New Models for Pseudo Self-Similar Traffic

Stephan Robert* Jean-Yves Le Boudec †

July 3, 1996

Abstract

After measurements on a LAN at Bellcore, it is known that data traffic is extremely variable on time scales ranging from milliseconds to days. The traffic behaves quite differently to what has been assumed until now; traffic sources were generally characterized by short term dependences but characteristics of the measured traffic have shown that it is long term dependent. Therefore, new models (such as Fractional Brownian motion, ARIMA processes and Chaotic maps) have been applied. Although they are not easily tractable, one big advantage of these models is that they give a good description of the traffic using few parameters. In this paper, we describe a Markov chain emulating self-similarity which is quite easy to manipulate and depends only on two parameters (plus the number of states in the Markov chain). An advantage of using it is that it is possible to re-use the well-known analytical queuing theory techniques developed in the past in order to evaluate network performance. The tests performed on 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. In addition, considerations about *pseudo long range dependences* are exposed.

1 Introduction

Recent studies of high quality traffic measurements have revealed that traffic behaves quite differently to what has been assumed until now. It has been observed that a large number of traffic sources produces a self-similar [1, 2] behavior over large time scales. Imagine a cluster composed of smaller clusters which look almost identical to the entire cluster, but scaled down by some factor. Each of these smaller clusters is again composed of smaller ones, and those of even smaller ones again. We can identify without difficulty, 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

*University of California at Berkeley, Department of EECS, CA 94720, USA

†Swiss Fereral Institute of Technology (EPFL), Laboratoire des réseaux de communication, CH-1015 Lausanne, Switzerland

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: $y(t) = rx(t/a)$ where $y(t)$ is the rescaled random function and $x(t)$ the original random function. For ordinary Brownian motion, we need to scale the 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 arbitrarily scale the vertical axis by a factor between 1 and 2. For small factors ($\simeq 1$), the graph appears 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 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 generalization 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 undistinguable (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. In this case, if $X(t_0) = 0$ for $t_0 = 0$, $X(t)$ and $X(rt)/r^H$ are statistically undistinguable. Then, if $X(t)$ is accelerated by a factor r: $X(rt)$, it is rescaled by dividing the amplitudes by r^H.

2 New model

2.1 Considerations on pseudo long-range dependences

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 dependent 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 category: the processes that have long-range dependences of index β, but do not have a degenerate correlation 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]and chaotic maps [7] which have long-range dependences. However, if we look at this definition, we see that a process having "long term dependences", but which is limited is considered as a short term dependences process. We see for example that Ethernet measurements have long term dependences, 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 stabilization of the index of dispersion [8] indicating a lack of self-similarity. So, according to the definition, a short term dependences 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.* Another problem which arises is the length of the measurements: no variable can be measured during an infinite amount of time. Especially in the networking 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 time-scales. In fact, this definition of long term dependences 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 will see in section 2.2, the auto-covariance function is a sum of exponentials. So we approximate a hyperbolic decaying auto-covariance function by a sum of exponentials. One myth is that it 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.

2.2 Suggested process

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 in one time slot, 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.

The k^{th} moment of the random variable X_t is by definition

$$E[X_t^k] = \sum_{i \geq 1} x_i^k prob(X_t = x_i) \tag{1}$$

X_t is the random variable representing the number of cells ($x_0 = 0, x_1 = 1$) arriving during the interval $[t-1,t)$. Here

$$E[X_t^k] = \vec{\pi}_t \mathbf{\Lambda}_t^{(k)} \vec{e} \tag{2}$$

with $\vec{\pi}_t = (\pi_{1_t}, \pi_{2_t}, ..., \pi_{n_t})$ and $\vec{e}$ which is the unity vector.

$$\mathbf{\Lambda}_t^k = diag(E(X_t^k|Y_t = 0), E(X_t^k|Y_t = 1), ..., E(X_t^k|Y_t = n)) \tag{3}$$

