A Markov Modulated process for self-similar traffic

Stephan Robert and Jean-Yves Le Boudec

Laboratoire de Réseaux de Communication (LRC),
EPFL CH-1015 Lausanne

September 23, 1995

Abstract

Recent papers have pointed out that data traffic exhibits self-similarity, but self-similarity is observed only on a finite timescale. In order to account for that, we introduce the concept of pseudo long-range dependencies. In this paper, we describe a Modulated Markov process producing self-similarity on a finite timescale; the process is quite easy to manipulate and depends only on three parameters (two real numbers and one integer). An advantage of using it is that it is possible to re-use the well-known analytical queueing theory techniques developed in the past in order to evaluate network performance. A quantitative method based on the *decomposability* theory of Courtois is given to evaluate the domain of validity where the process exhibits pseudo long-range dependencies. The tests performed to validate the model are the following: Hurst parameter (by the variances method) and the so-called 'visual' test. A method of fitting the model to measured data is also given. Finally, we describe a simulator generating traffic based on the Markov Modulated process.

Contents

1 Introduction

In this introduction, we want to present shortly the self-similarity concept before introducing the notion of *pseudo long-range dependencies.*

1.1 Background of self-similar processes

Recent studies of high quality traffic measurements have revealed that traffic behaves quite differently as what has been assumed until now. It has been observed that a large number of traffic sources produces a self-similar behavior in large time scales. The term "self-similar" is quite new (approximately 30 years old) [1, 2]. Intuitivelly, if macroscopically we see a form which is best described by a cluster. That cluster is composed of smaller clusters which looks almost identical to the entire cluster, however scaled down by some factor. Each of these smaller clusters is again composed of smaller ones, and those of even smaller ones again. Without difficulty, we can identify different generations of cluster on cluster; this is what we call self-similarity. Here, in the case of traffic behavior, self-similarity indicates that the behavior of a process is very similar (in a distributional sense). Mathematically, this concept is associated to the limit concept and therefore it will require some care. For time-series, the concept of "self-similarity" indicates a form of invariance with respect to changes of time-scales. Positive correlations for example are present at each time-scale, which is not the case for classical models. On the contrary, aggregated traffic become less bursty when the interval time grows. To summarize, a self-similar process can be sketched in the time direction by a factor a and in amplitude by r: $y(t) = rx(t/a)$, $y(t)$ is the rescaled random function and $x(t)$ the original random function. For ordinary Brownian motion, we need to scale amplitude by $\sqrt{2}$ when time is scaled by a factor of 2. Scaling amplitudes by other factors, such as 1 or 2, changes the statistical properties of the graphs. This Brownian motion is restrictive, therefore Mandelbrot introduced the fractional Brownian motion where it is possible to scale arbitrary the vertical axis by a factor between 1 and 2. For small factors ($\simeq 1$), the graph appears more rougher than the Brownian motion and for large factor ($\simeq 2$), the Brownian motion appears much smoother. For Brownian motion, this factor is set to 0.5. This factor is usually written as H and often called the Hurst exponent. Hurst [1] was an hydrologist who did some work with Mandelbrot on scaling properties of river fluctuation. The range for the exponent is from 0 (corresponding to very rough random fractal curves), to 1 corresponding to rather smooth looking random fractals. In fact, there exists a direct relation between H and the fractal dimension of the graph of a random fractal [3]. Ordinary Brownian motion is a process X(t) with Gaussian increments and $var(X(t_2) - X(t_1)) \propto |t_2 - t_1|^{2H}$ where $H = 1/2$. The generalisation of parameters $0 < H < 1$ is called fractional Brownian motion. If $X(t) - X(t_0)$ and $([X(rt) - X[t_0])/(r^H)$ are statistically indistinguable (they have the same finite dimensional joint distribution functions for any t_0 and $r > 0$), we say that they are statistically self-similar with parameter H. After this short introduction to self-similar processes, we suggest to consider pseudo long-range dependencies we meet with measured data traffic.

