Multiscale Fitting Procedure using Markov Modulated Poisson Processes

Paulo Salvador (salvador@av.it.pt)
University of Aveiro / Institute of Telecommunications, Aveiro, Portugal

Rui Valadas (rv@det.ua.pt)
University of Aveiro / Institute of Telecommunications, Aveiro, Portugal

António Pacheco (apacheco@math.ist.utl.pt)
Instituto Superior Técnico - Technical University of Lisbon, Mathematics Department, CEMAT and CLC, Lisboa, Portugal.

Received December 2001; in final form January 2003

Abstract. This paper proposes a parameter fitting procedure using Markov Modulated Poisson Processes (MMPPs) that leads to accurate estimates of queuing behavior for network traffic exhibiting LRD behavior. The procedure matches both the autocovariance and marginal distribution of the counting process. A major feature is that the number of states is not fixed a priori, and can be adapted to the particular trace being modeled. The MMPP is constructed as a superposition of L 2-MMPPs and one M-MMPP. The 2-MMPPs are designed to match the autocovariance and the M-MMPP to match the marginal distribution. Each 2-MMPP models a specific time-scale of the data. The procedure starts by approximating the autocovariance by a weighted sum of exponential functions that model the autocovariance of the 2-MMPPs. The autocovariance tail can be adjusted to capture the long-range dependence characteristics of the traffic, up to the time-scales of interest to the system under study. The procedure then fits the M-MMPP parameters in order to match the marginal distribution, within the constraints imposed by the autocovariance matching. The number of states is also determined as part of this step. The final MMPP with $M2^L$ states is obtained by superposing the L 2-MMPPs and the M-MMPP. We apply the inference procedure to traffic traces exhibiting long-range dependence and evaluate its queuing behavior through simulation. Very good results are obtained, both in terms of queuing behavior and number of states, for the traces used, which include the well-known Bellcore traces.

Keywords: Traffic modeling, autocorrelation, self-similar, long-range dependence, MMPP.

1. Introduction

Since the work by Leland *et al.* [1] several studies have shown that network traffic may exhibit properties of self-similarity and long-range dependence (LRD) [1–7]. These characteristics have significant impact on network performance. However, as pointed out in [8,9], matching the LRD is only required within the time-scales of interest to the system under study. For example, in order to analyze the queuing behavior, the selected traffic model needs only to capture the correlation structure of the source up to the so-called critical time-scale or correlation horizon, which is directly related to the maximum buffer size. One of the consequences of this result is that more traditional traffic

models such as Markov Modulated Poisson Processes (MMPPs) can still be used to model traffic exhibiting long-range dependence [10–13]. Moreover, providing a good match of the LRD behavior (through an accurate fitting of the autocovariance tail) is not enough for accurate prediction of the queuing behavior. For example, the work in [14] discusses the limitations of using only the mean and the autocorrelation function as statistical descriptors of the input process for the purpose of analyzing queuing performance. The authors show that the mean queue length can vary substantially when the parameters of the input process are varied, subject to the same mean and autocorrelation function. Thus, in general, accurate prediction of queuing behavior requires detailed modeling of the first-order statistics, not just the mean. While this issue is of major importance, we believe it remains largely neglected. The limitations of matching only the LRD are also in line with recent suggestions that multifractal models [15–17], or eventually hybrid models combining multifractal and self-similar behavior [18], may be more appropriate than self-similar ones for certain types of network traffic as they allow for incorporating time-dependent scaling laws.

The main goal of the present work is to develop a parameter fitting procedure using Markov Modulated Poisson Processes (MMPPs) that leads to accurate estimates of queuing behavior for network traffic exhibiting LRD behavior. In order to achieve this goal, the procedure matches closely both the autocovariance and the marginal distribution of the counting process. The consideration of MMPPs is motivated by the availability of theoretical results for analyzing the queuing behavior [19] and for determining effective bandwidths [20,21]. The work also seeks minimizing the number of states of the underlying model so as to reduce the complexity associated with the calculation of the performance metrics of interest. MMPPs have deserved considerable attention in the literature (see, e.g., [10–13,22–31]).

Matching simultaneously the autocovariance and the marginal distribution is a difficult task since every MMPP parameter has an influence on both characteristics. With the purpose of achieving some degree of decoupling when matching these two statistics, we construct the MMPP as a superposition of two MMPPs, where one MMPP (with 2^L states) is used to adjust the autocovariance and the other (with M states) is used to adjust the marginal distribution, taking into account the contribution of the first MMPP. We will denote the resulting process as M2^L-MMPP. The 2^L-MMPP matching the autocovariance is a superposition of L 2-MMPPs. Given that the autocovariance of a 2-MMPP is a single exponential, this approach allows matching the empirical autocovariance based on prior approximation by a weighted sum of exponentials, which results in a simple and accurate procedure. The M-MMPP matching the marginal distribution is forced to have null autocovariance at positive lags, to assure that the autocovariance of the M2^L-MMPP equals that of the superposition of the L 2-MMPPs; the marginal distribu-

tion of the M-MMPP is obtained through deconvolution of the L 2-MMPPs and M2^L-MMPP marginal distributions, thus ensuring that the contribution of the L 2-MMPPs is taken into account. The autocovariance modeling is such that each 2-MMPP (in the set of L 2-MMPPs) models a specific time-scale. Thus, in this work, the concept of time-scale is defined in the context of second-order statistics: each time-scale is associated with a characteristic time constant of the autocovariance function. An important feature of the procedure is that both L and M need not to be defined a priori, since they are determined as part of the procedure.