For a wide sense stationary process (WSS), the k^{th} moment of X_t can be written $E[X_t^k] = \vec{\pi}\mathbf{\Lambda}^{(k)}\vec{e}$. The general expression of the auto-covariance function of the random variable X_t is [9]

$$cov(X_{t+\tau}^s, X_t^k) = E[X_{t+\tau}^s, X_t^k] - E[X_{t+\tau}^s]E[X_t^k] \tag{4}$$

For our process, $cov(X_{t+\tau}^s, X_t^k)$ becomes

$$cov(X_{t+\tau}^s, X_t^k) = \gamma_\tau = \vec{\pi}_t \mathbf{\Lambda}_t^k (\mathbf{A}(t+\tau,t) - \vec{e}\vec{\pi}_{t+\tau}) \mathbf{\Lambda}_{t+\tau}^s \vec{e} \tag{5}$$

with

$$\mathbf{A}(t+\tau,t)_{ij} = a_{kj}(t+\tau,t) = prob(Y_{t+\tau} = j|Y_t = k) \quad 1 \leq k,j \leq n \tag{6}$$

which are the elements of the Markov chain transition matrix, and if the process is WSS

$$cov(X_\tau^s, X_0^k) = \vec{\pi}_t \mathbf{\Lambda}_t^k (\mathbf{A}^{|\tau|} - \vec{e}\vec{\pi}) \mathbf{\Lambda}^s \vec{e} \quad |\tau| \geq 1 \tag{7}$$

The Markov chains we suggest to use are the following

$$\mathbf{A} = \begin{pmatrix} 1 - 1/a - 1/a^2 - ... - 1/a^{n-1} & 1/a & 1/a^2 & \cdots & 1/a^{n-1} \\ q/a & 1 - q/a & 0 & \cdots & 0 \\ (q/a)^2 & 0 & 1 - (q/a)^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (q/a)^{n-1} & 0 & 0 & \cdots & 1 - (q/a)^{n-1} \end{pmatrix} \tag{8}$$

$$
\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{9}
$$

So, the Markov chain has only 3 parameters: a, q plus the number of states in the Markov chain n.

2.3 Foundations of the model

To build our model, we have considered a theory which was developed approximately 20 years ago by Courtois, the theory of *decomposability*. Courtois's analysis [10] is based on the important observation that large computing systems can effectively 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 between 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 [11]. 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 analyzed by a completely decomposable system if the intergroup dependences are sufficiently weak compared to intra-group 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 the inverse: we postulate that the LAN traffic is composed of different time-scales.

The Markov chain we propose to analyze here is in fact decomposable at several levels. In a first step, the development is done for only one level of decomposability. The development deviates from those of Simon, Ando and Courtois. As seen before, the Markov chain to study is characterized (section 2.2) 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. Let $\mathbf{A}^*$ be completely decomposable, then $\mathbf{A}^*$ is composed of squared sub-matrices placed on the diagonal:

$$
\mathbf{A}^* = \begin{pmatrix} \mathbf{A}_1^* & \cdots & 0 & 0 \\ 0 & \mathbf{A}_2^* & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \mathbf{A}_N^* \end{pmatrix} \tag{10}
$$

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 columns and $a_{i_J j_J}$ the element at the intersection of the ith row and the jth column 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 sub-matrix 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 [12]. With the matrix $\mathbf{A}$, the situation is different because only one eigenvalue ($\lambda(1_1)$) is equal to 1. We assume the eigenvalues are ordered as well.

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

$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{D}\mathbf{P} \tag{11}$$

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

$$\mathbf{D} = \sum_i \lambda_i \mathbf{P}_i = \sum_{i=1}^{n(I)} \sum_{I=1}^{N} \lambda(l_I)\mathbf{P}_I(l) \tag{12}$$

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

$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{P}_1(1)\mathbf{P} + \sum_{I=2}^{N} \lambda(1_I)\mathbf{P}^{-1}\mathbf{P}_I(1)\mathbf{P} + \sum_{I=1}^{N} \sum_{i=2}^{n(I)} \lambda(i_I)\mathbf{P}^{-1}\mathbf{P}_I(i)\mathbf{P} \tag{13}$$

$\lambda(i_I)\mathbf{P}^{-1}\mathbf{P}_I(i)\mathbf{P}$ can be replaced by $\mathbf{Z}(i_I)$ ($z_{kl}(i_I)$ are the elements of $\mathbf{Z}(i_I)$). The properties of $\mathbf{Z}(i_I)$ are given in [10].

