Is the apparent self-similarity of the broadband traffic due to non stationarity?

S.Vaton, H.Korezlioglu and T.Chonavel

ENST, 46 rue Barrault, 75634 Paris Cedex 13, France.
ENST Bretagne, BP832, 29285 Brest Cedex, France.

Abstract

This contribution concerns the statistical characterization of the traffic involved by a Local Area Network with an objective of flow control on Asynchronous Transfer Mode networks. Many authors investigate the traffic measured on broadband networks on long time scales and they all conclude that the measured traffic is self-similar. Our approach is rather different. We suspect that the apparent self-similarity of the traffic is due to the presence of non stationarities in a traffic with no self-similarity. To support our intuition we test the stationarity of the measured traffic on different time scales ranging from a few seconds to hours of traffic. We then propose to model the measured traffic as a locally stationary and semi-Markov processes. We study into details the estimation of the parameters of semi-Markov processes; both block and recursive procedures of estimation are exposed. We eventually generate three different non stationary and semi-Markov processes whose parameters are matching the estimates of the varying parameters of the measured traffic. We prove that if one relied on classical visual indexes of self-similarity one would conclude that the synthetic traffic is self-similar. These findings question the consensus about the possible self-similarity of the traffic and they lead the way to some realistic and tractable models.

Keywords: Ethernet, ATM, long range dependence, self-similarity, non stationarity, local stationarity, test for stationarity, hidden Markov model, Markov modulated Poisson process, block estimation, recursive estimation, variance-time analysis, R/S analysis.

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-0-387-35353-1_28

D. Kouvatsos (ed.), *Performance Analysis of ATM Networks*

1 Introduction

A large number of contributions have been devoted to the statistical analysis of the traffic measured on broadband networks after the pioneering paper by Willinger *et al.*[1], who presented for the very first time the analysis of high quality LAN traffic data, recorded at Bellcore. Soon afterwards, a large number of data became available (not only for LAN, but also WAN, VBR video traffic, and so on...).

It is interesting to note that most of the contributors used more or less the same statistical investigation methodology and came up with more or less the same conclusions. The authors affirm that the measured traffic is heavy tailed and long-range dependent that is to say $R(k) = L(k)k^{2H-2}, 0.5 < H < 1.0$, $R(k)$ denoting the autocorrelation of lag k of the series under study (inter-arrival times or block packet counts) and $L(x)$ being a slowly varying function at infinity i.e. $\forall x > 0, L(\tau x)/L(\tau) \overset{\tau \to +\infty}{\to} 1$. These findings are at variance with classical models of traffic such as the Poisson process or the Markov modulated Poisson process [2] for which dimensioning results have been established. It is known that quality measures such as the overflow probability or the delay are strongly underestimated by these traditional models if the real traffic has long-range dependence.

It is often difficult to find a physical interpretation for self-similarity. An explanation of the phenomenon that is significant for teletraffic applications has been proposed by Taqqu [3]; Taqqu proves that the aggregation of several ON/OFF renewal sources with Pareto distributed ON and/or OFF periods conducts asymptotically to a self-similar traffic. Other classical models displaying long-range dependence are the fractional Gaussian noise and the fractional (p, d, q) ARIMA process. More recently semi-Markovian models with an underlying Markov process on a countable infinite instead of finite space, enabling a richer probabilistic behavior for the underlying process have been proposed by Robert *et al.*[4] [5].

It is interesting to note that long-range dependence only specifies the autocovariance structure of the process for large lags, or alternatively, the behavior of the spectral density function at zero frequency. This is mainly because most of the works on long-range dependent processes have been devoted to Gaussian processes and more recently to linear processes: for these processes, second-order structure in some sense specifies all the dependence structure of the process. Very few is known on non-linear long-range dependent processes, at the exception of fractional bilinear and ARCH models.

As stressed above it is somewhat surprising that most of the investigators have adopted the methodology prescribed by Willinger *et al.*[1] without trying other possible routes. The most questionable part of their analyses is undoubtedly the fact that the traffic is supposed to be second-order stationary on these long time scales. Most of the analyses are done on hours of traffic. Based on these large time scale data, the usual method consists in 'proving' long-range dependence by computing several easily visualisable tests, at the very first place the variance-time plot, despite the fact that this index is known to be a poor estimator of the Hurst parameter from a statistical point of view. Heavy-tailness is proved by displaying the histogram of the data or by computing a goodness-of-fit measure, such as the Kolmogorov-Smirnov or the chi-square tests. Funilly, the critical levels of the tests are often computed as if the data were independent, though the first claim is that they are long-range dependent; this would of course largely modify the critical intervals and even the speed of convergence of the test statistics.

Our intuition is that the conclusions could have been different by adopting a perhaps more plausible assumption: the data are not stationary on a global scale but only locally i.e. stationary on a finer scale and non stationary on long time scales. It has been known ([6],[7],[8]) for years that deterministic jumps or trends in the mean of a time series without long-range dependence can mislead to the untrue conclusion that the series is long-range dependent if one relies on visual indexes of long-range dependence such as the variance time plot.

To support our intuition we analyze a common traffic trace. It was recorded at Bellcore and it was originally studied by Paxson and Floyd [9] who concluded that this stream exhibits a high degree of burstiness that can not be explained by any markovian model and who discussed how this burstiness might mesh with self-similar models of traffic. In Section 2 we test the stationarity of the measured traffic on different time scales ranging from a few seconds to hours of traffic. In Section 3 we introduce the Markov modulated Poisson process and a new model that we propose after careful analysis of different traffic data. In Section 4 we consider the estimation of the parameters of the proposed model: both block and recursive procedures of estimation are detailed. In Section 5 we prove by simulation that the abusive use of classical indexes of long-range dependence for non stationary series conducts to the false interpretation of long-range dependence.

2 Tests for stationarity

2.1 Principles

2.1.1 General Framework

Basically the tests of stationarity that we propose rely on the comparison of different empirical statistics calculated on two neighbour segments of finite length of the stream. The hypothesis of stationarity is rejected if the empirical statistics for the two neighbour segments are significantly different.

Denote by $\{X_t\}$ the sequence of the inter arrival times (IAT) from which a set of finite length $\{X_t\}_{1\leq t\leq T}$ is observed. Suppose that one aims at testing if this finite observation is strict sense stationary. Denote by τ_1 the presumed change point. For the sake of homogeneity we also define $\tau_0 = 0$ and $\tau_2 = T$. We do as if $\{X_t\}_{1\leq t\leq \tau_1}$ and $\{X_t\}_{\tau_1+1\leq t\leq T}$ were two realizations of finite length $T_i = \tau_i - \tau_{i-1}$ of two processes $\{X_t^1\}$ and $\{X_t^2\}$.

In what follows we test the stationarity of the IAT process in the sense of (i) the mean of the process (ii) the sampled cumulative distribution function $\mathbb{E}(g(X_t))$ where $g(x) = (\mathbb{I}_{\Delta_1}(x), \cdots, \mathbb{I}_{\Delta_N}(x))'$, $(\Delta_i)_{1\leq i\leq N}$ being a partition of $\mathbb{R}^+$ and (iii) the first covariance coefficients $\mathbb{E}((X_t^2, X_tX_{t+1}, \cdots, X_tX_{t+N-1})')$. We introduce a new time series $\{Z_t\}$ that is defined as (i) $Z_t = X_t$ (ii) $Z_t = g(X_t)$ or (iii) $Z_t = (X_t^2, X_tX_{t+1}, \cdots, X_tX_{t+N-1})'$ depending of the non stationarities that we want to detect.

2.1.2 A Central Limit Theorem

Different Assumptions are needed to establish a Central Limit Theorem for the vector of the empirical statistics.

Assumption 1 *$\{X_t\}$ is strict sense stationary.*

Assumption 2 $\frac{T_i}{T} \overset{T\to\infty}{\to} c_i > 0$.

Assumption 3 *$\{X_t\}$ is α-mixing with an α-mixing coefficient that verifies $\sum_{n=0}^{+\infty} \alpha_n^{\delta/(2+\delta)} < +\infty$ and with $\mathbb{E}(|X_t|^{2+\delta}) < +\infty$.*

Assumption 4 $q_t \overset{t\to\infty}{\to} +\infty$ *and* $q_t = o(t)$.

