An application of superpositions of two-state Markovian sources to the modelling of self-similar behaviour

Andersen, Allan T.; Nielsen, Bo Friis

Published in:
IEEE INFOCOM Kobe, Japan

Link to article, DOI:
10.1109/INFCOM.1997.635130

Publication date:
1997

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
An Application of Superpositions of two state Markovian Sources to the Modelling of Self-similar Behaviour

Allan T. Andersen and Bo Friis Nielsen
Department of Mathematical Modelling (IMM), Building 321
Technical University of Denmark
DK-2800 Lyngby, Denmark

Abstract
We present a modelling framework and a fitting method for modelling second order self-similar behaviour with the Markovian Arrival Process (MAP). The fitting method is based on fitting to the autocorrelation function of counts for a second order self-similar process. It is shown that with this fitting algorithm it is possible closely to match the autocorrelation function of counts for a second order self-similar process over 3-5 time-scales with 8-16 state MAPs with a very simple structure i.e. a superposition of 3 respectively 4 Interrupted Poisson Processes (IPP) and a Poisson process. The fitting method seems to work well over the entire range of the Hurst parameter.

1 Introduction
Recent measurement studies of packet traffic from Local Area Networks (LAN) [11] and Variable Bit Rate (VBR) video [6] reveal behaviour very different from Poisson. The measurements indicate that the number of packets arriving in a time-slot has long range dependencies and slowly decaying variances. Characteristics of the observed behaviour have in other contexts been associated with self-similar behaviour. Therefore the measurements motivate the assertion that certain types of packet traffic exhibit self-similar behaviour over several time-scales.

Until recently it has not been clear whether traditional Markov based models could be used to model this behaviour. It has been claimed that the necessary state space would be so large that Markov models would be inapplicable for all practical purposes. This has initiated a search for other models inherently more suitable for modelling self-similar behaviour e.g. chaotic maps [9] and Fractional Brownian Motions (FBM) [23]. For these models, however, the tools for analyzing queueing behaviour are still in an embryonic state.

In recent previous work [1],[2] and references therein the authors have applied Markovian Arrival Processes (MAP) to model self-similar behaviour. The results suggest that it is possible to model self-similar behaviour over several time-scales with very simple MAPs. However, in these references we had not yet developed a systematic method for the fitting of MAPs to a given self-similar structure. Such a fitting algorithm is the main contribution of this paper.

Related recent work regarding the use of Markovian processes to the modelling of self-similar behaviour can be found in [27]. Related work regarding the layered construction of models with self-similar behaviour was presented in [14].

This contribution can roughly be divided in two parts. The first part describes how the underlying Markovian environment of the MAP can be constructed as a superposition of two state Markov processes. This approach implies important analytical simplifications which are very useful when calculating moments of the counting process. The second part describes a method of choosing the parameters of a MAP which is to exhibit second-order self-similar behaviour. The fitting algorithm proposed is applicable for superpositions of two state sources and thus processes which are within the class discussed in the first part.

We shall not describe in detail the definitions of self-similarity and long-range dependence but refer the interested reader to e.g. [7], [15] and [21]. For self-similarity in the context of teletraffic modelling see e.g. [12] and [29]. For a description of MAP models in general see the original work done by M. F. Neuts and co-workers e.g. [19], [20], [22] and [26].

The paper is organized as follows: In section 2 we explain the idea of using superpositions of two state processes to model self-similar behaviour. Section 3 contains a description of an algorithm for the fitting of a MAP to match the autocorrelation function of counts of a second order self-similar process over sev-
Joint Markov chain of three independent two state Markov chains. Each of the transition intensity variables $a$, $b$ and $c$ can assume two different values depending on the direction of the transition i.e. $a = c_{12}$ or $a = c_{23}$, $b = c_{12}$ or $b = c_{22}$ and $c = c_{11}$ or $c = c_{21}$.

Figure 1:

2. **Superposition of two state processes**

We apply the continuous time Markovian Arrival Process (MAP) [19], [20] and [22] to model self-similar behaviour over several time scales. The MAP is a Markov renewal process whose transition probability matrix $F(*)$ is of the form

$$F(x) = \int_0^x e^{D_0 u} du D_1$$

where the matrices $D_0 = [D_{0ij}]$ and $D_1 = [D_{1ij}]$ are respectively a stable matrix and a non-negative matrix whose sum is an irreducible infinitesimal generator $D$ with stationary probability vector $\pi$.

