_x\tau_0))$ estimated on the traffic regulated by a TB policer fed with fGt$_R$ ($\alpha = 0.90$).

out of 10 independent simulations) on the traffic regulated by the policer with $\alpha = 0.90$ at input (cf. less accurate graphs in [13][14]). First, we notice that $\alpha_P < \alpha$. For $r/m_x \to \infty$ and $b/(\sigma_x\tau_0) \to \infty$, we have $\alpha_P \to \alpha$. Also, α_P decreases as r/m_x and $b/(\sigma_x\tau_0)$ are smaller. Finally, $\alpha_P(r, b)$ is non-monotonic: note the steep cliff for $1 < b/(\sigma_x\tau_0) < 3$ that takes, after the edge at $b/(\sigma_x\tau_0) = 3$, to the wide, gently hollow plateau.

We know from theory [4] that the queuing delay of fGt is distributed as eq. (1), where α determines the tail decrease rate. When the policer operates in normal conditions (i.e., $r/m_x > 1$ and $b/(\sigma_x\tau_0) > 10$), the traffic dropping ratio is low and $\alpha_P \cong \alpha$ (Fig. 2). Thus, we would expect intuitively that the queuing delay too of $x_P(t)$ is distributed likewise that of $x(t)$.

System simulation sharply contradicts this conjecture. For $r/m_x = 1.0$, 1.1, 1.2 and $b/(\sigma_x\tau_0) = 18.3$ (three dots marked in Fig. 2), we obtained about the following values for parameters (m_{xP}/m_x, σ_{xP}/σ_x, α_P) of the policed traffic $x_P(t)$: (0.90, 0.70, 0.806), (0.94, 0.80, 0.838) and (0.97, 0.88, 0.862), respectively. We fed $x_P(t)$ into the FIFO scheduler with capacity set to have uniform load $m_{xP}/C = 0.80$. The distributions of the resulting queuing delay in these three configurations are plotted in Fig. 3, as dotted lines labelled "policed traffic".

Then, we generated other fresh fGt$_R$ series $x_S(t)$, with same three parameters as those estimated on policed traffic $x_P(t)$ in the settings above, and fed them directly without regulation into the same FIFO scheduler. The resulting delay distributions are plotted in Fig. 3, as solid lines labelled "synthetic fGt$_R$".

The difference between delay distributions of $x_P(t)$ and $x_S(t)$ plotted in Fig. 3 is dramatic: the delay tail of policed traffic $x_P(t)$ is much shorter than that of fresh fGt$_R$ traffic $x_S(t)$ with same parameters $m_{xS} = m_{xP}$, $\sigma_{xS} = \sigma_{xP}$, $\alpha_S = \alpha_P$. For example, for $r/m_x = 1.2$, delay thresholds exceeded with probability 10^{-3} are about 6 ms and 7 s, respectively for $x_P(t)$ and $x_S(t)$, with traffic dropping ratio around 3%. In other words, the policer cuts down the downstream queuing delay by three orders of magnitude, by dropping just a little percentage of data, yet without diminishing traffic LRD significantly.

Thus, we proceeded to ascertain whether there exists any fGt$_R$ traffic $x_S(t)$ that yields same queuing delay, fed directly into the downstream FIFO buffer, than policed traffic $x_P(t)$ with $m_{xS} = m_{xP}$, $\sigma_{xS} = \sigma_{xP}$ for any $\alpha_S \neq \alpha_P$.

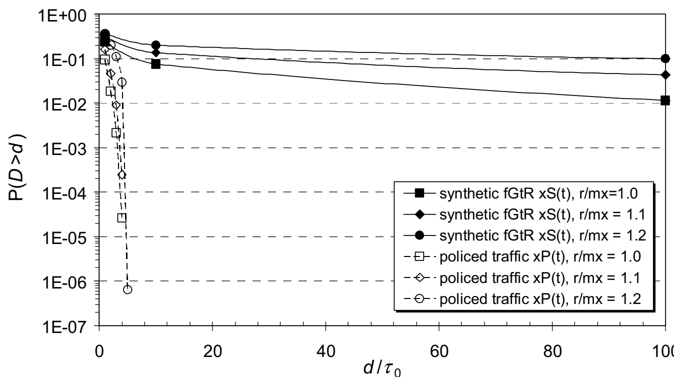

Fig. 3: Distribution of queuing delay in the FIFO buffer of policed traffic $x_P(t)$ and synthetic fGt$_R$ traffic $x_S(t)$ with same $m_{xS} = m_{xP}$, $\sigma_{xS} = \sigma_{xP}$, $\alpha_S = \alpha_P$.