We note that, in order to boost the computational efficiency of the fitting procedure, the 2-MMPPs used to fit the autocovariance function are interrupted Poisson processes, and that the value of M of the M-MMPP is chosen as the smallest value that provides a given degree of matching between the data and the marginal probability function of the fitted model. Moreover, the assumed independence between the M-MMPP and the L 2-MMPPs introduces further constraints on the parametric form of the fitted $M2^L$-MMPP.

When measuring network traffic data, we can either record the individual arrival instants or the number of arrivals in a predefined sampling (time) interval. The former approach brings more detail but the latter one has the advantage of producing a fixed amount of data that is known in advance. This allows the recording of longer traces, which clearly pays off the loss of detail in data recording, if the sampling interval is chosen appropriately. In this work, we consider discrete time MMPPs (dMMPPs) instead of continuous time MMPPs, since they are more natural models for data corresponding to the number of arrivals in a sampling interval. We note that discrete time and continuous time MMPPs are basically interchangeable (through a simple parameter rescaling) as models for arrival processes, whenever the sampling interval used for the discrete time version is small compared with the average sojourn times in the states of the modulating Markov chain.

The fitting procedure starts by approximating the autocovariance by a weighted sum of exponential functions. As part of this step, the relevant time-scales of the data are identified. After this, the procedure fits the M-dMMPP parameters in order to match the probability function, within the constraints imposed by the autocovariance matching. The final M2^L-dMMPP is obtained by superposing the L 2-dMMPPs and the M-dMMPP.

We apply the fitting procedure to traffic traces exhibiting LRD, including the well known, publicly available, Bellcore traces. The LRD characteristics are analyzed using the wavelet based estimator of [32]. Results show that the MMPPs obtained through the fitting procedure are capable of modeling the LRD behavior present in data. The fitting procedure is also assessed in terms of queuing behavior. Results show a very good agreement between the packet

loss ratio obtained with the original data traces and with traces generated from the fitted MMPPs.

This paper is organized as follows. Section 2 gives some background on discrete MMPPs. In section 3 we describe the fitting procedure. Section 4 presents the numerical results, which include applying the fitting procedure to measured traffic traces. In section 5 our work is compared with previously published ones. Finally, in section 6 we conclude the paper.

2. Background

The discrete time Markov Modulated Poisson Process (dMMPP) is the discrete time version of the popular (continuous time) MMPP and may be regarded as a Markov random walk where the increments in each instant have a Poisson distribution whose parameter is a function of the state of the modulator Markov chain. More precisely, the (homogeneous) Markov chain $(X, J) = \{(X_k, J_k),\ k = 0, 1, \ldots\}$ with state space $\mathbb{N}_0 \times S$ is a dMMPP if and only if for $k = 0, 1, \ldots$,

$$P(X_{k+1} = m,\ J_{k+1} = j | X_k = n,\ J_k = i) = \\ = \begin{cases} 0 & m < n \\ p_{ij}\, e^{-\lambda_i} \frac{\lambda_i^{m-n}}{(m-n)!} & m \geq n \end{cases} \tag{1}$$

for all $m, n \in \mathbb{N}_0$ and $i, j \in S$, with λ_i, $i \in S$, being nonnegative real constants and $\mathbf{P} = (p_{ij})$ being a irreducible stochastic matrix. Note that the distribution of $X_{k+1} - X_k$ given $J_k = j$ is Poisson with mean λ_j, so that λ_j represents the mean increment of the process X when the modulating Markov chain is in state j. The dMMPP is a particular case of the dBMAP (discrete time batch Markovian arrival process proposed by Blondia and Casals [33] and that has received some attention (see, e.g., [34–37] and references therein). In the dBMAP the distribution of the increments of the process at each time instant may depend on the state visited at the previous instant in addition to the current state.

Whenever (1) holds, we say that (X, J) is a dMMPP with set of modulating states S and parameter (matrices) $\mathbf{P}$ and $\mathbf{\Lambda}$, and write

$$(X, J) \sim \text{dMMPP}_S(\mathbf{P}, \mathbf{\Lambda}) \tag{2}$$

where $\mathbf{\Lambda} = (\lambda_{ij}) = (\lambda_i \delta_{ij})$. The matrix $\mathbf{P}$ is the transition probability matrix of the modulating Markov chain J, whereas $\mathbf{\Lambda}$ is the matrix of Poisson arrival rates. If S has cardinality r, we say that (X, J) is a dMMPP of order r (r-

dMMPP). When, in particular, $S = \{1, 2, \ldots, r\}$ for some $r \in I\!N$, then

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \ldots & p_{1r} \\ p_{21} & p_{22} & \ldots & p_{2r} \\ \ldots & \ldots & \ldots & \ldots \\ p_{r1} & p_{r2} & \ldots & p_{rr} \end{bmatrix} \text{ and } \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_r \end{bmatrix} \tag{3}$$

and we write simply that $(X, J) \sim \text{dMMPP}_r(\mathbf{P}, \mathbf{\Lambda})$.