We will explain the idea of the construction of MAPs with apparently self-similar behaviour qualitatively using figure 1. For a more comprehensive mathematical treatment see [1]. We consider the generator of the underlying Markov chain of a MAP $D$, without considering the special structure of the two component matrices $D_0$ and $D_1$. What we are looking for is a Markov chain model of a physical phenomenon exhibiting self-similar features. On the very macroscopic level what we see is just a constant rate. I.e. on a very large time-scale there is hardly any variability at all. This of course corresponds to time-scales where the self-similar behaviour does not appear. Imagine that after zooming in on the process, i.e. considering a finer time-scale, the process behaves roughly like a two state process. We model this behaviour with a two state Markov chain with mean sojourn times in each state chosen to be in accordance with the time-scale under consideration. Now zooming in again on each of the two states we can imagine that under a closer look each of the states again exhibit two state behaviour but with shorter mean sojourn times. Doing this repeatedly we obtain a model consisting of a hierarchy of nested two state processes. Of course for a finite dimensional Markov chain it is only possible to do this a limited number of times.

To exemplify the method consider again figure 1 with states labeled 1-8. Let the set of states 1-4 be the sub-states of one of the long term macro states and 5-8 be the sub-states of the other. It is understood that the mean sojourn time in each of the two states is long compared to the other time-constants in the system. This corresponds to the initial zooming in on the process. Each of these two sets of sub-states (states 1-4 and 5-8 respectively) can again be divided into two sets of sub-states namely states 1-2 and 3-4 respectively 5-6 and 7-8. Let each of these 4 sets of states be sub-states of the medium term macro states. Finally, each of these four sets of sub-states can be divided into two short term states which are the actual states in the Markov chain. The possible transitions are illustrated in Figure 1. From state 1 there are three possible transitions: a short term transition that will take us to state 2, a mid term transition to state 3 and finally a long term transition that will take us to state 5. Clearly, the most likely transition is to state 2. The mean sojourn time in a state is to a large extent determined by the fastest transition intensity.

To obtain analytical simplifications, we further assume that the sub Markov chains of any two corresponding states are identical. I.e. the transitions that take place on each level in the nested structure are independent. Clearly the underlying Markov chain under consideration is thus a superposition of a number of sub Markov chains. To avoid confusion it is emphasized that the MAP we are considering is not necessarily itself a superposition of two state MAPs, whether this is true or not depends on the way we choose the matrix $D_1$. The underlying Markov chain of each of the two state processes can be described in the following manner.

$$D_1^* = \begin{pmatrix} -c_{11} & c_{11} \\ c_{21} & -c_{21} \end{pmatrix}$$

Hence the superposition can be described as
where \(d\) is the number of superpositions (or depth of the model) and \(\bigoplus\) denotes the Kronecker sum [17] p. 412.

Here an illustration has been given of a methodology for constructing self-similar behaviour over several time-scales using two state processes as building blocks, however, this approach can easily be generalized to allowing more general processes as building blocks. This generalization, the General Multi Level Process (GMLP) is explained in e.g. [11] p. 411 ff.

The model is to find a MAP with a correlation structure which closely matches that of a second order self-similar process over several time-scales. Exact second order self-similarity is completely characterized by three parameters: the mean rate, an absolute measure of the variance and the Hurst parameter. The absolute measure of the variance and the Hurst parameter describe the correlation structure.

Second order self-similarity can be described in a number of equivalent ways e.g. the autocovariance function of the number of packets in a slot behaves asymptotically as \(\gamma_r(k) = \psi_{cov} k^{-\beta}\), where \(\psi_{cov}\) is an absolute measure of the variance (a positive constant), \(0 < \beta < 1\) (\(\beta = 2 - 2*H\)), and \(k\) is the lag. For the \(i\)th IPP the covariance between the number of events in two timeslots of size \(\Delta t\) with \(k - 1\) timeslots between them is expressed by \(k > 0\).

\[
\gamma_i(k) = \frac{\lambda^2 \delta_{11} \delta_{21} e^{-((\lambda_1 + \lambda_2)(k-1)\Delta t)}}{(\lambda_1 + \lambda_2)^k} \\
\times (1 - 2e^{-((\lambda_1 + \lambda_2)\Delta t)} + e^{-((\lambda_1 + \lambda_2)2\Delta t)})
\]

Clearly the corresponding covariance for a Poisson process.
process is zero. The covariance of a process composed of a superposition of independent subprocesses is found by summing the individual covariances.