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

$$\mathbf{A}^* = \mathbf{P}^{-1}\mathbf{P}_1(1)\mathbf{P} + \sum_{I=2}^{N} \lambda^*(1_I)\mathbf{P}^{-1}\mathbf{P}_I(1)\mathbf{P} + \sum_{I=1}^{N} \sum_{i=2}^{n(I)} \lambda^*(i_I)\mathbf{P}^{-1}\mathbf{P}_I(i)\mathbf{P} \tag{14}$$

Here we will give the first theorem of Simon and Ando [11] without demonstration:

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

$$\max_{k,l} |z_{kl}(i_I) - z^*_{kl}(i_I)| < \varsigma \tag{15}$$

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' ordinance, the first terms of 14 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 equilibrium 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 equilibrium is attained in the whole system. The whole nearly completely decomposable system moves towards equilibrium. The short-term equilibrium relative system moves towards equilibrium, but the short-term equilibrium relative values of the variables within each subsystem are approximately maintained. This dynamic behavior of a nearly decomposable matrix may be dissociated into four distinguishable periods that Simon and Ando [11] call, respectively, 1) short-term dynamics, $t < T_1$ 2) short-term equilibrium, $T_1 < t < T_2$ 3) long-term dynamics, $T_2 < t < T_3$ and 4) long-term equilibrium, $t > T_3$.

Here, we want to analyze matrices nearly completely decomposable with the form described in section 2.2. The general form of the matrix is

$$\mathbf{A} = \begin{pmatrix} 1-1/a-1/a^2-...-1/a^{n-1} & 1/a & 1/a^2 & \cdots & 1/a^{n-1} \\ q/a & 1-q/a & 0 & \cdots & 0 \\ (q/a)^2 & 0 & 1-(q/a)^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (q/a)^{n-1} & 0 & 0 & \cdots & 1-(q/a)^{n-1} \end{pmatrix} \tag{16}$$

and

$$\mathbf{A}^* = \begin{pmatrix} 1-1/a-1/a^2-...-1/a^{n-2} & 1/a & 1/a^2 & \cdots & 0 \\ q/a & 1-q/a & 0 & \cdots & 0 \\ (q/a)^2 & 0 & 1-(q/a)^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \tag{17}$$

with $q < a$.

$\mathbf{A}^*$ is a non-ergodic matrix. The comparison of the behaviors of $\vec{\pi}_t$ and $\vec{\pi}_t^*$ allow us to determine the "domain of validity" (by the variance method, section 3.2).

2.4 Characteristics of the process

For a WSS process, the variance is given by

$$var(N_m) = m(\vec{\pi}_m \mathbf{\Lambda}_m^{(2)} \vec{e}) - m^2(\vec{\pi}_m \mathbf{\Lambda}_m \vec{e})^2 + 2(\sum_{i=1}^{m-1}(m-i)(\vec{\pi}_m \mathbf{\Lambda}_m \mathbf{A}^i \mathbf{\Lambda}_m \vec{e})) \tag{18}$$