We will consider the superposition of (independent) L 2-dMMPPs

$$(X^{(l)}, J^{(l)}) \sim \text{dMMPP}_2(\mathbf{P}^{(l)}, \mathbf{\Lambda}^{(l)}), l = 1, 2, \ldots, L \tag{4}$$

and one M-dMMPP

$$(X^{(L+1)}, J^{(L+1)}) \sim \text{dMMPP}_M(\mathbf{P}^{(L+1)}, \mathbf{\Lambda}^{(L+1)}). \tag{5}$$

Note that, in particular, for $l = 1, 2, \ldots, L$,

$$\mathbf{P}^{(l)} = \begin{bmatrix} p_{11}^{(l)} & p_{12}^{(l)} \\ p_{21}^{(l)} & p_{22}^{(l)} \end{bmatrix}, \qquad \mathbf{\Lambda}^{(l)} = \begin{bmatrix} \lambda_1^{(l)} & 0 \\ 0 & \lambda_2^{(l)} \end{bmatrix} \tag{6}$$

and we assume that $p_{12}^{(l)} + p_{21}^{(l)} < 1$. In addition, we consider $J^{(1)}, J^{(2)}$, $\ldots, J^{(L+1)}$ to be ergodic chains in steady-state. For $l = 1, 2, \ldots, L$ we denote by $\pi^{(l)} = \begin{bmatrix} \pi_1^{(l)} \pi_2^{(l)} \end{bmatrix}$ the stationary distribution of $J^{(l)}$. Similarly, we denote by $\pi^{(L+1)} = [\pi_1^{(L+1)} \pi_2^{(L+1)} \ \ldots \ \pi_M^{(L+1)}]$ the stationary distribution of $J^{(L+1)}$.

The result of the superposition is the process

$$(X, J) = \left(\sum_{l=1}^{L+1} X^{(l)}, (J^{(1)}, J^{(2)}, \ldots, J^{(L+1)}) \right) \sim \text{dMMPP}_S(\mathbf{P}, \mathbf{\Lambda}) \tag{7}$$

where

$$S = \{1, 2\}^L \times \{1, 2, \ldots, M\} \tag{8}$$

$$\mathbf{P} = \mathbf{P}^{(1)} \otimes \mathbf{P}^{(2)} \otimes \ldots \otimes \mathbf{P}^{(L+1)} \tag{9}$$

$$\mathbf{\Lambda} = \mathbf{\Lambda}^{(1)} \oplus \mathbf{\Lambda}^{(2)} \oplus \ldots \oplus \mathbf{\Lambda}^{(L+1)} \tag{10}$$

with $\oplus$ and $\otimes$ denoting the Kronecker sum and the Kronecker product, respectively. Note that Markov chain J is also in steady-state. We refer to (X, J) as being the $M2^L$-dMMPP. The superposition of an M-dMMPP and L 2-dMMPPs is depicted in Figure 1.

In our approach L and M are not fixed *a priori* but instead are computed as part of the fitting procedure. However, in the rest of this section they

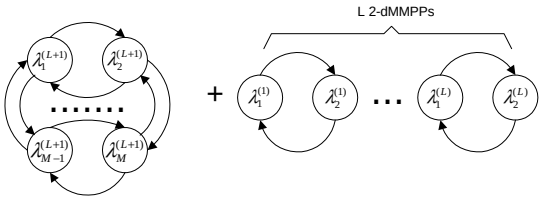

Figure 1. Superposition of an M-dMMPP and L 2-dMMPP models.

may be thought as being fixed. We want the L 2-dMMPPs to capture the autocovariance function of the increments of the arrival process (X). Also, discounting for the effect of the L 2-dMMPPs, we want the M-dMMPP to approximate the distribution of the increments of the arrival process. To explain how this may be accomplished, it is convenient to define the increment processes associated to $X^{(1)}, X^{(2)}, \ldots, X^{(L+1)}$, and X, which we denote by $Y^{(1)}, Y^{(2)}, \ldots, Y^{(L+1)}$, and Y, respectively. Thus,

$$Y_k^{(l)} = X_{k+1}^{(l)} - X_k^{(l)},\ l = 1, 2, \ldots, L+1 \tag{11}$$

and

$$Y_k = X_{k+1} - X_k \tag{12}$$

for $k = 0, 1, \ldots$. Note that Y_k is the (total) number of arrivals at sampling interval k and $Y_k^{(l)}$ is the number of arrivals that are due to the l-th arrival process, so that, in particular,

$$Y_k = \sum_{l=1}^{L+1} Y_k^{(l)}, \quad k = 0, 1, 2, \ldots. \tag{13}$$

Moreover $Y^{(1)}, Y^{(2)}, \ldots, Y^{(L+1)}$, and Y, are stationary sequences.