Let us recall that the α-mixing coefficient of the process $\{X_t\}$ is defined as $\alpha_n = \sup_{A,B} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$ the supremum being taken on all sets A in $\mathcal{M}_{-\infty}^t$ and B in $\mathcal{M}_{t+n}^{+\infty}$ where $\mathcal{M}_a^b = \sigma(X_t, a \le t \le b)$.

The Assumption 3 is verified by many usual processes and in particular by a large class of Markov processes. It is in particular verified by ARMA processes when the probability density function of the innovation is strictly positive on $\mathbb{R}$ and by any finite state irreducible Hidden Markov Chain.

Denote by $\hat{Z}_T^i = \frac{1}{T_i} \sum_{t=\tau_{i-1}+1}^{\tau_i} Z_t^i$ the empirical statistics for the segment of index i and denote by $\hat{Z}_T = ((\hat{Z}_T^1)'(\hat{Z}_T^2)')'$ the vector of the empirical statistics for the two segments.

Theorem 1 *Assume (A1-A2-A3). Then it holds that*

$$\sqrt{T}(\hat{Z}_T - (11)' \otimes \mathbb{E}(Z_t)) \sim \mathcal{AN}(0, \Gamma), \quad \Gamma = \begin{pmatrix} c_1^{-1}\Gamma_0 & 0 \\ 0 & c_2^{-1}\Gamma_0 \end{pmatrix}$$

where $\Gamma_0 = \sum_{\tau=-\infty}^{+\infty} \gamma_Z(\tau)$ with $\gamma_Z(\tau) = \mathbb{E}(Z_t Z_{t+\tau}') - \mathbb{E}(Z_t)\mathbb{E}(Z_t')$ and where $A \otimes B$ denotes the Kronecker product of A and B.

Note that Γ_0 is equal to the spectral density matrix of $\{Z_t\}$ at zero frequency. This remark permits the construction of a consistent estimator

$$\hat{\Gamma}_0 = \frac{1}{T} \sum_0^{+m_T} w(k) \Re((\sum_1^T (Y_t - \hat{\rho}) e^{j\frac{k+1}{T}t})'(\sum_1^T (Y_t - \hat{\rho}) e^{j\frac{k+1}{T}t}))$$

where $m_T = \sqrt{T}$ and $w(k) = \mathbb{I}_{k=0} + \frac{2}{2m_T+1}\mathbb{I}_{1\leq k\leq m_T}$.

The Demonstration of Theorem 1 can be found in Appendix 1.

2.1.3 Tests

- **Stationarity of the mean and of the marginal distribution**

  As stated above the test for stationarity consists in comparing $\hat{Z}_T^1$ and $\hat{Z}_T^2$. For the mean and for the marginal distribution of the process we consider the difference between $\hat{Z}_T^1$ and $\hat{Z}_T^2$ on the two neighbour segments $\hat{Z}_T^1 - \hat{Z}_T^2 = U\hat{Z}_T$, where $U = (1 - 1)$ in the test for stationarity of the mean of $\{X_t\}$ and $U = (1 - 1) \otimes I_N$ in the test for stationarity of the marginal distribution of $\{X_t\}$.

  Theorem 2 *Assume (A1-A2-A3). Then it holds that*

  $$\sqrt{T}U\hat{Z}_T \sim \mathcal{AN}(0, U\Gamma_0 U'), \quad T\hat{Z}_T(\Gamma^{-1/2})^H\Gamma^{-1/2}\hat{Z}_T \sim \chi^2(N)$$

  where $\Gamma^{\frac{1}{2}}$ *denotes a square root of* Γ, $\Gamma = \Gamma^{\frac{1}{2}}(\Gamma^{\frac{1}{2}})'$.

- **Stationarity of the first correlations**

  In [10] Mauchly introduces the sphericity statistics to test whether two gaussian random vectors have the same covariance matrix. Drouiche and Mokkadem ([11],[12]) generalize this test as a test for spectral adequacy. We propose to use this measure of similarity to test whether a process is second order stationary.

  For any positive sequence $\rho = (\rho_0, \rho_1, \cdots, \rho_{N-1})'$ denote by $T_N(\rho)$ the Toeplitz matrix $T_N(\rho) = \sum_{\tau=0}^{N-1} \rho_\tau M_\tau$ where M_τ is the matrix whose entry (i,j) is equal to $M_\tau(i,j) = \delta_\tau(|i-j|)$. For any two positive sequences μ and ν the sphericity is defined as the ratio

  $$S(\mu,\nu) = \frac{(det(T_N(\mu)T_N^{-1}(\nu)))^{1/N}}{\frac{1}{N}Tr(T_N(\mu)T_N^{-1}(\nu))}$$

Our idea is to derive the asymptotic distribution of $S(\hat{Z}_T^1, \hat{Z}_T^2)$ normalized by a factor that depends on the length T of the observation and to reject the hypothesis of stationarity if the obtained value is lower than a prescribed threshold determined by the false alarm probability α.

Theorem 3 *Assume (A1-A2-A3). Then it holds that*

$$2TS^*(\hat{Z}_T) \xrightarrow[(d)]{T\to\infty} Z^H \nabla^2 S((1,1)' \otimes \mathbb{E}(Z_t))Z, \quad Z \sim \mathcal{N}(0,\Gamma)$$

where $\nabla^2 S^*((1,1)' \otimes \mathbb{E}(Z_t))$ *denotes the Hessian of* S *at point* $(1,1)' \otimes \mathbb{E}(Z_t)$.

The demonstration of this result is based on a Taylor development of S at point $(\mathbb{E}(Z_t), \mathbb{E}(Z_t))$. As S is maximum at point $(\mathbb{E}(Z_t), \mathbb{E}(Z_t))$ a second order Taylor development is needed.

To establish the expression of $\nabla^2 S$ one needs the second order differential of $M \to \log|M|$ and of $M \to M^{-1}$:

$$(dA, dB) \to -Tr(M^{-1}dAM^{-1}dB)$$
$$(dA, dB) \to 2M^{-1}dAM^{-1}dBM^{-1}$$

- **Thresholds**

  Theorems 2 and 3 permit to reject the set of Assumptions (A1-A3) with a false alarm probability of α. If the obtained statistics is superior to the $(1-\alpha)$ quantile of the asymptotic distribution one concludes that (A1-A3) is wrong which means that A1 and A3 are mutually exclusive.

  Note that in Theorem 3 the asymptotic distribution is a quadratic form in a multidimensional Gaussian random variable and that the prescribed threshold is obtained by Monte-Carlo simulation.

2.2 Results

The simulations are replicated for thirteen time-scales ranging from six seconds to one hour and thirty minutes and for ten pairs of neighbour segments for each time-scale.

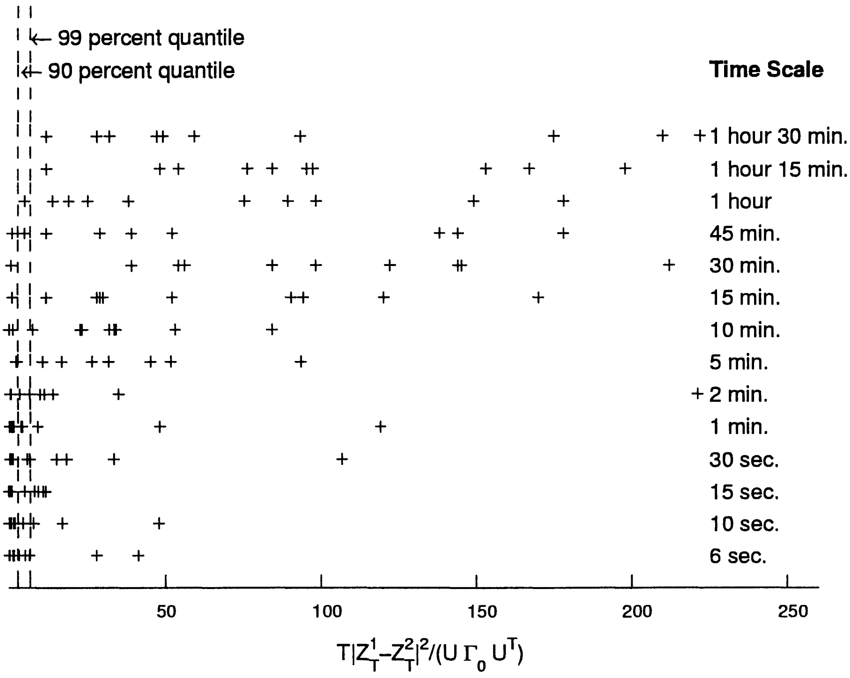

Figure 1: Stationarity of the mean

On the Figures 1 and 2 we plot $T\hat{Z}_T^H(\Gamma^{-1/2})^H\Gamma^{-1/2}\hat{Z}_T$ for all the pairs of neighbour segments. The 90% and 99% quantiles of $\chi^2(N)$ are represented in dotted lines. (A1-A3) is rejected when $T\hat{Z}_T^H(\Gamma^{-1/2})^H\Gamma^{-1/2}\hat{Z}_T$ is superior to the $(1-\alpha)$ quantile of $\chi^2(N)$.