Even with the relatively simple two state based models considered here there are potentially many parameters to be fitted. It is therefore convenient to limit the number of parameters by initially choosing the modulating parameters in the IPPs to satisfy $c_{11} = c_{2i} = a^{1-i}c_{11}$ for $i = 1, \ldots, d$ i.e. the modulating parameters are chosen to be logarithmically spaced with a factor $a$. This approach appears to be in good agreement with the constructive idea behind the models. For each fit we also initially choose the number $d$ of IPPs to be used.

A procedure is then needed to find a) the IPP arrival intensities $\lambda_i$, b) the Poisson arrival intensity $\lambda_P$, c) the absolute values of one of the modulating parameters e.g. $c_{11}$ and d) the logarithmic spacing parameter $a$. These quantities should be obtained from the following inputs i) the fundamental rate $\lambda^*$, ii) the Hurst parameter $H = 1 - \frac{\beta}{2}$, iii) an absolute variance measure and iii) the number of time-scales $n$ over which a fixed Hurst parameter is to be modelled.

The fitting procedure is a four step procedure. To avoid trivialities the shortest time-scale of interest is defined in relation to a reasonable level of packet traffic activity. The fundamental arrival rate is given relative to this time-scale and thus assumed to be between 1.0 and 10.0.

**Step one**

The first step is to find the logarithmic spacing parameter $a$. This step requires the number of IPPs $d$ and the number of time-scales $n$.

The parameter $a$ is determined from

$$a = 10^{\frac{n}{d}}, d > 1$$

The parameters $n$ and $d$ should be chosen so that $a \geq 5$ due to a fundamental assumption in step two.

**Step two**

The second step is to determine the arrival intensities $\lambda_i$ up to a normalizing constant, $\lambda = \kappa \cdot \lambda_i$. This is the main part of the algorithm. The overall correlation structure is determined here while the other steps more or less are used for adjusting and tuning.

The step requires the Hurst parameter $H (= 1 - \beta/2)$, the number of time-scales $n$ and the number of IPPs $d$.

The procedure is to determine the parameters, $\phi$, such that for some $k_0 > 0$

$$\gamma(k) = \sum_{i=1}^{d} \gamma_i(k) \approx \psi_{cov} k^{-\beta}, k_0 \leq k \leq k_010^n$$

Assuming that $\Delta t$ is chosen so that $(c_{11} + c_{2i}) \Delta t < 1, i = 1, \ldots, d$, a Taylor expansion of the $i$th IPPs covariance functions yields:

$$\gamma_i(k) \approx \frac{1}{4} (\Delta t)^2 \lambda_i^2 e^{-(c_{11} + c_{2i}) \Delta t k}$$

where it is used that $c_{11} = c_{2i}$. A key element of the fitting procedure is that for each individual IPP the covariance is fairly constant for all lags up to some $k : (c_{11} + c_{2i}) \Delta t k \approx 1$. From this point it decays rapidly and is magnitudes lower at lags where $(c_{11} + c_{2i}) \Delta t k > 5$. Due to the fact that the modulating intensities have been chosen to be logarithmically spaced initially this can be exploited to find the relative size of the IPP arrival intensities assuming that the logarithmic spacing parameter $a$ is not too small. Without loss of generality we define $\Delta t = 1$ i.e. equal to one unit on the smallest time-scale. The covariance function is fitted at $d$ different points defined by $(c_{11} + c_{2i}) k \approx 1$. With an arbitrary scaling of the covariance function i.e. $\psi_{cov}$ we obtain

$$(c_{1d} + c_{2d}) k = 1 :$$

$$\psi_{cov} a^{-(d-1)\beta} = \frac{1}{4} \sum_{j=1}^{d} (\phi_j)^2 e^{-a^{d-j}} \approx \frac{1}{4} (\phi_d)^2 e^{-1}$$

$$(c_{1(d-1)} + c_{2(d-1)}) k = 1 :$$

$$\psi_{cov} a^{-(d-2)\beta} = \frac{1}{4} \sum_{j=1}^{d} (\phi_j)^2 e^{-a^{d-2-j}} \approx \frac{1}{4} (\phi_d)^2 e^{-a-1} + \frac{1}{4} (\phi_{d-1})^2 e^{-1}$$

From these equations the relative magnitudes $(\phi_j)^2, i = 1, \ldots, d$ are obtained iteratively. The first equation gives $(\phi_d)^2$, the second equation gives $(\phi_{d-1})^2$ and so on. Not all the equations can necessarily be satisfied, in these cases the corresponding intensity $\phi_j$ is set to 0.