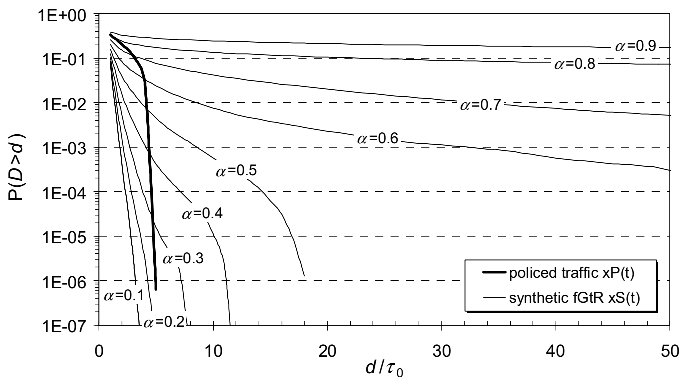

Fig. 4: Distribution of queuing delay in the FIFO buffer of policed traffic $x_P(t)$ and synthetic fGt$_R$ traffic $x_S(t)$ with same $m_{xS} = m_{xP}$, $\sigma_{xS} = \sigma_{xP}$ and various α_S.

In Fig. 3, we selected the delay distribution of $x_P(t)$, i.e. policed fGt$_R$ traffic with $\alpha = 0.90$ and $r/m_x = 1.2$ (empty round markers, $m_{xP}/m_x = 0.97$, $\sigma_{xP}/\sigma_x = 0.88$, $\alpha_P = 0.862$). This delay distribution is plotted again in Fig. 4 as a thick line. Then, we generated 9 new fresh fGt$_R$ series $x_S(t)$ with same parameters $m_{xS} = m_{xP}$, $\sigma_{xS} = \sigma_{xP}$ but varying $0.0 \le \alpha_S \le 0.9$. Their resulting delay distributions are plotted in Fig. 4 as thin lines. Fig. 4 shows that there is no fGt$_R$ traffic series $x_S(t)$ with same parameters $m_{xS} = m_{xP}$, $\sigma_{xS} = \sigma_{xP}$ that, for any value of α_S, yields similar queuing delay distribution than policed traffic $x_P(t)$.

In [13][14], we have shown that if the input traffic exhibits increasing LRD parameter α, still keeping same average rate and variance, policers deliver traffic downstream with α almost unaffected, causing possible disruptions of end-to-end delay SLA. While this is certainly true, now we can show also that the regulation action of the policer is however effective to improve the queuing performance in downstream schedulers.

We studied the distribution of queuing delay in the FIFO scheduler of fGt$_R$ traffic for various α, with and without policer regulation. The scheduler link capacity was set to have uniform load $m_x/C = 0.80$ or $m_{xP}/C = 0.80$ for both policed and unpoliced traffic series. The policer parameters were set $r/m_x = 1.2$ and $b/(\sigma_x\tau_0) = 18.3$. Simulation results are shown in Fig. 5. Even if the policer does not reduce significantly the α parameter of regulated traffic (Fig. 2), the resulting queuing delay in the downstream scheduler is much lower.

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.

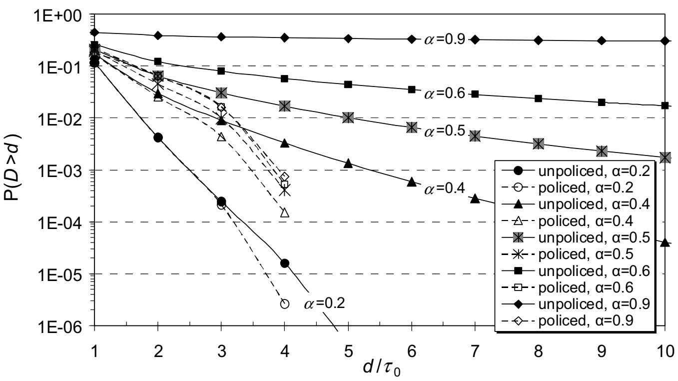

Fig. 5: Distribution of queuing delay in the FIFO buffer of unpoliced and policed fGt_R traffic for various α and $r/m_x = 1.2$, $b/(\sigma_x \tau_0) = 18.3$.