In order to characterize the marginal distributions of the processes $Y^{(1)}$, $Y^{(2)}, \ldots, Y^{(L+1)}$, and Y, we denote, respectively, by $\{f_l(k),\ k = 0, 1, 2, \ldots\}$, $l = 1, 2, \ldots, L+1$, and $\{f(k),\ k = 0, 1, 2, \ldots\}$, their (marginal) probability functions. As the univariate distributions of $Y^{(1)}$, $Y^{(2)}, \ldots, Y^{(L+1)}$ are mixtures of Poisson distributions, we denote the probability function of a Poisson random variable with mean μ by $\{g_\mu(k),\ k = 0, 1, 2, \ldots\}$, for $\mu \in [0, +\infty)$, so that

$$g_\mu(k) = e^{-\mu}\,\frac{\mu^k}{k!}, \quad k = 0, 1, 2, \ldots. \tag{14}$$

For $l = 1, 2 \ldots, L$, the marginal distribution of $Y^{(l)}$ (that is, the distribution of $Y_k^{(l)}$, for $k = 0, 1, \ldots$) is a mixture of two Poisson distributions with means $\lambda_1^{(l)}$ and $\lambda_2^{(l)}$ and weights $\pi_1^{(l)}$ and $\pi_2^{(l)}$, respectively. Thus the

probability functions of $Y^{(l)}$, $l = 1, 2, \ldots, L$, are given by

$$f_l(k) = \pi_1^{(l)}\, g_{\lambda_1^{(l)}}(k) + \pi_2^{(l)}\, g_{\lambda_2^{(l)}}(k), k = 0, 1, 2, \ldots \tag{15}$$

and their autocovariance functions are

$$\gamma_k^{(l)} = \text{ Cov}\,(Y_0^{(l)}, Y_k^{(l)}) = \pi_1^{(l)}\,\pi_2^{(l)}\,|\lambda_2^{(l)} - \lambda_1^{(l)}|^2\, e^{kc_l}, \quad k = 0, 1, 2, \ldots \tag{16}$$

where $c_l = \ln(1 - p_{12}^{(l)} - p_{21}^{(l)})$. Note that, in particular, the autocovariance functions of $Y^{(1)}, Y^{(2)}, \ldots, Y^{(L)}$ exhibit an exponential decay to zero.

As we want the M-dMMPP to approximate the distribution of the increments of the arrival process but to have no contribution to the autocovariance function of the increments of the M2^L-dMMPP, we choose to make $J^{(L+1)}$ a Markov chain with no memory whatsoever. This is accomplished by choosing

$$\mathbf{P}^{(L+1)} = \begin{bmatrix} \pi_1^{(L+1)} & \pi_2^{(L+1)} & \ldots & \pi_M^{(L+1)} \\ \pi_1^{(L+1)} & \pi_2^{(L+1)} & \ldots & \pi_M^{(L+1)} \\ \ldots & \ldots & \ldots & \ldots \\ \pi_1^{(L+1)} & \pi_2^{(L+1)} & \ldots & \pi_M^{(L+1)} \end{bmatrix}. \tag{17}$$

Note that this implies that $Y^{(L+1)}$ is an independent and identically distributed sequence of random variables whose distribution is a mixture of M Poisson random variables with means $\lambda_i^{(L+1)}$ and weights π_i^{L+1}, for $i = 1, 2, \ldots, M$. As a consequence, the probability function of $Y^{(L+1)}$ is given by

$$f_{L+1}(k) = \sum_{j=1}^{M} \pi_j^{(L+1)}\, g_{\lambda_j^{(L+1)}}(k),\ k = 0, 1, 2, \ldots \tag{18}$$

and the autocovariance function of $Y^{(L+1)}$ is null for all positive lags; i.e.,

$$\gamma_k^{(L+1)} = \text{ Cov}\,(Y_0^{(L+1)}, Y_k^{(L+1)}) = 0,\ k \geq 1. \tag{19}$$

Taking into account (13), it follows that the probability function of Y is given by:

$$f(k) = (f_1 \oplus f_2 \oplus \ldots \oplus f_{L+1})\,(k) \tag{20}$$

where $\oplus$ denotes the convolution of probability functions. Y is a sequence of random variables whose distribution is a mixture of Poisson random variables (note that the sum of independent mixtures of Poisson random variables is also a mixture of Poisson random variables), and the probability function of Y may be written in the following way

$$f(k) = \sum_{j_1=1}^{2} \sum_{j_2=1}^{2} \ldots \sum_{j_L=1}^{2} \sum_{j_{L+1}=1}^{M} \left(\prod_{l=1}^{L+1} \pi_{j_l}^{(l)} \right) g_{\sum_{l=1}^{L+1} \lambda_{j_l}^{(l)}}(k). \tag{21}$$

Moreover, from (13) and taking into account (16) and (19), we conclude that the autocovariance function of Y is given by

$$\begin{aligned}\gamma_k = \mathrm{Cov}\left(Y_0, Y_k\right) &= \sum_{l=1}^{L+1} \mathrm{Cov}\left(Y_0^{(l)}, Y_k^{(l)}\right) \\ &= \textstyle\sum_{l=1}^{L} \pi_1^{(l)}\, \pi_2^{(l)}\, |\lambda_2^{(l)} - \lambda_1^{(l)}|^2\, e^{kc_l}\end{aligned} \tag{22}$$

for $k = 1, 2, \ldots$.