1.2 Pseudo long-range dependancies

In this section, we want to show you that the mathematical definition of long-range dependencies is very strict and can not be applied in the context of measured data on a network simply because the measurement time is finite. In fact, mathematically [4], the difference between short-range and long-range dependencies is clear: for a short-range dependent process

- $\sum_{\tau=0}^{\infty} cov(X_t, X_{t+\tau})$ *is convergent*
- *spectrum at* 0 *is finite*
- $var(X^{(m)})$ *is for large* m *asymptotically of the form* $\frac{varX}{m}$
- *the averaged process* $X_k^(m)$ *tends to second order pure noise as* $m \to \infty$

and for a long-range dependant process:

- $\sum_{\tau=0}^{\infty} cov(X_t, X_{t+\tau})$ *is divergent*
- *spectrum at 0 is singular*
- $var(X^{(m)})$ *is for large* m *asymptotically of the form* $m^{-\beta}$
- *the averaged process* $X_k^(m)$ *don't tend to second order pure noise as* $m \to \infty$

There is another cathegory: the processes of long-range dependencies of index β, but they don't have a degenerate correlational structure as $m \to \infty$.

All stationary autoregressive-moving average processes of finite order, all finite Markov chains (including semi-Markov processes) are included in the first category. In the second category, we have the fractional Brownian motion [5], ARIMA processes [6], Chaotic maps [7] which have long-range dependancies. So the game is very clear but if we look at this definition, we see that a process having "long term dependancies", but which is limited is considered as a short term dependancies process. We see for example that Ethernet measurements have long term dependancies, at least over 4 or 5 orders of magnitude. In other words, if we represent the number of Ethernet packets arriving in a time-interval of 1 s, then the statistics of the number of packets looks the same for 10 s, 100 s, 1000 s, 10000 s and is distinctively different from a pure noise. However in the order of days, researchers at Bellcore have observed a stabilisation of the index of dispersion [8] indicating a lack of self-similarity. So, according to the definition, a short term dependancies process would be sufficient to model LAN traffic. The difference with other processes (Poisson, ON-OFF, ...) is striking and they should be categorized differently. Therefore, we propose to name them: *pseudo long-range dependent processes*. An other problem which arises is the length of the measurements: no variable can be measured during an infinite amount of time. Especially in the netwoking world (in dimensioning purposes) where the traffic is increasing day after day at an uncontrollable rate. A pseudo long-range dependent process is able to model (as well as an (exactly) long-range dependent process) aggregated traffic over several timescales. In fact, this definition of long term dependancies processes comes historically from the importance of self-similar processes which are able to give an elegant explanation to an empirical law (*Hurst effect*). In practice, we have always a finite set of data and asymptotic conditions are never met. We propose here to model LAN traffic with Markov chains. As we have seen in chapter 2, the autocovariance function is a sum of exponentials. So we approximate a hyperbolic decaying autocovariance function by a sum of exponentials. A myth is to believe that is too complicated because of the number of parameters to manage [6]. In this paper, we give an example where it is sufficient to consider only three parameters and there are surely other solutions to resolve this problem. Because of this difficulty, markovian models have not been considered yet. In the next section, we propose a simple Markov modulated process drived by only three parameters and able to produce self-similar traffic on a finite timescale.

2 Markov model

In this paper, we investigate the use of a simple, discrete time Markov modulated model for representing self-similar data traffic. We consider the process of cell arrivals on a slotted link; call X_t the random variable representing the number of cells (assumed to be 0 or 1) during the t^{th} time slot, namely during time interval $[t-1,t)$. Let $Y_t = i$ be the modulator's state i, $i \in 1,2,3,...,n$ at time t. The arrivals of cells are modulated by a n-state discrete time Markov chain with transition probabilities $a_{ij}(t_1,t_2) = prob(Y_{t_2} = j|Y_{t_1} = i)$. Let ϕ_{ij} denote the probability of having j cells given that the modulator's state is i, more specifically $\phi_{ij} = pr(X_t = j|Y_t = i)$. The Markov modulated chain state probabilities are noted as $\pi_{i_t} = prob(Y_t = i)$, i is referred to the modulator's state and t to the time. The Markov modulated chain is assumed stationary and homogeneous. In this case, all the moments are equal and no more dependent of the time.

$$E[X_t^k] = E[X] = \vec{\pi}\mathbf{\Lambda}\vec{e}, \forall k, \forall t \tag{1}$$

with $\vec{\pi} = (\pi_1, \pi_2, ..., \pi_n)$ and $\vec{e}$ the unity vector. Λ is defined as

$$\Lambda = diag(E(X|Y=0), E(X|Y=1), ..., E(X|Y=n)) \tag{2}$$

If N_m represents the number of arrivals in a window of m time intervals, the variance of N_m can be written

$$var(N_m) = mE[X] - m^2(E[X])^2 + 2(\sum_{i=1}^{m-1}(m-i)(\vec{\pi}\Lambda\mathbf{A}^{i+1}\Lambda\vec{e})) \tag{3}$$