On the Figure 3 we plot the cumulative distribution function $P(X \leq 2TS(\hat{Z}_T^1, \hat{Z}_T^2))$ for the asymptotic distribution
$X = Z^H \nabla^2 S \mid_{(\mathbb{E}(Z_t),\mathbb{E}(Z_t))} Z$ where $Z \sim \mathcal{N}(0,\Gamma)$. The 90% and 99% fractiles for the distribution of X are represented in dotted lines. (A1-A3) is rejected if $P(X \leq 2TS(\hat{Z}_T^1, \hat{Z}_T^2))$ is superior to $(1-\alpha)$.

The conclusions of our simulations is that (A1-A3) is wrong for most pairs of neighbour segments for long time-scales. Classical models such as the stationary Poisson process or the stationary Markov Modulated Poisson Process are consequently not adapted to the traffic that we investigate on these long time-scales.

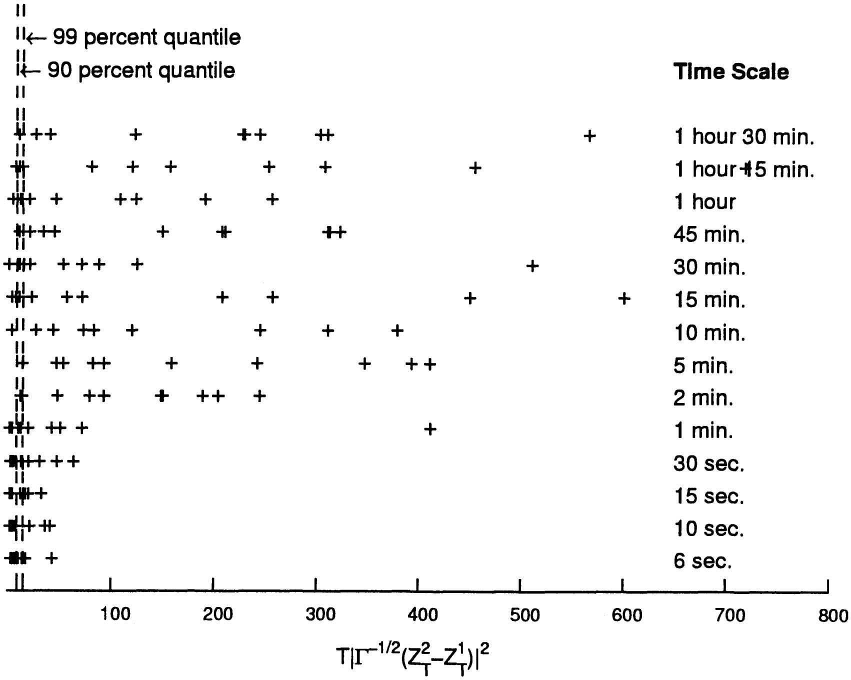

Figure 2: Stationarity of the marginal distribution

Note that the Assumption A3 is wrong for long-range dependent processes such as the Fractional Gaussian Noise or the fractionally integrated autoregressive moving average process. Consequently the tests developed do not permit to reject A1 for long-range dependent processes. The difficulty to decide between long-range dependence and non stationarities has already been discussed by Duffield *et al.*in [7].

It is thus difficult to decide if the evidences of auto-similarity mentioned by many authors result from a real auto-similarity of the traffic or from some non-stationarities that might have mislead to the conclusion of auto-similarity ([6],[7],[8]) or from the coexistence of both phenomena.

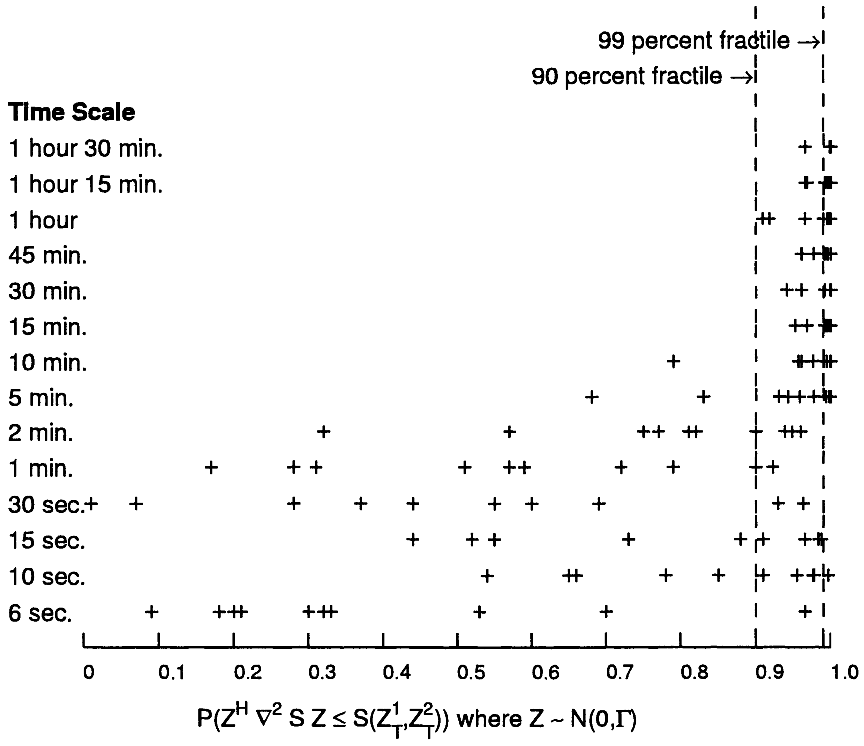

Figure 3: Stationarity of the first five correlations

3 Locally stationary and semi-Markov models

3.1 Vindication of the locally stationary and semi-Markov models

As mentioned in Section 2 it is impossible to decide between the hypothesis of self-similarity and the hypothesis of local stationarity on the simple basis of the tests of stationarity that we have developed. In what follows we favor the hypothesis of local stationarity by comparison with the hypothesis of long-range dependence.

As one can see on Figure 4 the measured traffic is not significantly correlated if one considers different slides of traffic that last approximately ten seconds. This ten seconds time scale corresponds to the shortest period of reallication of the bandwith at the User Network Interface that one can afford; for shorter periods of reallocation the control automata would be overloaded [13]. A mixture of exponentials or of shifted exponentials seems to be a reasonable model for the marginal distribution of the inter-arrival times on this ten seconds time-scale. We think that the classical Markov modulated Poisson process (MMPP;[2]) or a new model that we call the shifted exponential hidden Markov model (SEHMM) are both reasonable and tractable models of traffic on short time scales.

A solution to model time series that are suspected to be locally stationary consists in proposing a parametric model whose parameters are varying slowly with time. On longer time scales we propose to model the measured traffic as a semi-Markov process whose parameters are varying slowly with time. This solution has many advantages among which the existence of recursive procedures of estimation of the parameters of the process and the existence of queuing results for such input processes.

3.2 The Markov modulated Poisson process

The Markov modulated Poisson (MMPP) process is a generalization of the classical Poisson process that is commonly used as a model for teletraffic data. This model permits to a certain extent to take the observed burstiness of the traffic into account while keeping in a markovian framework. For a review about the Markov modulated Poisson process we refer the reader to the tutorial paper of Meier-hellstern [2].