2b.3.4
Step three
The third step is to determine the constant $\kappa$ from the previous step and the intensity of the Poisson process $\lambda_P$. This step requires the one step correlation $\rho$, the fundamental rate $\lambda^*$ and $\phi$ obtained in the previous step.

With the assumption $(c_{1i} + c_{2i})\Delta t << 1$, $i = 1, \ldots, d$ a Taylor expansion of the IPP variance function [13] respectively the covariance function yields the following Taylor approximation for the one step correlation $\rho$. (Note that in our case the IPPs are “symmetric” i.e. $c_{1i} = c_{2i}$ and that $\Delta t = 1$).

$$\rho \approx \frac{\frac{1}{2} \sum_{j=1}^{d} \lambda_j^2}{\lambda^* + \frac{1}{2} \sum_{j=1}^{d} \lambda_j^2} = \frac{\frac{1}{2} \sum_{j=1}^{d} \lambda_j^2}{\lambda^*} = \frac{2\kappa \sum_{j=1}^{d} \phi_j^2}{\lambda^* + \kappa^2 \sum_{j=1}^{d} \phi_j^2}$$

Isolating $\kappa$ gives

$$\kappa \approx 2 \sqrt{\frac{\rho \lambda^*}{(1-\rho) \sum_{j=1}^{d} \phi_j^2}}$$

The $\kappa$ value obtained here is not necessarily feasible since it must satisfy

$$\kappa \leq 2 \frac{\lambda^*}{\sum_{j=1}^{d} \phi_j}$$

In the infeasible case $\kappa$ is determined by the nearest match

$$\kappa = 2 \frac{\lambda^*}{\sum_{j=1}^{d} \phi_j}$$

Clearly $\lambda_P = 0$ in this case.

If $\kappa$ is feasible $\lambda_P$ is readily found as

$$\lambda_P = \lambda^* - \frac{1}{2} \kappa \sum_{j=1}^{d} \phi_j$$

An infeasible $\kappa$ corresponds to the case where the desired value of $\rho$ is to large to be met with this procedure instead the largest possible value of $\rho$ is chosen.

Step four
This step is to determine the value of $c_{11}$. $c_{11}$ is determined by a brute force least squares fit of the correlation function of the model to $\rho \ast k^{-\beta}$. The fit is performed over the time-scales from $k = 1$ to $k = 10^n$.

To summarize, we end up with a fitting algorithm for parameterizing a MAP to imitate self-similar behaviour over several time scales which requires 5 parameters.

1. Mean rate of process
2. Lag 1 correlation
3. Hurst parameter
4. Number of IPPs
5. Number of time-scales to be modelled

Some remarks regarding the fitting procedure

- The accuracy of the Taylor approximations of the IPP variance and covariance functions are of course dependent on $t$ satisfying $t(c_{1i} + c_{2i}) << 1$, $i = 1, \ldots, d$. Particularly for the IPP $i = 1$ this does not necessarily hold true which makes the lag 1 correlation inaccurate.

- We use the one step correlation $\rho$ as a surrogate for an absolute variance measure since with the outlined procedure fitting to the correlation structure $\rho \ast k^{-\beta}$ completely determines the fit and hence also the absolute variance.

- When fitting with the algorithm described above it turns out that particularly for $H$ not to small, approx. $H > 0.8$, that the correlation function fit to $\rho \ast k^{-\beta}$ can be improved by allowing the $d$th IPPs modulating parameters, $c_{1d} = c_{2d}$, to be altered in an additional least squares fit.

- As mentioned in the last sentence of step 2 it is possible that a number, say $d_0$, $0 \leq d_0 < d$, of the $\phi_j$ is set to 0. This occurs when the Hurst parameter is relatively high, approx. $H > 0.8$. The number of “active” IPPs in the fitted model is thus given by $d - d_0$ and the corresponding MAP model has $2^{d-d_0}$ states. One of the input parameters to the fitting algorithm is the number of IPPs to be used. In our implementation we interpret this parameter as the number of “active” IPPs $d^*$ to be used. We always start the fitting procedure with $d = d^*$. If it turns out that $d_0 = 0$ the fitting procedure can be completed in one iteration. If $d_0 > 0$ we increment $d$ by one and carry out steps 1 - 4 again. This last procedure is continued until $d - d_0 = d^*$. This simple procedure has worked in all the cases we have examined.
4 Numerical Results