Now, we propose a family of models

$$\mathbf{A} = \begin{pmatrix} 1-1/a-1/a^2-...-1/a^n & 1/a & 1/a^2 & \cdots & 1/a^n \\ q/a & 1-q/a & 0 & \cdots & 0 \\ (q/a)^2 & 0 & 1-(q/a)^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (q/a)^n & 0 & 0 & \cdots & 1-(q/a)^n \end{pmatrix} \tag{4}$$

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{5}$$

So, the Markov chain is drived by only 3 parameters: a, q plus the number of states in the Markov chain n. Here, the expectation is given by

$$E[X] = \frac{1 - (1/q)}{1 - (1/q)^n} \tag{6}$$

3 Domain of validity where the process exhibits pseudo long-range dependencies

After an introduction to the Courtois's theory, we want to analyse matrices nearly completely decomposable (section 2) and completely decomposable (which will be constructed in section 3.2). The comparison of the behaviors of $\vec{\pi}_t$ and $\vec{\pi}_t^*$ allows us to determine the domain of validity where the process exhibits pseudo long-range dependencies.

3.1 Background on the *decomposability*

To build our model, we have considered a theory which was developped approximately 20 years ago by Courtois, the theory of *decomposability*. Courtois's analysis [9] is based on the important observation that large computing systems can usefully be regarded as nearly completely decomposable systems. Systems are arranged in a hierarchy of components and subcomponents with strong interactions within components at the same level and lower interactions with other components. Near decomposability has been observed in other domains than computing: in economics, in biology, genetics, social sciences. The pioneers in this domain are Simon and Ando who studied several study-cases in economics and in physics [10]. What they stated is that aggregation of variables in a nearly decomposable system must separate the analysis of the short term and long term dynamics. They proved two major theorems. The first says that a nearly-decomposable system can be analysed by a completely decomposable system if the intergroup dependancies are sufficently weak compared to intragroup ones. The second theorem says that even in the long term, the results obtained in the short term will remain approximately valid in the long term, as far as the relative behavior of the variables of the same group is concerned. In our study, the problem is inverse: we postulate the LAN traffic is composed of different timescales.

The Markov chain we propose to analyse here is in fact decomposable at several levels. In a first step, the development is done for only one level of decomposability. The presented development deviates from those of Simon, Ando and Courtois.

As seen before, the Markov chain to study is characterized (section **??**) by its transition matrix $(n * n)$ $\mathbf{A}$ and its state probabilities $\vec{\pi}$ ($\vec{\pi}_{t+1} = \vec{\pi}_t\mathbf{A}$), $\mathbf{A}$ is nearly decomposable for small $(q/a)^n$. Let $\mathbf{A}^*$ be completely decomposable, then $\mathbf{A}^*$ is composed of squared submatrices placed on the diagonal:

$$\mathbf{A}^* = \begin{pmatrix} \mathbf{A}_1^* & \cdots & 0 & 0 \\ 0 & \mathbf{A}_2^* & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \mathbf{A}_N^* \end{pmatrix} \tag{7}$$

The remaining elements are equal to zero. $\mathbf{A}_{IJ}^*$ is a sub-matrix of $\mathbf{A}$ at the intersection of the Ith set of rows and the Jth set of colomns and $a_{i_J j_J}$ the element at the intersection of the ith row and the jth colomn of $\mathbf{A}_{IJ}$. To $\mathbf{A}^*$ is associated a new $\vec{\pi}^*$: $\vec{\pi}_{t+1}^* = \vec{\pi}_t^* \mathbf{A}^*$. We have adopted the following notation: $\vec{\pi}^*$ is a horizontal vector the same size as $\mathbf{A}_{II}^*, I = 1, ..., N$. For simplification $\mathbf{A}_{II} = \mathbf{A}_i, i = 1, ..., N$ and is a square matrix $n(i) * n(i)$ with $\sum_{i=1}^{N} n(i) = n$. Each submatrix of $\mathbf{A}_i^*$ has its own set of eigenvalues $\lambda^*(i_I)$. For conviniance, we suppose they are ordered: $\lambda^*(1_I) = 1 > \lambda^*(2_I) \geq ... \geq \lambda^*(n(I)_I), I = 1, ..., N$. $\lambda^*(1_I) = 1$ because the matrix is stochastic [11]. With the matrix $\mathbf{A}$, the situation is different because only one eigenvalue $(\lambda(1_1))$ is equal to 1. We suppose the eigenvalues are ordered as well.