3.2.1 Representation of the MMPP as a DSPP

The doubly stochastic Poisson process (DSPP) is a generalization of the classical Poisson process. In the case of the doubly stochastic Poisson process the intensity of the Poisson process is a non negative stochastic process. Denote by $(\lambda(t))$ a continuous time non negative stochastic process, denote by (T_i) the successive instants of arrival and denote by $N(t)$ the associated count process $N(t) == \sum_{i=0}^{\infty} \mathbb{I}(T_i \leq t)$. We suppose that $T_0 = 0$ and $N(0) = 0$. The doubly stochastic Poisson process is defined by

$$\begin{aligned} &(\mathbf{i})\forall 0 \leq t1 \leq t2 \leq t3, \forall (p,q) \in \mathbb{N}^2, \\ &\mathbb{P}(N_{t_3} - N_{t_2} = p, N_{t_2} - N_{t_1} = q \mid \mathcal{F}_0) \\ &= \mathbb{P}(N_{t_3} - N_{t_2} = p \mid \mathcal{F}_0)\mathbb{P}(N_{t_2} - N_{t_1} = q \mid \mathcal{F}_0) \end{aligned} \tag{1}$$

where $\mathcal{F}_\infty = \sigma(\lambda_t, t \geq 0)$ is the complete filtration associated to $(\lambda_t)_{t\in\mathbb{R}}$.

$$\begin{aligned} &(\mathbf{ii})\forall t > s, \forall n \in \mathbb{N}, \mathbb{P}(N_t - N_s = n \mid \mathcal{F}_\infty, \mathcal{G}_s) \\ &= \exp(-\int_s^t \lambda_u \, du)\frac{(\int_s^t \lambda_u \, du)^n}{n!} \end{aligned} \tag{2}$$

where $\mathcal{G}_s = \sigma(N_t; t \leq s)$.

The Markov modulated Poisson process is the particular case of the doubly stochastic Poisson process when the intensity process is a finite-state and continuous time Markov process

$$\forall t > s, \mathbb{P}(\lambda_t = \lambda_i \mid \mathcal{F}_s) = \mathbb{P}(\lambda_t = \lambda_i \mid \lambda_s) \tag{3}$$

where $\mathcal{F}_s = \sigma(\lambda_t; t \leq s)$.

$(\lambda_1, \lambda_2, \cdots, \lambda_K)$ is the set of states of the intensity process and K is the order of the Markov modulated Poisson process. The most common parameterization of the Markov modulated Poisson process is (Q, Λ) where Q is the stochastic transition matrix of the finite state continuous time Markov intensity process and where Λ is the diagonal matrix whose entry (i, i) is equal to λ_i. The Markov modulated Poisson process is then referred to as the (Q, Λ) source.

3.2.2 Representation of the MMPP as a MRP

A Markov renewal process (MRP) is a discrete time stochastic process (X_n, Y_n) such that **(i)** $(X_n)_{n\in\mathbb{N}}$ is a finite state Markov chain.

$$\textbf{(ii)}\mathbb{P}(Y_n \leq t, X_n = j \mid \mathcal{G}_{n-1}) = \mathbb{P}(Y_n \leq t, X_n = j \mid X_{n-1}) \qquad (4)$$

where $\mathcal{G}_n = \sigma((X_m, Y_m); m \leq n)$ is the filtration associated to (X_n, Y_n).

The most common parameterization of the Markov renewal process is its stochastic transition matrix $F(t) = (F_{ij}(t))_{1\leq i,j\leq K}$ with $F_{ij}(t) = \mathbb{P}(X_{n+1} = j, Y_{n+1} \leq t \mid X_n = i)$. The transition matrix of the imbedded Markov chain $(X_n)_{n\in\mathbb{N}}$ is then $P = F(+\infty)$.

Denote by $U_i = T_i - T_{i-1}$ the successive inter-arrival times and denote by $\bar{\lambda}_i = \lambda_{T_i}$ the state of the underlying Markov intensity process at time T_i. We demonstrate in Appendix 2 that $(\bar{\lambda}_i, U_i)$ is a Markov renewal process with Markov transition matrix $F(t) = \int_0^t \exp((Q - \Lambda)u)\, du\Lambda$.

3.2.3 Representation of the MMPP as a HMM

A hidden Markov model is a stochastic process (X_n, Y_n) such that
(i) $(X_n)_{n\in\mathbb{N}}$ is a finite state Markov chain.

$$\textbf{(ii)}\mathbb{P}(Y_n \leq t \mid X_n = i, \mathcal{G}_{n-1}) = \mathbb{P}(Y_n \leq t \mid X_n = i) \qquad (5)$$

where $\mathcal{G}_n = \sigma((X_m, Y_m); m \leq n)$ is the filtration associated to (X_n, Y_n).

The terminology "hidden" stems from the fact that the imbedded Markov chain $(X_n)_{n\in\mathbb{N}}$ is not observed directly and that any statistical treatment should be based on the observation of $(Y_n)_{n\in\mathbb{N}}$ only.

The hidden Markov model is parameterized by the transition matrix of the imbedded Markov chain $P = (P_{ij})_{1\leq i,j\leq K}$ where $P_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$ and by the either discrete or continuous distribution of Y_n conditionally to $X_n = i$.

Remark that any Markov renewal process (X_n, Y_n) of order K can be represented as a hidden Markov model of order K^2. Denote by (X_n, Y_n) a Markov renewal process with Markov renewal matrix $F(t)$. Denote by $\bar{X}_n = (X_{n-1}, X_n)$ the successive transitions of the underlying Markov chain. Then $(\bar{X}_n, Y_n)$ is a hidden Markov chain: the transition matrix P of the imbedded Markov chain $(\bar{X}_n)_{n\in\mathbb{N}}$ is a structured transition matrix with entries equal to

$$P_{(i_1,i_2)(j_1,j_2)} = \mathbb{1}_{i_2=j_1} F_{j_1 j_2}(+\infty)$$

and the cumulative ditribution function of Y_n conditionally to $\bar{X}_n = (i,j)$ is $F_{ij}(t)$.

3.3 The shifted exponential HMM

In this Section we briefly expose a new model that we have developed after carefully analyzing different measures of traffic on broadband networks [9] [13].

On Figure 4 we display the spectral density function of the inter-arrival times and the histogram of the natural logarithm of the inter-arrival times for five distant frames corresponding roughly to ten seconds of traffic. The spectral density function is estimated using either standard smoothed periodogram techniques or the Burg maximum entropy method [14, 15, 16, 17]. The logarithm transformation permits to visualize the distribution of the inter-arrival times in the region of short inter-arrivals.

From a simple visual inspection of this plot it is seen that the spectral density function of the process evolves with time. It is noteworthy on the non parametric estimate of the sdf that the spectral density in the neighbourhood of zero frequency seems compatible with short-range dependence. Note that the maximum entropy estimate of the sdf cannot reveal long-range dependence: the maximum entropy spectral estimation amounts to fit to the p first autocovariance coefficients of an autoregressive (AR) model and to

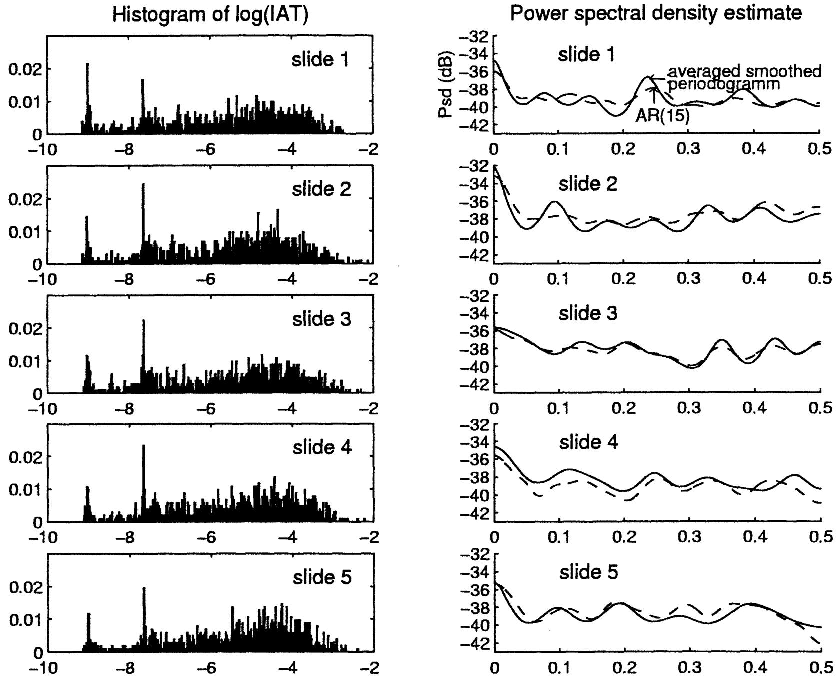

Figure 4: Histogram and power spectral density of five distant slides of traffic lasting ten seconds

extrapolate the autocovariance coefficients after time index p as if the autocovariance sequence were that of an order p AR model. Thus the extrapolated autocovariance coefficients decrease as an exponential for large lags. It is also interesting to note that the difference between the maximum and the minimum of the spectral density function is somewhat reduced, yet data are not compatible with a Poisson process since the inter-arrival times are significantly correlated.