and the expectation is given by

$$E[X] = \frac{1-(1/q)}{1-(1/q)^n} \tag{19}$$

2.5 Simulator

The simulation of such a Markov chain requires 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. If this Markov chain is simulated with classical methods, the random number generator must be very good. If the transitions probabilities ratio is bigger, no random number generator is able to generate traffic according to the Markov chain. Therefore, we used an iterative method. Suppose for example that $a = 10$ and that the Markov chain is in the first state, then a number between 0 and 1 is generated with a reliable random number generator [13]. If the next number is less than $1/a = 1/10$, the next state is not 2 but between 3 and n. The procedure continues until the random number is more than $1/a$ or until the last state is reached. To resume, the probability of being in the next step in the state i knowing that we are in the first state is $prob(Y_{t+1} = i|Y_t = 1) = (1/a)^{i-1}$ with $i = 2, 3, \cdots, n$ and the probability of staying in the first state is given by $prob(Y_{t+1} = 1|Y_t = 1) = 1 - \sum_{i=2}^{n}(1/a)^{i-1}$. So, we are not limited by the number of states in the Markov chain. We have seen how it is possible to change the state if we are in the first state but if we are in another state, the transition probability ratios can be very high too. If we take the preceeding example, in the eighth state, the transition probability $prob(Y_{t+1} = 1|Y_t = 8) = (q/a)^8 = 4.3 * 10^{-9}$ and $prob(Y_{t+1} = 8|Y_t = 8) = 1 - 4.3 * 10^{-9}$. The problem of big differences between transition probabilities remains. One technique would be to use the method used for the state in having a threshold for example, but instead of that, we used another, more efficient and quicker method. We don't chose at each slot if we remain or not in the same state but we calculate how long we stay in the present state. The time we stay in the same state is geometrically distributed. The probabilities of staying in the same state are near 1, therefore, it is useful to test the generated distribution. This was tested for parameters varying from 0.5 to $1 - 0.5^{-8}$. For each parameter, 10000 samples have been generated and the maximal error found between the expectation of the distribution and the mean of the simulations is 2.3%. The modulator changes state according to the iterative method seen before and when the modulator is in a state, the sojourn time is found with the random number generator. So, at each iteration, the modulator changes state.

3 Tests of self-similarity

3.1 Visual test

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. Figure 1 shows that 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 a natural length of burst in the first case as in the second. At each timescale, we observe bursty periods separated by less bursty subperiods like in the measured LAN traffic [14, 6]. In contrast, the classical sources behave very differently because the correlation is present in general only at one timescale. Even through the traffic could be very

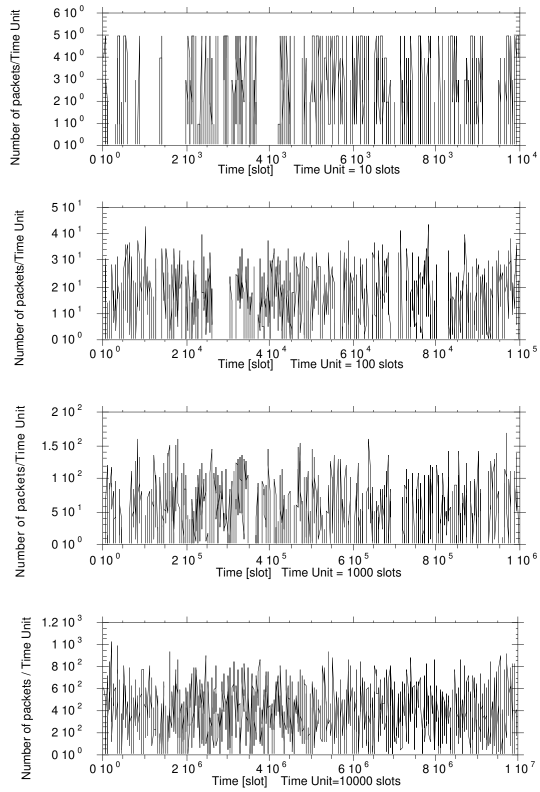

Figure 1: 5-state Markov chain time evolution on 4 different timescales, $H_l = 0.78$

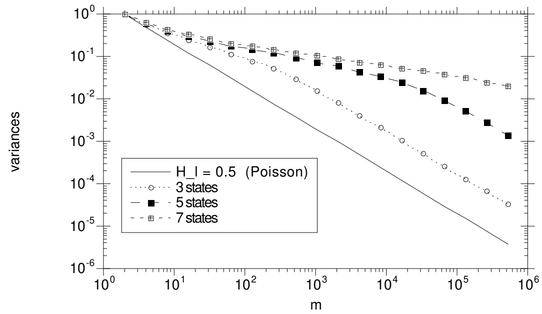

Figure 2: a = 10, q = 0.9, number of states = n = 3/5/7

bursty at low timescales, it is not sufficient to reproduce real characteristics of measured traffic because of the lack of burstiness at higher scales [14].