Suppose now $\mathbf{A}$ is diagonalisable, so

$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{D}\mathbf{P} \tag{8}$$

$\mathbf{P}$ is the passage matrix and $\mathbf{D}$ can be written

$$\mathbf{D} = \sum_l \lambda_l \mathbf{P}_l = \sum_l \sum_I \lambda(l_I)\mathbf{P}_I(l) \tag{9}$$

with $\mathbf{P}_l$ = projector (i.e. $p_{ij} = 0\ \forall i, j \neq l$, $p_{ll} = 1$). So

$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{P}_1(1)\mathbf{P} + \sum_{I=2}^{N} \lambda(1_I)\mathbf{P}^{-1}\mathbf{P}_I(1)\mathbf{P} + \sum_{I=1}^{N}\sum_{i=2}^{n(I)} \lambda(i_I)\mathbf{P}^{-1}\mathbf{P}_1(1)\mathbf{P} \tag{10}$$

$\lambda(i_I)\mathbf{P}^{-1}\mathbf{P}_1(1)\mathbf{P}$ can be replaced by $Z(i_I)$. The properties of $Z(i_I)$ are given in [9].

Similarly for $\mathbf{A}^*$, we have

$$\mathbf{A}^* = \mathbf{P}^{-1}\mathbf{P}_1(1)\mathbf{P} + \sum_{I=2}^{N} \lambda^*(1_I)\mathbf{P}^{-1}\mathbf{P}_I(1)\mathbf{P} + \sum_{I=1}^{N}\sum_{i=2}^{n(I)} \lambda^*(i_I)\mathbf{P}^{-1}\mathbf{P}_1(1)\mathbf{P} \tag{11}$$

Here we will give the first theorem of Simon and Ando [10] without demonstration:

Theorem 1 *For an arbitrary positive number ς, there exists a number ϵ_ς such that for $\epsilon \leq \epsilon_\varsigma$,*

$$\max_{k,l} |z_{kl}(i_I) - z_{kl}^*(i_I)| < \varsigma \tag{12}$$

with $2 \leq i \leq n(I), 1 \leq I \leq N, 1 \leq k, l \leq n$

Let us now focus our attention on the implication of this theorem. The discussion is intuitive but very important in this context. The time behavior of π_t and the comparison with the time behavior of π_t^* is at the center of the debate. Due to the eigenvalues' ordonance, the first terms of (11) will not imply a big variation in the short term ($t < T_2$), because the $\lambda(1_I)$ $I = 1, ..., N$ are close to unity. Thus, for $t < T_1$, the predominantly varying term of $\mathbf{A}$ is the last one, so π_t and π_t^* evolve similarly. For $T_1 < t < T_2$, the time behaviors of π_t and π_t^* are defined by the last terms of $\mathbf{A}$ and $\mathbf{A}^*$

respectively, a similar equilibrum is being reached within each subsystem of $\mathbf{A}$ and $\mathbf{A}^*$. For $T_2 < t < T_3$, the preponderantly varying term of $\mathbf{A}$ is the second one. For $t > T_3$, the first term of $\mathbf{A}$ dominates all the others. A global equilibrum is attained in the whole system. The whole nearly completely decomposable system moves towards equilibrum, but the short-term equilibrum relative system moves towards equilibrum, but the short-term equilibrum relative values of the variables within each subsystem are approximately maintained. This dynamic behavior of a nearly decomposable matrix may be dissociated into four distinguable periods that Simon and Ando [10] call, respectively, 1) short-term dynamics, $t < T_1$ 2) short-term equilibrum, $T_1 < t < T_2$ 3) long-term dynamics, $T_2 < t < T_3$ and 4) long-term equilibrum, $t > T_3$.