The analysis of the histogram of the logarithm of the inter-arrival times and of its evolution also reveals interesting features. It seems that a mixture of exponential distributions with different offsets. Our model of marginal distribution is close to the model proposed by Kofman *et al.*[18]. According to these authors who analyzed the traffic measured by Jain and Routhier [19] the distribution of the inter-arrival times is a mixture of exponentials.

On the basis of these findings we propose to model the measured traffic as a hidden Markov model (X_n, U_n), the distribution of U_n conditionally to $X_n = i$ being an exponential distribution with an offset time s_i

$$\begin{array}{l} \mathbb{P}(X_t = i \mid \mathcal{F}_{t-1}, \mathcal{G}_{t-1}) = P(X_t = i \mid X_{t-1}) \\ \mathbb{P}(U_t \in A \mid X_t = i, \mathcal{F}_{t-1}, \mathcal{G}_{t-1}) = \int_A \mathbb{I}_{[s_i, +\infty[}(u) \lambda_i e^{-\lambda_i (u - s_i)} du \end{array} \tag{6}$$

where $\mathcal{F}_t = \sigma(U_s; s \leq t)$ and by $\mathcal{G}_t = \sigma(X_s; s \leq t)$.

4 Estimation of the parameters of hidden Markov models

In the previous section we have proposed different semi-Markov processes to model the traffic measured on broadband networks. One of these models (SEHMM) is a hidden Markov model. The second model (MMPP) can be represented as a hidden Markov model with structured stochastic transition matrix as we demonstrate in the previous section. In this Section we detail both block and recursive procedures of estimation of the parameters of a hidden Markov model. The block procedure is used on the time scales on which the measured traffic can be considered as stationary; the recursive procedures of estimation can be used to track the slow variations of the measured traffic on longer time scales.

Remark that renewal processes such as the simple Poisson process are particular cases of hidden Markov models when the order of the model is $K = 1$; the procedures of estimation that we develop in this Section apply to renewal processes. We verify that the offsets $(s_i)_{1 \leq i \leq K}$ do not vary significantly with time and we consequently do not consider the problem of the estimation of the offsets $(s_i)_{1 \leq i \leq K}$; in what follows we assign the offsets $(s_i)_{1 \leq i \leq K}$ a constant value.

4.1 Unconstrained parameterization

One should remark that the parameters of the Shifted Exponential Hidden Markov Model are subject to the following constraints

$$
\begin{array}{rll}
 & \forall 1 \leq i,j \leq K & P_{ij} \geq 0 \\
 & \forall 1 \leq i \leq K & \sum_{j=1}^{K} P_{ij} = 1 \\
\text{and} & \forall 1 \leq i \leq K & \lambda_i > 0
\end{array} \tag{7}
$$

These constraints must be taken into account in the optimization steps of the procedures of estimation that we expose in what follows. Another solution consists in proposing an unconstrained parameterization of the model and in using standard optimization techniques for the problem without constraints. The second solution is retained in what follows.

To overcome the constraints of positivity $\lambda_i > 0, 1 \leq i \leq K$ one can substitue λ_i for $\Phi(\lambda_i)$ where Φ is a bijection from $\mathbb{R}^+$ into $\mathbb{R}$ e.g. $\lambda_i \rightarrow \log \lambda_i$.

The constraints concerning $P_{i,:} = (P_{i1}, \cdots, P_{iK})$ are equivalent to setting $P_{i,j} = (u^i_j)^2$ where $u^i = (u^i_j)_{1 \leq j \leq K}$ is a unitary vector of $\mathbb{R}^K$ that is to say verifies $\sum_{j=1}^{K} (u^i_j)^2 = 1$. This remark permits to parameterize $P_{i,:}$ by $(K-1)$ Givens angles $(\alpha^i_j)_{1 \leq j \leq K-1}$.

Proposition 1 *Any unitary vector u in $\mathbb{R}^K$ is the image vector of e_K where $e = (e_1, \cdots, e_K)$ is the canonical basis of $\mathbb{R}^K$ by a product of $(K-1)$ Givens rotations in $\mathbb{R}^K$:*

$$
u = \prod_{k=1}^{K-1} G_{k,K-1}(\alpha_k)[0 \cdots 01]^T,
$$

$0 \leq \gamma_k < 2\pi$ *and*

$$
G_{k,K}(\alpha_k) = \begin{pmatrix}
I_{K-k-1} & & & \\
 & \cos \alpha_k & & \sin \alpha_k \\
 & & I_{k-1} & \\
 & -\sin \alpha_k & & \cos \alpha_k
\end{pmatrix}
$$

represents the matrix of rotation of angle α_k in the subspace (e_k, e_K) of $\mathbb{R}^K$.

4.2 Off line estimation of the parameters

In this Section we briefly recall the basic principles of the Expectation Maximization (EM) algorithm [20]. The EM algorithm is a constructive method of the maximum likelihood estimate of the parameters of a mixture or of a finite state Hidden Markov Model. Denote by $\{(Y_t, X_t)\}_{1\leq t\leq T}$ a finite length realization of a hidden Markov chain. The basic idea of the EM algorithm consists in constructing a sequence $(\hat{\theta}_k)_{k\in\mathbb{N}}$ that verifies $Q(\hat{\theta}_{k+1}, \hat{\theta}_k) \geq Q(\hat{\theta}_k, \hat{\theta}_k)$ when $Q(\theta_1, \theta_2) = \mathbb{E}(L(X_{1:T}, Y_{1:T}; \theta_1) \mid X_{1:T} = x_{1:T}, \hat{\theta}_2)$. A consequence of Jensen's inequality [21] is that the likelihood of the observed $\{X_t\}_{1\leq t\leq T}$ is an increasing sequence

$$L(x_{1:T}; \hat{\theta}_{k+1}) \geq L(x_{1:T}; \hat{\theta}_k)$$

and thus that it converges on a local maximum of the log-likelihood $\theta \to L(x_{1:T}; \theta)$ since $L(x_{1:T}; \theta)$ is always inferior to zero.

Each iteration of the EM algorithm can be decomposed into two steps.

- **Expectation Step:**

  This step permits to get the distribution of the imbedded Markov chain $Y_{1:T}$ conditionally to the observed $X_{1:T} = x_{1:T}$ with an algorithm of linear complexity. It is based on the recursive construction of two lattices

$$\begin{aligned}
\alpha_t(i) &= \mathbb{P}(X_{1:t}, Y_t = i \mid X_{1:T} = x_{1:T}; \hat{\theta}_k) \\
\beta_t(i) &= \mathbb{P}(X_{t+1:T} = x_{t+1:T}) \mid X_{1:T} = x_{1:T}; \hat{\theta}_k) \\
\alpha_{t+1}(i) &= \textstyle\sum_j \alpha_t(j) P_{ji} p_i(x_{t+1}) / \sum_{i,j} P_{ji} \alpha_t(j) p_i(x_{t+1}) \\
\beta_{t-1}(i) &= \textstyle\sum_j (x_t) P_{ij} \beta_t(j) / \sum_{i,j} P_{ij} \beta_t(j) p_j(x_t)
\end{aligned} \tag{8}$$

  $p_i(\bullet)$ denotes the probability density function of X_t conditionally to $Y_t = i$.