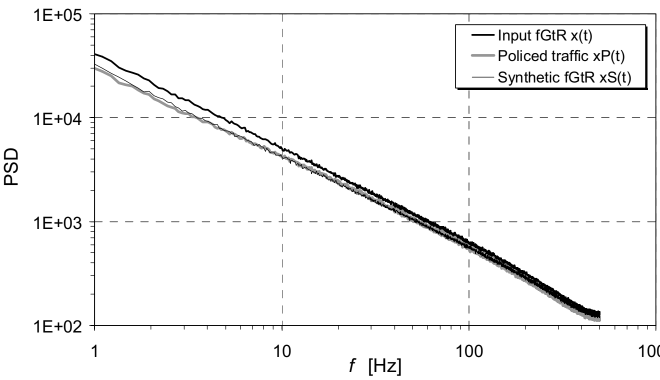

Fig. 6: PSD of a fGt_R series $x(t)$ with $\alpha = 0.9$, of $x_P(t)$ obtained by policing $x(t)$ with $r/m_x = 1.2$ and $b/(\sigma_x \tau_0) = 18.3$, and of synthetic fGt_R $x_S(t)$ with same parameters $m_{xS} = m_{xP}$, $\sigma_{xS} = \sigma_{xP}$, $\alpha_S = \alpha_P$ as of $x_P(t)$.

For example, let us assume that the SLA specifies $P(D > 4\text{ ms}) \leq 10^{-4}$, while the customer supplies fGt_R traffic with $\alpha = 0.2$. By inspection of Fig. 5 ($\tau_0 = 1$ ms), we see that this SLA is fulfilled. Nevertheless, if input traffic $x(t)$ keeps same m_x and σ_x but increases LRD to $\alpha = 0.6$, we observe again from Fig. 5 that the SLA is now violated, resulting to disruption of the required QoS. However, after this LRD increase, it is worth noticing that $P(D > 4\text{ ms}) = 5 \cdot 10^{-4}$ for policed traffic, while without regulation it is greater by two orders of magnitude. This positive impact of the policer is even sharper assuming LRD increase to $\alpha = 0.9$. In Fig. 5, it is also worth noticing that any LRD increase for $\alpha > 0.5$ does not augment $P(D > d)$ significantly on the policed traffic, while the effect on unpoliced traffic is dramatic.

To summarize, Figs. 3 and 4 show that the queuing delay of policed traffic $x_P(t)$ is much lower than that of synthetic fGt_R $x_S(t)$ with same parameters $m_{xS} = m_{xP}$, $\sigma_{xS} = \sigma_{xP}$, $\alpha_S = \alpha_P$. In particular, Fig. 4 shows that there is no fGt_R traffic series $x_S(t)$ with same $m_{xS} = m_{xP}$, $\sigma_{xS} = \sigma_{xP}$ that, for any value of α_S, yields similar queuing delay distribution than policed traffic $x_P(t)$.

Based on this evidence, we conclude that the plain fGt_R model is not adequate to describe the policed traffic $x_P(t)$, even when the policer operates in normal conditions (i.e., $r/m_x > 1$ and $b/(\sigma_x \tau_0) > 10$), the traffic dropping ratio is low and $\alpha_P \cong \alpha$ (Fig. 2), at least as far as evaluation of queuing performance in downstream schedulers is concerned.