3.2 Method of the variances to test the self-similarity of the process

The problem we are trying to solve in this Section 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 $cov(X_t, X_{t+\tau})$ with $cov(X_t, X_t) = var(X_t)$. 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 will analyse, we are not exactly in this case because all finite Markov chains have a limit. Therefore, we propose to use the name of "local" Hurst H_l parameter instead of Hurst parameter for our Markov chains. The Markov chains we examine have pseudo long-range dependences.

- If $cov(X_t^{(m)}, X_{t+\tau}^{(m)}) = cov(X_t, X_{t+\tau})$ for $\tau \geq 0$ 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 $cov(X_t^{(m)}, X_{t+\tau}^{(m)}) = Cov(X_t, X_{t+\tau})$ as $\tau \to \infty$ and $var(X^{(m)}) = \sigma^2 m^{-\beta}$ for $m \to \infty$, the process X is called (asymptotically) second-order 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 $var(X^{(m)}) = var(X)/m$. 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 which is given by the slope of the diagram $\log_{10}(var(X^{(m)})/\sigma^2)$ against $\log_{10}(m)$. The estimated slope β gives an estimate of $H : H = 1 - \beta/2$. Note that other methods are available to estimate H: R/S statistics proposed by Mandelbrot [15] and periodograms [6]. In our context, we estimate the local Hurst parameter H_l. In Figures 2, 3 and 4, the variance-time plot is given for a Markov chain having the structure described in 2.3. In addition, a line indicates the slope of -1 (Hurst=0.5). According to the Figures, we see that the local Hurst parameter

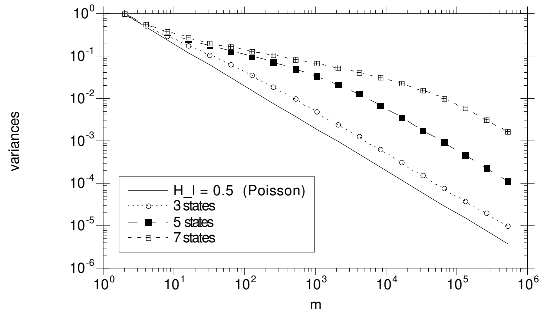

Figure 3: a = 5, q = 0.9, number of states = n = 3/5/7

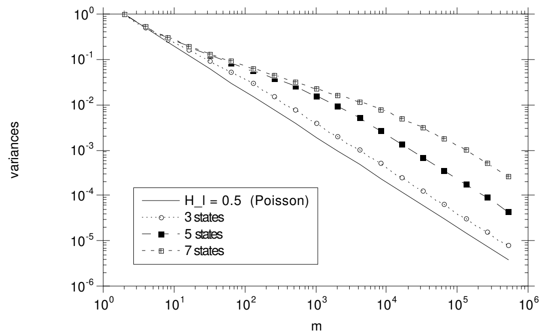

Figure 4: a = 10, q = 2, number of states = n = 3/5/7

varies with the number of states of the Markov chain. Note that the expectation varies also with the number of states of the process. Here we introduce the "domain of validity" which is the domain where $H_l \neq 0.5$, which is very appearent for the 3 and 5 state processes in Figures 2, 3 and 4.

4 Fitting

Based on measurements, a lot of fitting procedures have been proposed in the literature (see for example [16], [17]). Ours is based on the Markov chain described in (section 2.2). Here, we fit only two parameters: mean and Hurst parameter (plus the number of states in the Markov chain). As seen in (section 2.4), $E[X]$ equals $(1-(1/q))/(1-(1/q)^n)$. For a given $E[X]$, it is quite easy to find q iteratively (algorithm 1). 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 adjust 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.