  Straightforward manipulations conduct to

$$\begin{aligned}
\gamma_t(i) &= \mathbb{P}(X_{t+1:T} = x_{t+1:T} \mid Y_t = i; \hat{\theta}_k) = \alpha_t(i)\beta_t(i) \\
\zeta_t(i) &= \mathbb{P}(Y_t = i, Y_{t+1} = j \mid X_{1:T} = x_{1:T}; \hat{\theta}_k) \\
&= \frac{\alpha_t(i) P_{ij} p_j(x_{t+1}) \beta_{t+1}(j)}{\sum_{i,j} \alpha_t(i) P_{ij} p_j(x_{t+1}) \beta_{t+1}(j)}
\end{aligned} \tag{9}$$

- **Maximization Step:**

  This step consists in maximizing $Q(\theta, \hat{\theta}_{k+1})$ with respect to θ. Note that $Q(\theta, \hat{\theta}_k)$ can be decomposed into two terms $Q(\theta, \hat{\theta}_k) = Q_1(\theta, \hat{\theta}_k) + Q_2(\theta, \hat{\theta}_k)$ where

$$\begin{array}{rcl} Q_1(\theta, \hat{\theta}_k) & = & \mathbb{E}(L(Y;\theta) \mid X = x; \hat{\theta}_k) = \sum_i \log p_i \alpha_1(i) \\ & & + \sum_{i,j} \log P_{ij} \sum_{t=2}^{T} \zeta_t(i,j) \\ Q_2(\theta, \hat{\theta}_k) & = & \mathbb{E}(L(X \mid Y;\theta) \mid X = x; \hat{\theta}_k) \\ & = & \sum_i \gamma_t(i)(\log \lambda_i - \lambda_i(x_t - s_i)) \end{array} \tag{10}$$

  depend separately from $(P_{ij})_{1 \leq i,j \leq K}$ and from $(\lambda_i)_{1 \leq i \leq K}$. $Q(\theta, \hat{\theta}_k)$ can thus be maximized separately with respect to $(P_{ij})_{1 \leq i,j \leq K}$ and with respect to $(\lambda_i)_{1 \leq i \leq K}$:

$$\begin{array}{rcl} \hat{P}_{ij}^{(k+1)} & = & \frac{\sum_{t=2}^{T} \zeta_t(i,j)}{\sum_j \sum_t \zeta_t(i,j)} \\ \hat{\lambda}_i^{(k+1)} & = & \frac{\sum_t \gamma_t(i)}{\sum_t \gamma_t(i)(x_t - s_i)} \end{array} \tag{11}$$

  Note that for these formulae of actualisation the constraints (7) are automatically verified and though that one can use the most common parameterization $\theta = ((P_{ij})_{1 \leq i,j \leq K}, (\lambda_i)_{1 \leq i \leq K})$.

Leroux [22] proved the consistency of the maximum likelihood estimator under mild conditions in the case of hidden Markov models and later Bickel and Ritov [23] proved the local asymptotic normality of this estimator.

4.3 On line estimation of the parameters

Procedures of recursive estimation of hidden Markov models parameters have been considered by Holst and Lindgren [24], Krishnamurty and Moore [25], Ryden [26] and Mevel [27]. In the papers of Holst *et al.*and of Lindgren *et al.*no convergence results are established although simulation studies show that their algorithms often work out well in practice. Ryden and Mevel prove that their recursive estimator converges to the set of minima of the Kullback-Leibler divergence and is asymptotically normal under suitable conditions. In this Section we expose the procedure of Mevel which is used in the sequel to follow the slow variations of the parameters of the non stationary hidden Markov model.

Note that the log-likelihood can be decomposed into

$$L(x_{1:T};\theta) = \sum_{t=1}^{T} \log p(x_t \mid x_{1:t-1};\theta) = \sum_{t=1}^{T} \log \sum_{i=1}^{K} b_i(t;\theta) p_i(x_t;\theta) \tag{12}$$

where $p_i(\bullet;\theta)$ denotes the probability density function of X_t conditionally to $Y_t = i$ and where

$$b_i(t;\theta) = \mathbb{P}(Y_t = i \mid X_{1:t-1} = x_{1:t-1};\theta) \tag{13}$$

denotes the one-step ahead prediction filter.

Mevel proposes an algorithm of linear complexity to approximate the one-step ahead prediciton filter as well as its gradient w.r.t. θ

$$b_i(t+1;\theta) = \frac{\sum_j P_{ji}(\theta) p_j(x_t;\theta) b_j(t;\theta)}{\sum_{ij} P_{ji}(\theta) p_j(x_t;\theta) b_j(t;\theta)} = F(b_\bullet(t;\theta), x_t) \tag{14}$$

and similarly

$$\frac{\partial}{\partial\theta} b_i(t+1;\theta) = G(b_\bullet(t;\theta), \frac{\partial}{\partial\theta} b_\bullet(t;\theta), x_t) \tag{15}$$

Mevel proves that one can estimate the true parameter θ_{tr} by looking for an at least local minimum of the Kullback-Leibler divergence $K(\theta) \overset{p.s.;\theta_{tr}}{=} \lim_{n\to+\infty} \frac{1}{n} \log \frac{p_n(X_{1:n};\theta_{tr})}{p_n(X_{1:n};\theta)}$ with a recursive estimator

$$\hat\theta_t = P_G(\hat\theta_{t-1} + \gamma_t \frac{\partial}{\partial\theta} \log p(x_t \mid x_{1:t-1};\theta_t)) \tag{16}$$

where G is a closed, bounded and convex subset of the parameters space and P_G is a projection on this convex set and where

$\gamma_t = \gamma_0 t^{-\alpha}$ where $\frac{1}{2} < \alpha \leq 1$. Note that in the non stationary context of tracking of the slowly varying parameters then $\gamma_t = \gamma$ and none of the convergence results mentioned here above holds.

The lattice procedure of estimation of the one-step ahead prediction filter and of its gradient permits to estimate

$$\begin{array}{l}\frac{\partial}{\partial\theta}\log p(x_t \mid x_{1:t-1};\theta) = (\log \sum_{i=1}^{K} b_i(t;\theta)p_i(x_t;\theta))^{-1} \\ (\sum_{i=1}^{K} b_i(t;\theta)\frac{\partial p_i(x_t;\theta)}{\partial\theta} + \sum_{i=1}^{K} p_i(x_t;\theta)\frac{\partial b_i(t;\theta)}{\partial\theta})\end{array} \tag{17}$$

Another solution consists in replacing the most common constrained parameters with the unconstrained parameters $\theta = ((\alpha_j^i)_{1\leq i\leq K, 1\leq j\leq K-1}, (\log\lambda_i)_{1\leq i\leq K})$ as exposed in Section 4.1 so that no projection is needed.

5 Indexes for self-similarity

5.1 Variance time plot

Denote by $\{X_t\}$ a series with long-range dependence i.e.

$$R(k) = L(k)k^{2H-2} \quad 0.5 < H < 1.0 \tag{18}$$

where $R(k) = \mathbb{E}((X_t - \bar{X})(X_{t+k} - \bar{X}))$ denotes the autocorrelation of lag k of $\{X_t\}$ and where $L(x)$ is a slowly varying function at infinity that is to say $L(\tau x)/L(\tau) \overset{\tau\to+\infty}{\to} 1$ for all $x > 0$.

It is well known [28] that if $\{X_t^{(m)}\}$ denotes the aggregated series $X_t^{(m)} = \frac{1}{m}\sum_{k=(t-1)m+1}^{tm} X_k$ the sample variance $\mathrm{var}X^{(m)}$ of the aggregated series is $\mathrm{var}X^{(m)} \sim \sigma_0^2 m^{2H-2}$ as $m \to +\infty$. This permits the construction of a visual index of long-range dependence. One plots $\log \mathrm{var}X^{(m)}$ versus $\log m$ for various aggregation levels. If the series is long-range dependent the graphic fits a straight line with a slope $-1 < \beta = 2H - 2 < 0$. The slope of this straight line provides an estimate of the Hurst parameter H.