3.2 Use in our case

Here, we want to analyse matrices nearly completely decomposable with the form described in (section 2). The general form of the matrix is

$$\mathbf{A} = \begin{pmatrix} 1 - 1/a - 1/a^2 - ... - 1/a^n & 1/a & 1/a^2 & \cdots & 1/a^n \\ q/a & 0 & 1 - q/a & \cdots & 0 \\ (q/a)^2 & 0 & 1 - (q/a)^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (q/a)^n & 0 & 0 & \cdots & 1 - (q/a)^n \end{pmatrix} \tag{13}$$

and $\mathbf{A}^*$ built based on the Courtois's theory seen in the previous section.

$$\mathbf{A}^* = \begin{pmatrix} 1 - 1/a - 1/a^2 - ... - 1/a^N & 1/a & 1/a^2 & \cdots & 0 \\ q/a & 1 - q/a & 0 & \cdots & 0 \\ (q/a)^2 & 0 & 1 - (q/a)^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \tag{14}$$

Here, we have only two submatrices placed on the diagonal, $\mathbf{A}_1^*$ and $\mathbf{A}_2^*$. $\mathbf{A}_1^*$ is a square matrix $(n-1)*(n-1)$ and $\mathbf{A}_2^*$ a square matrix $1*1$. In this case, $n(1) = n - 1$ and $n(2) = 1$, $n(1) + n(2) = n$, $N = 2$. Note that $q < a$ and that $\mathbf{A}^*$ is a non-ergodic matrix.

3.3 comparison of the behavior of $\vec{\pi}_t^*$ and $\vec{\pi}_t$

to be completed

4 tests of self-similarity

4.1 Variance-time curve

The problem we are trying to solve in this chapter is to test the self-similarity and to measure the Hurst parameter H. In a first step, we want to examine second order stationary processes with autocovariance function $\gamma_\tau = cov(X_t, X_{t+\tau})$ with $\gamma_0 = \sigma = variance$. By definition [4], a process having only one Hurst parameter H to describe it is called exactly second-order self-similar. The process X_t and the averaged processes $X_t^{(m)}$ have identical correlational structures. With the Markov chains we analyse, we are not exactly in this case because all finite Markov chains have a limit. The Markov chains we examine have pseudo long-range dependancies. More formally, let $X = (X_t : t = 0, 1, 2, 3, ...)$ be a covariance stationary stochastic process with an autocovariance function γ_τ and $X^{(m)} = (X_k^m : k = 0, 1, 2, 3, ...)$ a new averaged series over non-overlapping blocks of size m (m is a time interval in our case) with an autocovariance function $\gamma_\tau^{(m)}$ $X_k^{(m)} = \frac{1}{m}\sum_{t=km-m+1}^{km} X_t$

- *If $\gamma_\tau^{(m)} = \gamma_\tau$ for $h \geq 0$ and $var(X^{(m)}) = \sigma^2 m^{-\beta}$ for all $m = 1, 2, 3, ...$ the process X is called (exactly) second-order self-similar with self-similarity parameter $H = 1 - \beta/2$ [6].*

- *If $\gamma_\tau^{(m)} = \gamma_\tau$ as $h \to \infty$ and $var(X^{(m)}) = \sigma^2 m^{-\beta}$ for $m \to \infty$, the process X is called (asymptotically) second-order self-similar.*

  We propose here a new definition

- *If $\gamma_\tau^{(m)} = \gamma_\tau$ for $h \geq 0$ and $var(X^{(m)}) = \sigma^2 m^{-\beta}$ for all $m = 1, 2, 3, ...m_{max}$, the process X is called (exactly) second-order pseudo self-similar.*

Notice that $0 < \beta < 2$ but here we are interested in the range $0 < \beta < 1$. The case $\beta = 1$ gives second-order pure noise with $varX^{(m)} = \frac{varX}{m}$. All finite Markov processes behave in this way when $m \to \infty$. $0 < \beta < 1$ indicates positive correlation through time-scales. To estimate the Hurst parameter of a second order self-similar process, it is sufficient to estimate β (estimate of β:$\hat{\beta}$) which is given by the slope of the diagram $\log_{10}(\frac{var(X^{(m)})}{\sigma^2})$ against $\log_{10}(m)$. The estimated slope $\hat{\beta}$ gives an estimate of H : $\hat{H} = 1 - \hat{\beta}/2$. Note that other methods are available to estimate $\hat{H}$: R/S statistics proposed by Mandelbrot [12] and periodograms [6].

4.2 Examples

In figures (1, 2 and 3), the variance-time plot is given for a Markov chain having the structure described in 3. In addition, a line indicates the slope of -1 (Hurst=0.5).

Let us see at examples with various parameters.