Algorithm 1

Initialisation

Input of $E[X]$

$q = q_{init} = q_{old}$

repeat

$E[X]_{found} = (1 - 1/q)/(1 - (1/q)^n)$

if $(E[X]_{found} > E[X])$ **then**

$q_{old} = q$

$q = q/2$

else

$q = (q_{old} - q)/2$

until $(E[X] - E[X]_{found} < \epsilon)$

end

The validity domain (where $H_l \neq 0.5$) is determined by the method described in (section 2.3). Notice in figure 5 that the domain where the local Hurst parameter is $\neq 0.5$ becomes larger when $H_l \rightarrow 1$. In other words, the Markov chain needs more states for self-similarity having a low local Hurst parameter than a high one. The knee is very apparent, see for example at the curve where $H_l = 0.8$. For higher local Hurst parameters, the knee is no more apparent because more computational time would be required to represent it. For extreme values of local Hurst parameter ($H_l = 0.9, 0.95$), the variance-time curve becomes a little bit wavy. Table 1 shows us some values found with the algorithm.

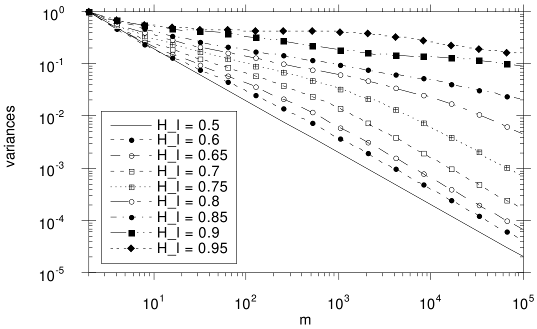

Figure 5: variance-time plot for a Markov process of 5 states, local Hurst parameter varies from 0.6 to 0.95

n	H_l	a	q
4	0.70	2.65	0.4418
	0.75	4.97	
	0.80	10.27	
	0.85	28.49	
5	0.70	2.60	0.5764
	0.75	4.38	
	0.80	6.70	
	0.85	12.97	
	0.90	85.61	
6	0.70	1.99	0.6737
	0.75	3.22	
	0.80	4.80	
	0.85	8.21	
	0.90	50.02	

Table 1: a and q as a function of H_l and n, $E[X] = 0.05$

5 Conclusions

In this paper, we have described a Markov chain emulating self-similarity which is quite easy to manipulate and depends only on three. The tests performed on 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 was also given.

6 Acknowledgments

This work is supported by the Swiss PTT Telecom.

References

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

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

[3] H.-O. Peitgen, H. Jurgens, and D. Saupe, *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," Tech. Rep. TD(92)041, COST 242, 1992.

[6] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," *IEEE/ACM Transactions on Networking*, vol. 2, February 1994.

[7] A. Erramilli, "Chaotic maps as models of packet traffic," in *ITC 14, The Fundamental Role of Teletraffic in the Evolution of Telecommunications Networks* (J. Labetoulle and J. Roberts, eds.), (Antibes Juan-Les-Pins, France), Elsevier Science Publishers B.V. (North-Holland), June 6-10 1994.

[8] W. Leland and D. Wilson, "High Time–Resolution Measurement and Analysis of LAN Traffic: Implications for LAN Interconnection," in *IEEE Infocom*, (Bal Harbor, FL), April 1991.

[9] A. Papoulis, *Probability, Random Variables, and Stochastic Processes.* Mc Graw Hill, 1984.

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

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

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

[13] P. L'Ecuyer, "Efficient and Portable Combined Random Number Generators," *Communications of the ACM*, vol. 31, pp. 742–774, June 1988.

[14] S. Robert and J.-Y. LeBoudec, "Can self-similar traffic be modelled by Markovian processes ?," in *1996 International Zurich Seminar on Digital Communications, IZS'96*, (ETH-Zentrum, Zurich, Switzerland), February 19-23 1996.

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

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

[17] R. Grünenfelder and S. Robert, "Wich Arrival Law Parameters Are Decisive for Queuing System Performance ?," in *ITC 14, The Fundamental Role of Teletraffic in the Evolution of Telecommunications Networks* (J. Labetoulle and J. Roberts, eds.), (Antibes Juan-Les-Pins, France), pp. 377–386, Elsevier Science Publishers B.V. (North-Holland), June 6-10, 1994.