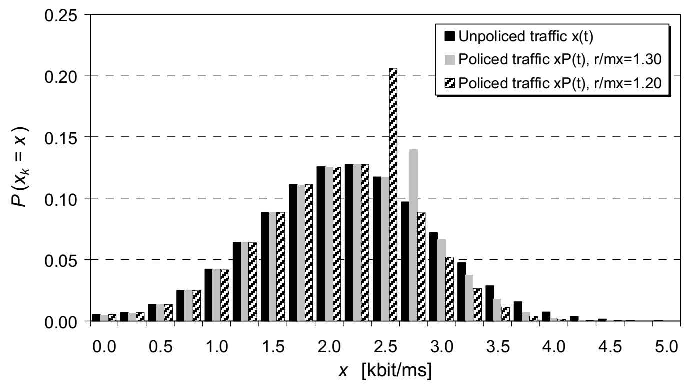

Fig. 7: Probability distributions of samples of the input fGt_R series $\{x_k\}$ ($m_x = 2.279$ kbit/ms, $\sigma_x = 773.9$ bit/ms, $\alpha = 0.9$) and of the policed traffic series $\{x_{Pk}\}$ for r/m_x=1.2, 1.3 and b/σ_x=18.3.

To prove further that the policer in such configuration does not affect much the spectral characteristics of regulated traffic (i.e., its second-order statistics), in Fig. 6 we compare the PSDs computed on a fGt_R series $x(t)$ with $\alpha = 0.9$, on traffic $x_P(t)$ obtained by policing $x(t)$ with $r/m_x = 1.2$ and $b/(\sigma_x \tau_0) = 18.3$, and on another fGt_R series $x_S(t)$ synthesized with same parameters $m_{xS} = m_{xP}$, $\sigma_{xS} = \sigma_{xP}$, $\alpha_S = \alpha_P$ as of $x_P(t)$. PSD has been estimated by FFT-based periodogram over 8192 points, having divided the series in 1024 segments with Welch data windowing [22]. We note that the PSDs in Fig. 6 are nearly identical, except the slight α reduction already pointed out.

In spite of its $1/f^{\alpha_P}$ spectrum, as a matter of fact, the policed traffic $x_P(t)$ does not obey a plain fGtR model under a number of aspects.

First, the policed traffic $x_P(t)$ is not Gaussian. The policer alters noticeably the probability distribution of traffic samples x_k. Fig. 7 compares the probability distributions $P(x_k = x)$ of samples of the input fGt_R series $\{x_k\}$ with parameters ($m_x = 2.279$ kbit/ms, $\sigma_x = 773.9$ bit/ms, $\alpha = 0.9$) and of the policed traffic series $\{x_{Pk}\}$ for $r/m_x = 1.2$, 1.3 and $b/\sigma_x = 18.3$. In the policed traffic distribution, note the concentrated probability for $x_k = r$, due to arrival of r or more bits of traffic in a time unit when the token counter is empty.

In addition, we noticed also that the policer alters significantly the third-order statistical correlation of the process $x(t)$. The third-order covariance of a stationary process $x(t)$ may be defined as the two-lags function (cf. eq. (3))

$$C_x(\tau_1, \tau_2) = E\left[\left(x(\tau) - m_x\right)\left(x(\tau + \tau_1) - m_x\right)\left(x(\tau + \tau_2) - m_x\right)\right] \quad (6).$$

We generated 1000 independent fGt_R sequences $\{x_k\}$ with $\alpha = 0.9$. We regulated them by a policer with $r/m_x = 1.2$ and $b/(\sigma_x \tau_0) = 18.3$. Then, we estimated the covariances $C_x(k_1, k_2)$ and $C_{xP}(k_1, k_2)$ of the input and policed traffic, respectively, by averaging the 1000 covariances estimated on single sequences. The results are plotted in Figs. 8 and 9. We observe that the third-order covariance of the traffic series has been altered substantially by the policer, although the PSD has been affected only slightly (Fig. 7).

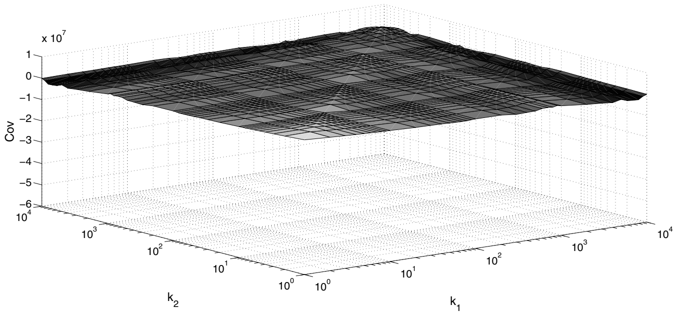

Fig. 8: Third-order covariance $C_x(k_1, k_2)$ of input fGt$_R$ traffic $x(t)$ ($\alpha = 0.9$).