3. Inference Procedure

In the rest of the paper we will refer to $Y^{(1)}, Y^{(2)}, \ldots, Y^{(L)}$ as the 2-dMMPPs, to $Y^{(L+1)}$ as the M-dMMPP, and to Y as the M2^L-dMMPP.

The inference procedure can be divided in four steps: (i) approximation of the empirical autocovariance by a weighted sum of exponentials and identification of time-scales, (ii) inference of the M-dMMPP probability function and of the 2-dMMPPs parameters, (iii) inference of the M-dMMPP Poisson arrival rates and transition probabilities and (iv) calculation of the final M2^L-dMMPP parameters. The flow diagram of this procedure is represented in Figure 2. In the following subsections we describe these steps in detail.

3.1. Autocovariance approximation and time-scales identification

Our approach is to approximate the autocovariance by a large number of exponentials and then aggregate exponentials with a similar decay into the same time-scale. This is close to the approaches considered in [10,13,38]. As a first step, we approximate the empirical autocovariance by a sum of K exponentials with real positive weights and negative real time constants. We chose K as $\sqrt{k_{max}}$, where k_{max} represents the number of points of the empirical autocovariance. This is accomplished through a modified Prony algorithm [39]. The Prony algorithm returns two vectors,

$$\vec{a} = [\ a_1\ \ \ldots\ \ a_K\ ] \qquad \vec{b} = [\ b_1\ \ \ldots\ \ b_K\ ]$$

which correspond to the approximating function

$$C_K\left(\vec{a}, \vec{b}\right) = \sum_{i=1}^{K} a_i e^{-b_i k}, \qquad k = 1, 2, 3, \ldots \tag{23}$$

At this point we identify the components of the autocovariance that characterize the different time-scales. We define L different time-scales, in which the autocovariance decays, $b_i, i = 1, \ldots, K$, fall in the same logarithmic scale.

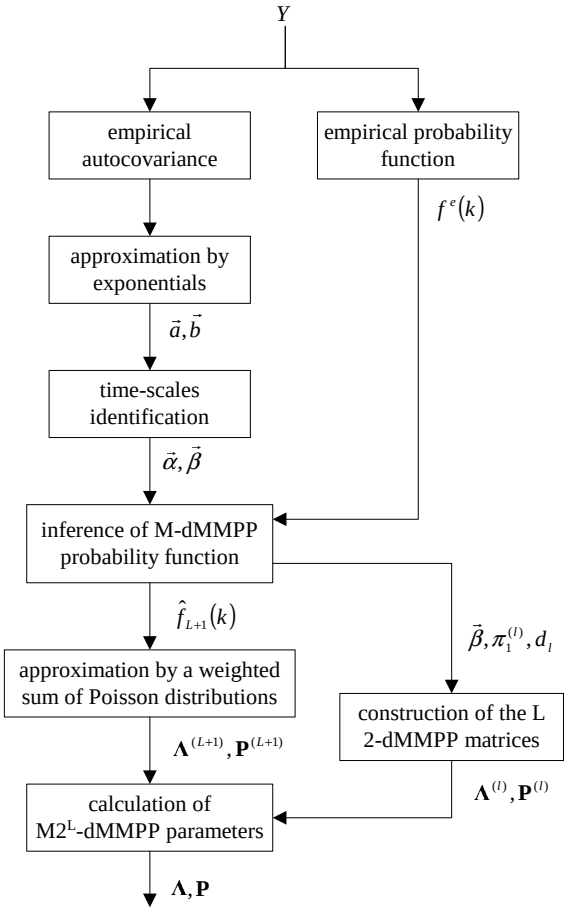

Figure 2. Flow diagram of the inference procedure.

To explain how this is accomplished it is useful to assume that $b_j \leq b_{j+1}, 1 \leq j \leq K-1$, and $\lceil \ldots \rceil$ represents the integer round towards plus infinity. The value L is computed through the following iterative process. Starting with $l = 1$ and $i_l = 1$ compute i_{l+1} through

$$i_{l+1} = \min \left\{ K+1, \inf \left\{ j : i_l < j \leq K \wedge \lceil \log_{10}(b_j) \rceil > \lceil \log_{10}(b_{j-1}) \rceil \right\} \right\}.$$

If $i_{l+1} > K$ then make $L = l$ and stop; otherwise increment l by one and repeat the process. Note that, in particular,

$$\lceil \log_{10}(b_{i_l}) \rceil = \lceil \log_{10}(b_{i_l+1}) \rceil = \ldots = \lceil \log_{10}(b_{i_{l+1}-1}) \rceil$$

but, if $i_{l+1} \leq K$,

$$\lceil \log_{10}(b_{i_l}) \rceil < \lceil \log_{10}(b_{i_{l+1}}) \rceil .$$

For $j = 1, 2, \ldots, L$, we consider that the decays b_{i_l} to $b_{i_{l+1}-1}$ characterize the same traffic time scale and we aggregate the $n_l = |i_{l+1} - i_l|$ components in one component with the following parameters:

$$\alpha_l = \sum_{k=i_l}^{i_{l+1}-1} a_k \qquad \text{and} \qquad \beta_l = -\frac{\sum_{k=i_l}^{i_{l+1}-1} a_k b_k}{\alpha_l}. \tag{24}$$