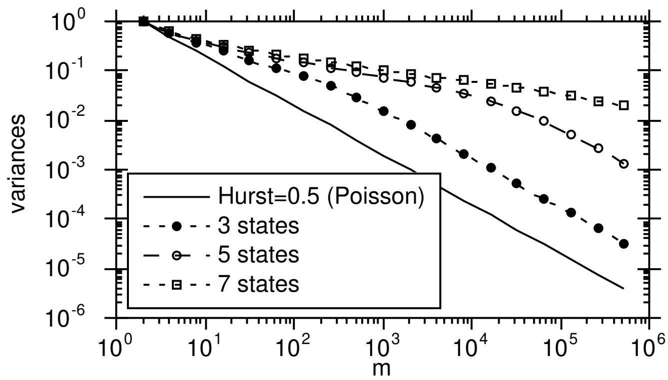

Figure 1: a=10 q=0.9 number of states=N=3/5/7

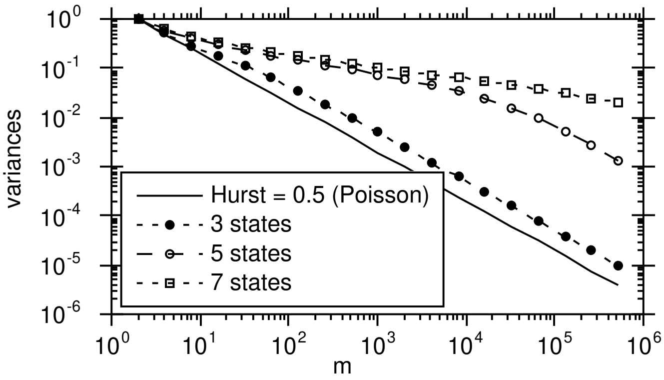

Figure 2: a=5 q=0.9 number of states=N=3/5/7

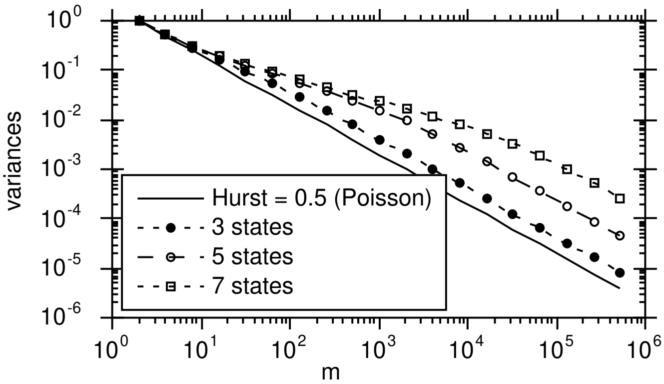

Figure 3: a=10 q=2 number of states=N=3/5/7

According to the previous figures, we see that the Hurst parameter varies with the number of states of the Markov chain. Note that the mean varies also with the number of states of the process. Here we introduce the "domain of validity" which is the domain where the Hurst parameter is not equal to 0.5, which is very appearent for the 3 and 5 state processes.

4.3 Visual test

In the introduction, we said that self-similarity is bounded with the concept of statistical invariance over timescales. So the easiest test to detect self-similarity is to trace measured traffic (like in [8]) or simulated traffic over timescales. This method is in fact very rough because it is difficult for example, to visually quantify the degree of similarity between scales ($\hat{\beta}$ can change over timescales for example). As observed in real traffic, we see similitudes over timescales. What is very different between our model and the ON-OFF source is the absence of natural length of burst in the first case like in the second. At each timscale, we observe bursty periods separated by less bursty subperiods like in the measured LAN-traffic [13] [6]. In contrast, the classical sources behave very differently because the correlation is present in general only at one timescale (ON-OFF sources for example). Even through the traffic could be very bursty at low timescales, it is not sufficient to reproduce real characteristics of measured traffic because of the lack of burstiness at higher scales [13].