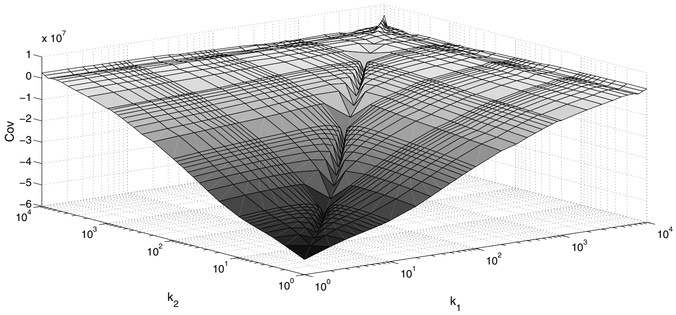

Fig. 9: Third-order covariance $C_{xP}(k_1, k_2)$ of the policed traffic $x_P(t)$ ($\alpha = 0.9$, $r/m_x = 1.2$, $b/(\sigma_x \tau_0) = 18.3$).

VI. Conclusions

In this paper, we investigated by simulation the queuing performance of LRD traffic regulated by token-bucket policers. In particular, we compared the distributions of the delay in a downstream scheduler buffer of regulated and unregulated LRD flows, in order to determine whether they present different probability of violating delay thresholds.

First, more extensive simulations confirmed and refined results already presented in [13][14]:

- TB policers hardly reduce traffic LRD (Fig. 2);
- in normal operation (i.e., $r/m_x > 1$ and $b/(\sigma_x \tau_0) > 10$, the customer meets the TCA and few traffic is dropped), the policer does not alter much the $1/f^{\alpha}$ PSD of traffic (Fig. 6), i.e. its 2nd-order statistics, and its LRD parameter $\alpha_P \cong \alpha$ (Fig. 2).

In spite of this, policers reduce significantly the queuing delay in downstream schedulers. The delay of policed traffic $x_P(t)$ is much lower than that of input traffic $x(t)$ and of synthetic fGt$_R$ $x_S(t)$ with same parameters as of $x_P(t)$ (Figs. 3, 4 and 5). Furthermore, there is no fGt$_R$ traffic, for any value of α, which yields similar delay distribution as policed traffic with same mean and variance (Fig. 4).

Based on this evidence, we concluded that the policed traffic $x_P(t)$ does not obey a plain fGt$_R$ model, even when the policer operates in normal conditions and the PSD of $x_P(t)$ is $\sim 1/f^{\alpha_P}$, at least as far as evaluation of queuing performance in downstream schedulers is concerned.

Actually, the policed traffic $x_P(t)$ is not Gaussian anymore (Fig. 7). Moreover, the policer alters significantly also the third-order two-lags covariance of the traffic (Figs. 8 and 9).

Finally, we point out that the positive effect of the policer on traffic queuing performance is significant: although traffic LRD is reduced only slightly, the probability of exceeding delay thresholds may be cut down even by some orders of magnitude, thus reducing noticeably the negative effects of possible LRD increase in input traffic.

References

[1] K. Park, W. Willinger, "Self-Similar Network Traffic: An Overview"; P. Abry, P. Flandrin, M. S. Taqqu, D. Veitch, "Wavelets for the Analysis, Estimation, and Synthesis of Scaling Data"; in *Self-Similar Network Traffic and Performance Evaluation*, K. Park, W. Willinger, Eds. Chichester, UK: John Wiley & Sons, 2000, pp. 1-88.

[2] V. Paxson, S. Floyd, "Wide-Area Traffic: the Failure of Poisson Modeling", *IEEE/ACM Trans. Networking,* vol. 3, no. 6, pp. 226-244, 1995.

[3] N. G. Duffield, Neil O'Connell, "Large Deviations and Overflow Probabilities for The General Single Server Queue, with Applications", *Math. Proc. Cam. Phil. Soc.*, vol. 118, pp. 363-374, 1995.