Taking in account (16), (23) and (24), these parameters are used to fit the autocovariance function of the 2-dMMPP $Y^{(l)}$. Through these relations, it results that

$$\alpha_l = d_l^2 \pi_1^{(l)} \pi_2^{(l)} \qquad \text{and} \qquad \beta_l = c_l \tag{25}$$

where $d_l = |\lambda_2^{(l)} - \lambda_1^{(l)}|$, i.e., the fitted autocovariance function of $Y_1 + Y_2 + \ldots + Y_L$ is

$$\sum_{l=1}^{L} \alpha_l e^{k\beta_l}, \quad k = 1, 2, \ldots. \tag{26}$$

3.2. Inference of the M-dMMPP probability function and of the L 2-dMMPP parameters

The next step is the inference of the M-dMMPP probability function from the empirical probability function of the original data trace. The relation between the probability functions of the 2-dMMPPs, the M-dMMPP and the M2^L-dMMPP is defined by (20).

In order to simplify the deconvolution of $f_{L+1}(k)$ and $f_l(k), l = 1, ..., L$, we consider that the Poisson arrival rate is zero in one state of each 2-dMMPP source; that is, $\lambda_1^{(l)} = 0$ and $\lambda_2^{(l)} = d_l$, for $l = 1, ..., L$. From (25),

$$d_l = \sqrt{\frac{\alpha_l}{\pi_1^{(l)} \pi_2^{(l)}}}, \quad l = 1, 2, \ldots, L. \tag{27}$$

The probability function of the M-dMMPP, f_{L+1}, is inferred from the empirical probability function of the data, denoted by f^e, and the L 2-dMMPP probability functions, denoted by $\hat{f}_l$, $l = 1, 2, \ldots, L$, based on the fitted parameters, after fixing the probabilities $\pi_1^{(l)}$, $l = 1, 2, \ldots, L$, through (20). More precisely, f_{L+1} is fitted jointly with the parameters $\pi_1^{(l)}$, $l = 1, ..., L$, through the following constrained minimization process:

$$\min_{\{\pi_1^{(l)}, l=1,...,L\}, \{f_{L+1}(k), k=0,1,...\}} \sum_k |o^e(k)| \tag{28}$$

where

$$o^e(k) = f^e(k) - \left(\hat{f}_1 \oplus ... \oplus \hat{f}_L \oplus f_{L+1}\right)(k) \tag{29}$$

subject to (25) and

$$\begin{array}{c} 0 < \pi_1^{(l)} < 1,\ l = 1, 2, \ldots, L, \\ f_{L+1}(k) > 0,\ k = 0, 1, \ldots, \\ \text{and} \quad \sum_{k=0}^{+\infty} f_{L+1}(k) = 1. \end{array} \tag{30}$$

We denote by $\hat{f}_{L+1}$ the fitted probability function of the M-dMMPP. Note that $\pi_1^{(l)}$ is not allowed to be 0 or 1 because, in both cases, the l^{th} 2-dMMPP would degenerate into a Poisson process. The constrained minimization process given by (28)–(30) is a non-linear programming problem and in general, it is computationally demanding to obtain the global optimal solution. Accordingly, to solve this problem we consider two approximations: (i) we make $\pi_1^{(l)} = \pi_1^{(l+1)}$, $l = 1, \ldots, L-1$ and (ii) we restrict the range of possible $\pi_1^{(l)}$ solutions to be discrete and such that $\pi_1^{(l)} = 0.001k$, $k = 1, \ldots, 999$. Then a search process is used to find the minimum value of the objective function. These approximations have had negligible impact on the results obtained so far with the fitting procedure, in particular on those presented in section 4.

At this point all parameters of the 2-dMMPPs, $Y^{(1)}, Y^{(2)}, \ldots, Y^{(L)}$, have been determined and their corresponding 2-dMMPP matrices

$$\{(\mathbf{P}^{(l)}, \mathbf{\Lambda}^{(l)}), l = 1, 2, \ldots, L\}$$

can be constructed in the following way:

$$\mathbf{P}^{(l)} = \begin{bmatrix} 1 - \pi_2^{(l)}(1 - e^{\beta_l}) & \pi_2^{(l)}(1 - e^{\beta_l}) \\ \pi_1^{(l)}(1 - e^{\beta_l}) & 1 - \pi_1^{(l)}(1 - e^{\beta_l}) \end{bmatrix}$$

and

$$\mathbf{\Lambda}^{(l)} = \begin{bmatrix} 0 & 0 \\ 0 & d_l \end{bmatrix}.$$

3.3. Inference of M-dMMPP parameters

The next step is the inference of the number of states and Poisson arrival rates of the M-dMMPP from $\hat{f}_{L+1}$. To do this we infer $\hat{f}_{L+1}$ as a weighted sum of Poisson probability functions, i.e., as the probability function of a finite Poisson mixture with an unknown number of components.