5.2 R/S analysis

Another classical result is that a series with long-range dependence satisfies to the Hurst effect. Consider $X_{1:n}$ a finite set of

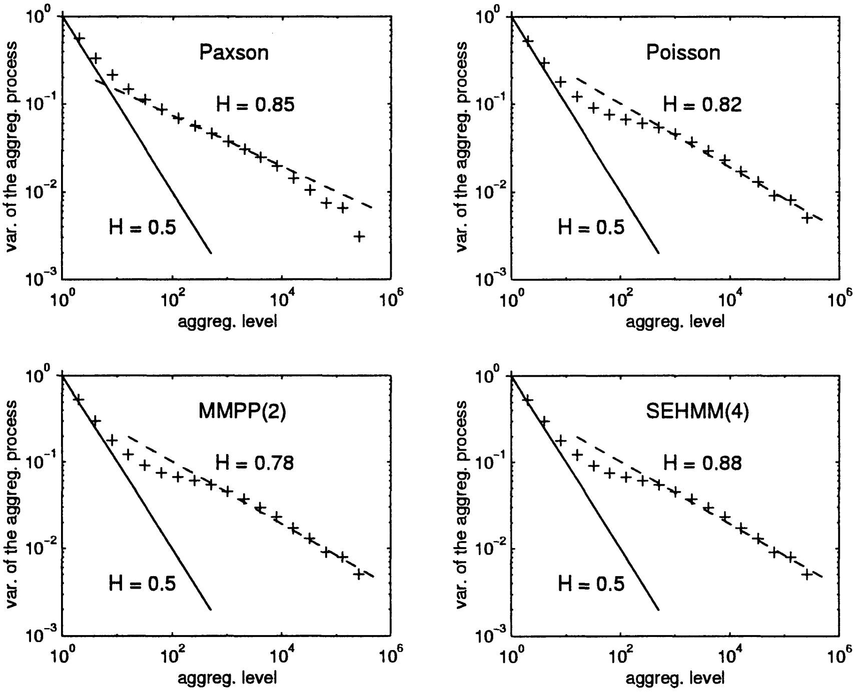

Figure 5: Variance-time plot

observations of length n. Denote by $\bar{X} = \frac{1/n}{\sum}_{1\leq k\leq n} X_k$ and $S = \frac{1}{n}\sum_{1\leq k\leq n}(X_k - \bar{X})^2$ the sample mean and the sample variance of $X_1 : n$; then the range of $X_{1:n}$ is defined as $R = \max(0, W_1, \cdots, W_n) - \min(0, W_1, \cdots, W_n)$ where $W_t = \sum_{1\leq k\leq t}(X_t - \bar{X}), 1 \leq t \leq n$.

The series is said to be self-similar if the rescaled adjusted range $\mathbb{E}((R/S)(n))$ is

$$\mathbb{E}(\tfrac{R}{S}(n)) \overset{n\to+\infty}{\sim} cn^H \quad c > 0, 0.5 < H < 1.0 \tag{19}$$

This phenomenon is known as the Hurst effect. Broadly speaking this result means that the integrated process visits more space when the increment process has long-range dependence than when it does not have any long-range dependence. This phenomenon was obsedrved by Hurst [29] in 1951 in a study on the Nile floods. This result permits the construction of a second visual index of

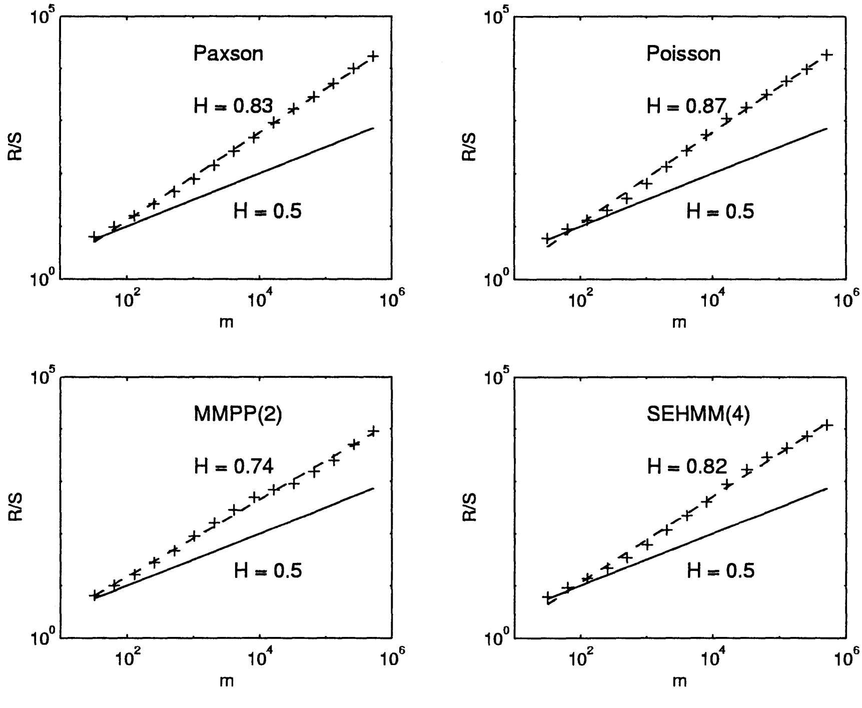

Figure 6: R/S analysis

long-range dependence. One plots $\log \mathbb{E}(R/S)$ versus $\log n$ and the series is self-similar if the graphic fits a straight line with slope $0.5 < H < 1.0$. The slope of this straight line provides an estimate of the Hurst parameter H.

Other procedures of estimation of the Hurst parameter are periodogramm regression techniques [30] and Whittle's maximum likelihood estimate.

5.3 Results

Our intuition is that the apparent self-similarity that many authors observe may be due to some non stationarities in a time series without long-range dependence. To support our intuition we simulate different non stationary and semi-Markov traffics whose parameters are matching the estimates of the parameters of the real traffic and we visualize the variance time plot and the R/S indexes for self-similarity of the traffic that we simulate. Three different non stationary models are proposed: the non stationary Poisson process, the non stationary Markov modulated Poisson process of order 2 (MMPP(2)) and the non stationary shifted Exponential hidden Markov model of order 4 (SEHMM(4)).

We display on Figures 5 and 6 the results of the analyses for the real traffic [9] and for the three synthetic locally stationary synthetic traffic without self-similarity. As one can see on Figures 5 and 6 if one relied on the simple visual indexes for self-similarity one would conclude that the synthetic traffics with no self-similarity but with some non stationarities are self-similar. The estimates of the Hurst parameter that one deduces from the variance time analysis and from the R/S analysis are close if one takes into account the bad quality of these estimators.

6 Conclusion

In this contribution we have demonstrated that the apparent self-similarity of the traffic measured on broadband networks might stem from some undetected non stationarities in a traffic without self-similarity rather than from a real self-similarity of the trafffic.This is a major finding since that means that one can take advantage of the numerous results that have been obtained for years for semi Markov traffics. On the contrary very few is known if the real traffic happens to be effectively self-similar. In the future we plan to support our findings by intensive simulation on other measures of Ethernet traffic and of high bit rate video data.

A Demonstration of Theorem 1

The demonstration of Theorem 1 requires to establish the following Central Limit Theorem

Lemma 1 *Assume (A1-A2-A3). Then it holds that*

$$\sqrt{T}\left(\frac{1}{T_i}\sum_{t=\tau_{i-1}+1}^{\tau_i} Z_t - \mathbb{E}(Z_t)\right) \sim \mathcal{AN}(0, c_i^{-1}\Gamma_0), \quad 1 \leq i \leq 2$$

where $\Gamma_0 = \sum_{\tau=-\infty}^{+\infty} \gamma_Z(\tau)$ *with* $\gamma_Z(\tau) = \mathbb{E}(Z_t Z'_{t+\tau}) - \mathbb{E}(Z_t)\mathbb{E}(Z'_t)$

The demonstration of Lemma 1 can be found in for example [21].

To prove Theorem 1 we mimic the approach of Epps [31]. We define a new estimator $\tilde{Z}_T^i$ where the first q_T terms are removed

$$\tilde{Z}_T^i = \frac{1}{T_i}\sum_{\tau_{i-1}+q_T+1}^{\tau_i} Z_t.$$

The basic idea consists in proving that $\sqrt{T}\hat{Z}_T^i$ and $\sqrt{T}\tilde{Z}_T^i$ converge in distribution to the same normal distribution and in proving that $\sqrt{T}\tilde{Z}_T^1$ and $\sqrt{T}\tilde{Z}_T^2$ are asymptotically independent, in the sense that

$$T(\mathbb{E}(e^{i\tilde{Z}_T^1 u^H + i\tilde{Z}_T^2 v^H}) - \mathbb{E}(e^{i\tilde{Z}_T^1 u^H})\mathbb{E}(e^{i\tilde{Z}_T^2 v^H})) \stackrel{T\to\infty}{\to} 0$$

It results from the Davydov Theorem [21] that

$$|\mathbb{E}(e^{i(\tilde{Z}_T^1 u^H + \tilde{Z}_T^2 v^H)}) - \mathbb{E}(e^{i\tilde{Z}_T^1 u^H})\mathbb{E}(e^{i\tilde{Z}_T^2 v^H})| \leq \alpha(q_T) \stackrel{T\to+\infty}{\to} 0$$

and though $\tilde{Z}_T^1$ and $\tilde{Z}_T^2$ are asymptotically independent.