In this section we will present numerical results obtained from our fitting algorithm which are representative of the entire range of Hurst parameters i.e. we present results for $H = 0.60$, $H = 0.75$ respectively $H = 0.90$. We present two different sets of fits. First we fit to imitate self-similar behaviour over 4 time-scales using 3 IPPs resulting in a simple MAP model with 8 states. Secondly we fit to imitate self-similar behaviour over 5 time-scales using 4 IPPs resulting in a simple MAP model with 16 states. Both sets of fits are obtained for all three representative Hurst parameters. In all 6 cases a mean rate of 3.0 and a one step correlation of 0.4 has be chosen.

The three figures 3, 4 and 5 show the three different fits with 8 state models to the asymptotic autocorrelation function $\rho \cdot k^{-\beta}$ in a log-log plot. It is clearly seen from all three curves that it is possible to imitate a certain Hurst parameter within a quite narrow margin over 4 time scales using a fitted 8 state MAP model with a simple structure. The Index of Dispersion of Counts (IDC) curves of the three models are jointly displayed in figure 6.

The three figures 7, 8 and 9 show the three different fits with 16 state models to the asymptotic autocorrelation function $\rho \cdot k^{-\beta}$ in a log-log plot. It is clearly seen from all three curves that it is possible to imitate a certain Hurst parameter within a quite narrow margin over 5 time scales using a fitted 16 state MAP model with a simple structure. The Index of Dispersion of Counts (IDC) curves of the three models are jointly displayed in figure 10.
Figures 8, 9, 10, 11, 12: Various autocorrelation and time dynamic behaviour plots for 16 state MAPs with different Hurst parameters.

Plots correspond to similar plots in e.g. [18] and [29] for real LAN traffic data. In particular for the higher Hurst parameters our plots seem at least as "visually" self-similar as the ones presented in [18] and [29].

Here it is important to recall what self-similarity implies. Let \( X_k, k = 1, 2, 3, \ldots \) denote a stochastic process with a mean and a variance. Let \( mX_k \) denote the aggregate stochastic process \( mX_k = X_{m(k-1)+1} + X_{m(k-1)+2} + \ldots + X_{mk} \). Then exact self-similarity implies that \( mX \) and \( m^H X \) have the same finite dimensional distributions for \( k = 1, 2, 3, \ldots \). For an exact second order self-similar process this implies that \( mX \) and \( m^H X \) have the same variance and autocorrelation. When considering figures 11 - 13 it is thus not surprising that figure 13 seems "visually" the most self-similar since \( E\{mX\} = E\{mX\} \) while \( \text{Var}\{mX\} \)
The queueing implications of the fitted models on both infinite and finite queues are of major interest. Preliminary work in this area has been reported in [1] and [2], indicating queueing behaviour similar to what has been demonstrated with simulated traces of LAN-measurements [8].

It is well known, however, that the second order properties of counts for the arrival process are not sufficient for predicting queueing performance, see e.g. [2], [5], [10] and [24]. In this respect it is still not sufficient to be able to detect and model second order self-similarity of the counting process. It is possible that low order descriptors of the counting process e.g. IDC or the correlation structure in conjunction with other measures like e.g. the inter-arrival time correlation structure or the IDI [28], Index of Dispersion of Intervals will suffice. Another possibility is to include higher order properties of counts such as the bispectrum which has been suggested in e.g. [25].

In [5] it is shown how a continuum of 2 state MMPPs with the same first and second order properties of the counting process can be constructed. In our models this approach can be used to fit to additional descriptors while maintaining the first and second order properties of the counting process. Applying this approach we have in [3] tentatively tried to compare the queueing behaviour of our models with that of publicly available Bellcore LAN traces. Besides matching first and second order properties of the counting process in [3] we have additionally attempted to match the inter-arrival time correlations. While the results in [3] provide for a certain optimism it is clear that a more elaborate fitting procedure to e.g. the inter-arrival time correlation structure is desirable.

We are currently looking further into these issues, results will be reported elsewhere.

Acknowledgements

The first author would like to acknowledge the financial support from the Technical Research Council of Denmark under grant no. (9500947). This work was partly performed while the second author was on sabbatical leave at Bell Communications Research - Bellcore. He is very thankful for the support and hospitality provided by Bellcore and for the support by the Technical Research Council of Denmark under grant no. (9400273).

References