The matching is carried out through an algorithm that progressively subtracts a Poisson probability function from $\hat{f}_{L+1}$, which is described in the flowchart of Figure 3. We represent the i^{th} Poisson probability function with mean φ_i by $g_{\varphi_i}(k)$. We define $h^{(i)}(k)$ as the difference between $\hat{f}_{L+1}(k)$ and the weighted sum of Poisson probability functions at the i^{th} iteration. Initially, we set $h^{(1)}(k) = \hat{f}_{L+1}(k)$. In each step, we first detect the maximum of $h^{(i)}(k)$. The corresponding k-value, $\varphi_i = [h^{(i)}]^{-1}\left(\max h^{(i)}(k)\right)$,

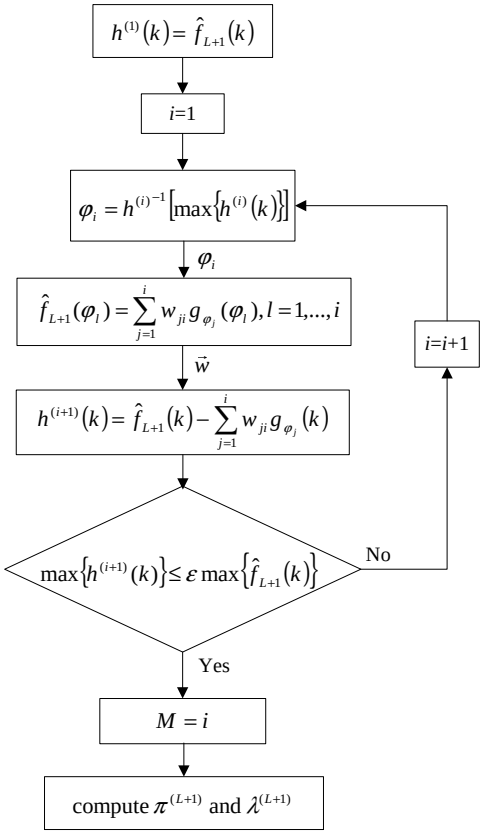

Figure 3. Algorithm for calculation of the number of states and Poisson arrival rates of the M-dMMPP.

will be considered the i^{th} Poisson rate of the M-dMMPP. We then calculate the weights of each Poisson probability function, $\vec{w}_i = [w_{1i}, w_{2i}, ..., w_{ii}]$, through the following set of linear equations:

$$\hat{f}_{L+1}(\varphi_l) = \sum_{j=1}^{i} w_{ji} g_{\varphi_j}(\varphi_l), \qquad l = 1, ..., i.$$

This assures that the fitting between $\hat{f}_{L+1}(k)$ and the weighted sum of Poisson probability functions is exact at φ_l points, for $l = 1, 2, \ldots, i$. The final step in each iteration is the calculation of the new difference function

$$h^{(i)}(k) = \hat{f}_{L+1}(k) - \sum_{j=1}^{i} w_{ji} g_{\varphi_j}(k).$$

The algorithm stops when the maximum of $h^{(i)}(k)$ is lower than a pre-defined percentage of the maximum of $\hat{f}_{L+1}(k)$ and M is made equal to i.

Other methods for parameter estimation of finite Poisson mixtures with an unknown number of components, have been proposed, e.g., based on moment estimation or maximum likelihood (see, e.g., [40] and references therein). These methods lead to the solution of a system of nonlinear equations with a number of variables equal to twice the number of components of the Poisson mixture model. By contrast, in the algorithm described above the rates of the Poisson mixture model are fitted directly and the weights are fitted by solving a system of linear equations with the weights as variables, thus obtaining a more computationally effective procedure. Moreover, the proposed procedure leads to the fitting of a small number of components for the fitted IP traces. Note also that the proposed procedure is in line with the global heuristic approach followed in the paper, which seeks a fast method to fit a particular class of MMPPs instead of a purely statistical fitting of an MMPP.

After M has been determined, the parameters of the M-dMMPP, $\{(\pi_j^{(L+1)}, \lambda_j^{(L+1)}), j = 1, 2, \ldots, M\}$, are then set equal to

$$\pi_j^{(L+1)} = w_{jM} \qquad \text{and} \qquad \lambda_j^{(L+1)} = \varphi_j.$$

3.4. M2^L-dMMPP MODEL CONSTRUCTION

Finally, the M2^L-dMMPP process can be constructed using the equations (9) and (10), where $\mathbf{\Lambda}^{(L+1)}$, $\mathbf{P}^{(L+1)}$, $\mathbf{\Lambda}^{(i)}$ and $\mathbf{P}^{(i)}$, $i = 1, ..., L$, were calculated in the last two subsections.