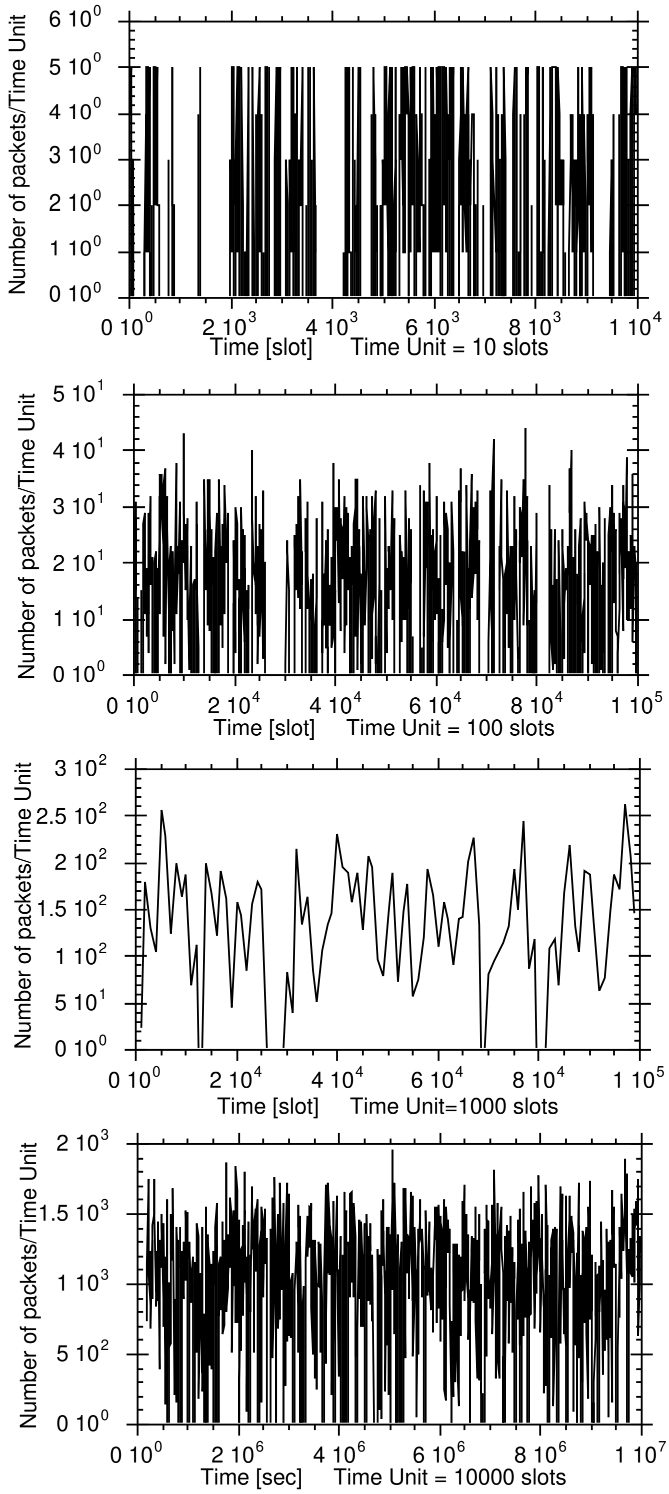

Figure 4: Behavior of the Markov process over 4 timescales, a=10 q=0.9 number of states=5

5 Fitting

Based on measurements, a lot of fitting procedures have been proposed in the litterature (see for example [14], [15]). Ours is based on the Markov chain described in (section 2). Here, we fit only two parameters: mean and Hurst parameter (plus the number of states in the Markov chain). As seen in (section **??**), EX equals $(1-(1/q))/(1-(1/q)^n)$. For a given EX, it is quite easy to find q iteratively. With the method described, it is possible to find the domain where the process is self-similar. This domain can be enlarged by increasing the number of states in the Markov chain, a is used to ajust the slope β in order to have the desired Hurst parameter (a is found iteratively with the Newton-Raphson method and $\hat{\beta}$ is estimated with least squares).

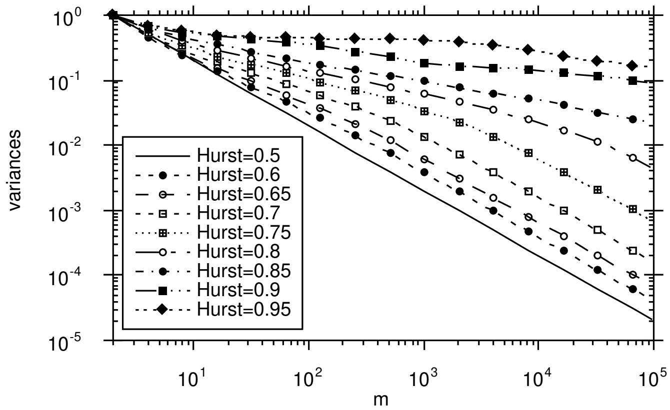

Figure 5: variance-plot for a Markov process of 5 states, Hurst parameter varies from 0.6 to 0.95

The validity domain (where $H \neq 0.5$) is determined by the method described in (section 3). Notice in figure 5 that the domain where the Hurst parameter is $\neq 0.5$ becomes larger when $H \rightarrow 1$. In other words, the Markov chain needs more states for self-similarity having a low Hurst parameter than a high one. The knee is very apparent, see for example at the curve where $H = 0.8$. For higher Hurst parameters, the knee is no more apparent because more computational time would be requiered to represent it. For extreme values of Hurst parameter ($H = 0.9, 0.95$), the variance-time curve becomes a little bit wavy.