Denote by $D_T^i = \sqrt{T}(\hat{Z}_T^i - \tilde{Z}_T^i) = \frac{\sqrt{T}}{T_i} \sum_{t=\tau_{i-1}+1}^{\tau_{i-1}+q_T} Z_t$ the difference between $\sqrt{T}(\hat{Z}_T^i$ and $\sqrt{T}\tilde{Z}_T^i$. It results from (A2-A4) that the covariance matrix of D_T^i tends to zero as T tends to infinity. It then results from Lemma 1 and from the Slutski Theorem [17] that $\sqrt{T}(\tilde{Z}_T - (1,1)' \otimes \mathbb{E}(Z_T)) \sim \mathcal{AN}(0, \Gamma)$ which concludes the proof of Theorem 1.

B Representation of the (Q, Λ) source as a Markov renewal process

Consider the (Q, Λ) source introduced in Section 3.2. We are going to demonstrate that this (Q, Λ) source can be represented as a discrete time Markov renewal process with Markov renewal matrix $F(t) = \int_o^t \exp((Q - \Lambda)u)\, du\Lambda$. For that matter we closely follow the demonstration of Meier-Hellstern [2].

Denote by $M(t)$ the matrix whose entry (i,j) is equal to $M_{ij}(t) = \mathbb{P}(N(t) = 0, \lambda_t = \lambda_j \mid \lambda_0 = \lambda_i)$.

$M(t)$ satisfied to the following Chapman-Kolmogorov equation:

$$\begin{array}{lcl} M_{ij}(dt) & = & \sum_{k \neq j} M_{ik}(t) Q_{kj} dt + M_{ij}(t)(Q_{jj}(t) - \lambda_j) dt \\ M_{ij}(dt) & = & \sum_k M_{ik}(t)(Q_{kj} - \Lambda_{kj}) dt \\ M'(t) & = & M(t)(Q - \Lambda) \end{array}$$

with $M(0) = I_K$ so that $M(t) = \exp((Q - |Lambda)t)$.

Remark now that

$$F_{ij}(dt) = P(U_{n+1} \in dt, \bar{\lambda}_{n+1} = \lambda_j \mid \bar{\lambda}_0 = \lambda_i) = M_{ij}(t)\lambda_j dt$$

that is to say that

$$F'(t) = M(t)\Lambda = exp((Q - \Lambda)t)\Lambda$$

with the initial condition $F(0) = 0$ and though $F(t) = \int_0^t \exp((Q - \Lambda)u)\, du\Lambda = (\Lambda - Q)^{-1}(I_K - \exp((Q - \Lambda)x)\, \Lambda$.

References

[1] W. Willinger, M. Taqqu, R. Sherman, and D.V. Wilson, "Self-similarity through high variability: Statistical analysis of ethernet lan traffic at the source level," in *SIGCOMM'95*, Cambridge, USA, jan 1995.

[2] W. Fischer and K. Meier-Hellstern, "The markov modulated poisson process (mmpp) cookbook," *Performance Evaluation*, vol. 18, no. 2, pp. 149–171, 1993.

[3] M.S. Taqqu and J. Levy, "Using renewal processes to generate long-range dependence and high variability," in *Dependence in Probability and Statistics*, E. Eberlein and M.S. Taqqu, Eds., pp. 73–89. Birkhauser, Boston, 1986.

[4] S. Robert and J.Y. Le Boudec, "On a markov modulated chain exhibiting self-similarities over finite time-scales," *Performance Evaluation*, vol. 27, pp. 159–173, 1996.

[5] S. Robert and J.Y. Le Boudec, "New models for pseudo self-similar traffic," *Performance Evaluation*, vol. 30, pp. 57–68, 1997.

[6] V. Klemes, "The hurst phenomenon: a puzzle?," *Water Resour. Res.*, vol. 10, pp. 675–678, 1974.

[7] N. Duffield, G. Lewis, N. O'Connel, and F. Toomey, "Statistical issues raised by the bellcore data," Preprint.

[8] V. Teverovski and M. Taqqu, "Testing for long-range dependence in the presence of shifting means or a slowly declining trends, using a variance type estimator," *J. of Time Series Analysis*, vol. 18, no. 3, pp. 279–304, 1997.

[9] V. Paxson and S. Floyd, "Wide-area traffic: The failure of poisson modeling," in *SIGCOMM'94*, London, England, aug 1994, pp. 257–268.

[10] A. Mauchly, "Significance test for sphericity of a normal n-variate distribution," *Ann. Math. Stat.*, vol. 11, pp. 204–209, 1940.

[11] K. Drouiche, "Measuring randomness," Preprint. To appear in the Journal of the Royal Society B.

[12] A. Mokkadem, "On some new tests for time series," *Stochastic Processes and Their Applications*, vol. 318, pp. 755–758, 1997.

[13] F. Clerot, S. Bengio, A. Gravey, and D. Collobert, "Dynamical resource reservation schemes in an atm network using neural network-based traffic prediction," in *ATM'97*, Bradford, England, july 1997, pp. 03/1–11.

[14] M. Rosenblatt, *Stationary Sequences and Random Fields*, Birkhauser, 1985.

[15] D. Brillinger, *Time Series: Data Analysis and Theory (2nd edition)*, Applications of Mathematics. Holden D, 1981.

[16] R. Azencott and D. Dacunha-Castelle, *Séries d'Observations Irrégulières: Modélisation et Prévision*, Techniques Stochastiques. Masson, 1984.

[17] P.J. Brockwell and R.A. Davies, *Time Series: Theory and Methods; Second Edition*, Springer Series in Statistics. Springer-Verlag, 1991.

[18] D. Kofman, H. Korezlioglu, and M. Gagnaire, "An analytical description of packet trains model for local area network traffic characterization," in *2nd IFIP Workshop on Performance Modelling and Evaluation of ATM Networks*, july 1994.

[19] R.Jain and S.A. Routhier, "Packet trains: and a new model for computer network traffic," *IEEE-JSAC*, vol. 4, no. 6, pp. 986–995, sep 1986.

[20] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the —cal EM algorithm," *Journal of the Royal Statistical Society B*, vol. 29, pp. 1–38, 1977.

[21] P. Billingsley, *Probability and measure*, Wiley series in Probability and Mathematical Statistics. Wiley, 1979.

[22] B.G. Leroux, "Maximum likelihood estimation for hidden markov models," *Stochastic processes and their applications*, pp. 127–143, 1992.

[23] P.J. Bickel and Y. Ritov, "Inference in hidden markov models i: local asymptotic normality," *Bernoulli 2*, pp. 199–228, 1996.

[24] U. Holst and G. Lindgren, "Recursive estimation in mixture models with markov regime," *IEEE Trans. on Inform. Theory*, vol. 37, pp. 1683–1690, 1991.

[25] V. Krishnamurty and J.B. Moore, "On-line estimation of hidden markov models parameters, based on the kullback-leibler information measure," *IEEE Trans. on Signal Processing*, vol. 41, pp. 2557–2573, 1993.

[26] T. Ryden, "On recursive estimation for hidden markov models," *Stoch. Proc. and their Applic.*, vol. 66, pp. 79–96, 1997.

[27] L. Mevel, *Statistique Asymptotique pour les Modèles de Markov Cachés*, Ph.D. thesis, Université de Rennes 1, 1997.

[28] J. Beran, *Statistics for long memory processes*, Chapmann and Hall, 1994.

[29] H.E. Hurst, "Long term stotage capacity of reservoirs," *trans. Am. Sc. Civil. Eng.*, vol. 116, pp. 770–799, 1951.

[30] P.M. Robinson, "Gaussian semi-parametric estimation of long-range dependence," *The Annals of Statistics*, vol. 33, no. 5, pp. 1630, oct 1995.

[31] T.W. Epps, "Testing that a gaussian process is stationary," *The Annals of Statistics*, vol. 16, no. 4, pp. 1667, dec 1988.