[4] I. Norros, "On The Use of Fractional Brownian Motion in the Theory of Connectionless Networks", *IEEE J. Select. Areas Commun.*, vol. 13, no. 6, pp. 953–962, Aug. 1995.

[5] Y. Cheng, W. Zhuang, "Calculation of Loss Probability in a Partitioned Buffer with Self-Similar Input Traffic", *Proc. IEEE GLOBECOM 2004*.

[6] J. Choe, N. B. Shroff, "Queueing Analysis of High-Speed Multiplexers Including Long-Range Dependent Arrival Processes", *Proc. IEEE INFOCOM 1999*, New York, NY, USA, March 1999.

[7] S. Vamvakos, V. Anantharam, "On the Departure Process of a Leaky Bucket System with Long-Range Dependent Input Traffic", *Queueing Systems*, vol. 28, pp. 191-214, 1998.

[8] B. Gonzalez-Arevalo, "Performance of a leaky bucket system with long-range dependent input traffic", Queueing Systems, Vol. 46, No. 3-4 pp. 439-459, mar-apr 2004.

[9] S. Molnár and A. Vidács, "On Modeling and Shaping Self-Similar ATM Traffic", *Proc. 15th International Teletraffic Congress (ITC15)*, Washington, DC, USA, July 1997.

[10] K. Christensen, V. Ballingam, "Reduction of Self-Similarity by Application-level Traffic Shaping", *Proc. IEEE 22nd Annual Conference on Local Computer Networks*, Minneapolis, MN, USA, Nov. 1997.

[11] F. C. Harmantzis, D. Hatzinakos, I. Katzela, "Shaping and Policing of Fractal α-Stable Broadband Traffic", *Proc. IEEE Canadian Conf. on Elec. and Comp. Eng.*, May 2001.

[12] F. de M. Pereira, N. L. S. da Fonseca, D. S. Arantes, "Fractional Bounded Arrival Process for Self-Similar Traffic Policing", *Proc. IEEE ICC 2003*, Anchorage, AK, USA, May 2003.

[13] S. Bregni, P. Giacomazzi, G. Saddemi, "Properties of the Traffic Output by a Leaky-Bucket Policer with Long-Range Dependent Input Traffic", *Proc. IEEE ICC 2007*, Glasgow, UK, June 2007.

[14] S. Bregni, P. Giacomazzi, G. Saddemi, "Output Traffic Characterization of Policers and Shapers with Long-Range Dependent Input Traffic", *Proc. IEEE GLOBECOM 2007*, Washington, DC, USA, Nov. 2007.

[15] J. Qiu, E. Knightly, "Inter-Class Resource Sharing Using Statistical Service Envelopes", *Proc. IEEE INFOCOM '99*, New York, Mar. 1999.

[16] S. Shenker, J. Wroclawski, "General Characterization Parameters for Integrated Service Network Elements", *IETF RFC 2215*, Sept. 1997.

[17] S. Bregni, L. Jmoda, "Accurate Estimation of the Hurst Parameter of Long-Range Dependent Traffic Using Modified Allan and Hadamard Variances", to appear in *IEEE Trans. Commun.*, 2008. Extended version available: http://home.dei.polimi.it/bregni/public.htm.

[18] D. W. Allan, J. A. Barnes, "A Modified Allan Variance with Increased Oscillator Characterization Ability", *35th Freq. Contr. Symp.*, 1981.

[19] P. Lesage, T. Ayi, "Characterization of Frequency Stability: Analysis of the Modified Allan Variance and Properties of Its Estimate", *IEEE Trans. Instrum. Meas.*, vol. 33, no. 4, pp. 332-336, Dec. 1984.

[20] L. G. Bernier, "Theoretical Analysis of the Modified Allan Variance", *Proc. 41st Annual Freq. Contr. Symp.*, 1987.

[21] S. Bregni, "Chapter 5 - Characterization and Modelling of Clocks", in *Synchronization of Digital Telecommunications Networks*. Chichester, UK: John Wiley & Sons, 2002, pp. 203-281.

[22] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vettering, *"Numerical Recipes in C – The Art of Scientific Computing. 2nd Edition"*, Cambridge, UK: Cambridge University Press, 2002.