6 Simulator

The simulation of such a Markov chain requiers care, especially because of the big differences we can find between two transition probabilities. If $a = 10$ and the Markov chain (8 states for example) is in the first state, $a_{12} = 10^{-1}$ but $a_{18} = 10^{-7}$. The ratio between the two extreme transition probabilities is in the order of 10^6. The number generator must be very good to generate such Markov chains. Note that the situation can be worse and no number generator would be able to generate traffic according to the given Markov chain. Because of that, it has been chosen to generate traffic iteratively.

Another problem is the goodness of the distribution (geometrical in our case) in each state, for parameters approaching 1. To test the generator, we have built a little test program where the geometrical distribution parameter varies from 0.5 to $1 - 0.5 * 10^{-8}$. The results show that the error between the mean (10000 samples) and the theorical value ($1/parameter$) is below 2.3 percent in the entire range. Outside of this range (for values nearer of 1 than $1 - 0.5 * 10^{-8}$), we got integer overflow problems.

7 Conclusions

In this paper, we have described a Markov chain producing self-similarity on a finite timescale which is quite easy to manipulate and depends only on three parameters. The tests performed on the model are the following: Hurst parameter (by the variances method) and the so-called 'visual' test. A method for fitting the model to measured data is also given. We think the theory of decomposability provides a fundamental explanation of the observed self-similarity on data networks. Furthermore, to consider only one relation between correlations over timescales seems to be irrealistic. It is surely necessary to take into account "local" self-similarities.

8 Acknowledgements

This work is supported by the Swiss PTT Telecom.

References

[1] R. P. B. H. E. Hurst and Y. M. Sinaika, "Long Term Storage in Reservoirs. An Experimental Study," *Constable, London*, 1965.

[2] J. W. V. N. B. Mandelbrot, "Fractional Brownian Motions, Frational Noises and Applications," *SIAM Review*, vol. 10, October 1968.

[3] D. S. H.-O. Peitgen, H. Jurgens, *Chaos and Fractals. New Frontiers of Science*. Springer-Verlag, 1992.

[4] D. R. Cox, "Long-range dependance: A review," *Statistics: An Appraisal*, 1984.

[5] I. Norros, "Studies on a model for connectionless traffic, based on fractional Brownian motion," *COST 242TD(92)041*, 1992.

[6] W. E. Leland, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," *IEEE/ACM Transactions on Networking*, February 1994.

[7] A. Erramilli, "Chaotic maps as models of packet traffic," *ITC-14*, June, Antibes Juan-Les-Bains 1994.

[8] W. E. Leland and D. V. Wilson, "High Time–Resolution Measurement and Analysis of LAN Traffic: Implications for lan Interconnection," *IEEE Infocom*, April 1991.

[9] P. J. Courtois, *Decomposability.* ACM Monograph Series, 1977.

[10] H. A. Simon and A. Ando, "Aggregation of variables in dynamic systems," *Econometrica*, no. 29, 1961.

[11] S. Friedberg, A. Insel, and L. Spence, *Linear Algebra.* Illinois State University: Prentice-Hall, 1989.

[12] B. Mandelbrot, "Robust R/S analysis of long run serial correlation," in *Proc. 42nd Session ISI*, pp. 69–99, 1979.

[13] S. Robert and J.-Y. L. Boudec, "Can self-similar traffic be modelled by Markovian processes ?," *COST 242TD(95)26*, May 1995.

[14] R. Gusella, "A Measurement Study of Diskless Workstation Traffic on an Ethernet," *IEEE Transactions on Communications*, vol. 38, September 1990.

[15] R. Grünenfelder and S. Robert, "Wich Arrival Law Parameters Are Decisive for Queuing System Performance ?," *ITC 14*, pp. 377–386, 6-10 June, Antibes, France 1994.