4. Numerical Results

We apply our fitting procedure to several traffic traces: (i) the publicly available Bellcore LAN traces [1] and (ii) a set of traces of IP traffic measured at the University of Aveiro (UA). The UA traces are representative of Internet access traffic produced within a University campus environment, and span over five different day periods. The UA network provides connection between the 29 campus buildings, and includes three backbone technologies: (i) an old 10Base5 Ethernet, (ii) an optical fiber FDDI ring and (iii) an optical fiber ATM network. The interconnection is provided by 15 IP/IPX routers spread over the campus. The network allows asynchronous accesses from the exterior, using modems (analogue/ISDN), which provides Internet access to external users. The UA network is connected to the Internet through an ATM connection at 10 Mbits/sec. The measurements were carried out in a 100 Mb/s Ethernet LAN on the back of the Internet access router. The main characteristics of the UA traces are described in Table I. Each UA trace corresponds

to a different day period. All measurements captured 20 million packets. However, to assure stationarity, traces UA6b and UA10b were truncated to 10 million packets. The traffic analyzer was a 1.2 GHz AMD Athlon PC, with 1.5 Gbytes of RAM, running WinDump. The measurements recorded the arrival instant and the IP header of each packet. The UA traces and the fitting software, which is based on MATLAB, are available at http://www.av.it.pt.

We assess the fitting procedure by comparing the empirical probability and autocovariance functions of the original data traces and the theoretical ones corresponding to the fitted dMMPPs.

We analyze the presence of LRD behavior, in both the original and the fitted data traces, using the method described in [32]. This method resorts to the so-called Logscale Diagram which consists in the graph of y_j against j, together with confidence intervals about the y_j, where y_j is a function of the wavelet discrete transform coefficients at scale j. Traffic is said to be LRD if, within the limits of the confidence intervals, the y_j fall on a straight line, in a range of scales from some initial value j_1 up to the largest one present in data.

Table II summarizes some of the characteristics and results of the fitting process. For each trace, it indicates the number of states of the fitted dMMPP, the stationary probabilities of 2-MMPPs' first state (π_1), the number of 2-dMMPPs modeling the autocovariance (L), the computational efficiency measured both in terms of time and floating point operations (FLOPS), obtained with MATLAB, and the errors relative to the fitting of the probability function (PF) and of the autocovariance function (AF). The fitting error is defined in terms of the Inequality Coefficient (IC). The results indicate that a close match was obtained in all cases.

We also analyze the queuing behavior by comparing the packet loss ratio, obtained through trace-driven simulation, using the original data traces and simulated traces obtained from the fitted dMMPPs. To calculate the packet loss ratio, we assume a fixed packet size equal to the mean packet size. The sampling interval of the counting process was 0.1 seconds for all traces. In each case, results are shown for two average link utilization values.

To determine if the fitted dMMPPs are able to capture the LRD behavior up to the time-scales of interest, we calculate the so-called correlation horizon (CH), using the method of [8]. The CH is the autocovariance lag that separates relevant and irrelevant time-scales for the purpose of assessing queuing behavior, and is a function of the input data characteristics and of the system parameters.

Table I. Main characteristics of measured traces.

Trace name	Capture date	Capture interval	Trace size (pkts)	Mean rate (pkts/sec)	Mean pkt size (bytes)
UA17	Thu, June 7^{th} 2001	2.40pm to 6.10pm	20 millions	1114	536
UA1b	Tue, July 3^{th} 2001	8.00pm to 3.15am	20 millions	766	598
UA6b	Thu, July 5^{th} 2001	3.58am to 9.11am	10 millions	533	692
UA10b	Fri, July 6^{th} 2001	12.41pm to 3.16pm	10 millions	1074	600
UA19b	Tue, July 10^{th} 2001	10.15am to 3.08pm	20 millions	1138	557

Table II. Fitting results for the packet arrival process.

Trace name	Inferred model	π_1	L	MFLOPS	Fitting time (sec)	PF IC (%)	AF IC (%)
pAug	56-dMMPP	0.1	3	95.35	38.25	7.25	29.05
pOct	28-dMMPP	0.4	2	28.41	50.67	9.21	7.37
UA17	12-dMMPP	0.2	2	78.24	14.49	7.74	8.24
UA1b	24-dMMPP	0.1	3	74.33	15.12	10.02	7.31
UA6b	12-dMMPP	0.2	2	67.99	12.52	13.62	5.70
UA10b	16-dMMPP	0.1	2	130.06	21.93	9.70	17.88
UA19b	12-dMMPP	0.2	2	95.51	13.72	10.75	5.23

4.1. BELLCORE TRACES

The fitting procedure was applied to the Bellcore traces pOct.TL and pAug.TL, both with 1 million samples. These traces were fitted to a 28-dMMPP and a 56-dMMPP respectively.

The pOct.TL trace autocovariance was fitted by two exponential functions with parameters $\vec{\alpha} = [185.9 \quad 180.8]$ and $\vec{\beta} = [-3.51 \times 10^{-2} \quad -3.27 \times 10^{-4}]$, and the pAug.TL trace autocovariance by three exponentials with parameters $\vec{\alpha} = [75.3 \quad 91.57 \quad 43.78]$ and $\vec{\beta} = [-8.43 \times 10^{-1} \quad -9.15 \times 10^{-2} \quad -1.2 \times 10^{-3}]$. Thus, in these traces, the number of identified time-scales was two ($L = 2$) and three ($L = 3$), respectively.

Figure 4 shows the fitting results for the probability function (pOct.TL trace). In this case, the fitting was performed with a very small approximation error. From Figure 5 it can be seen that the autocovariance of the fitted model is able to reproduce the average behavior of the empirical autocovariance (but not its oscillatory behavior).

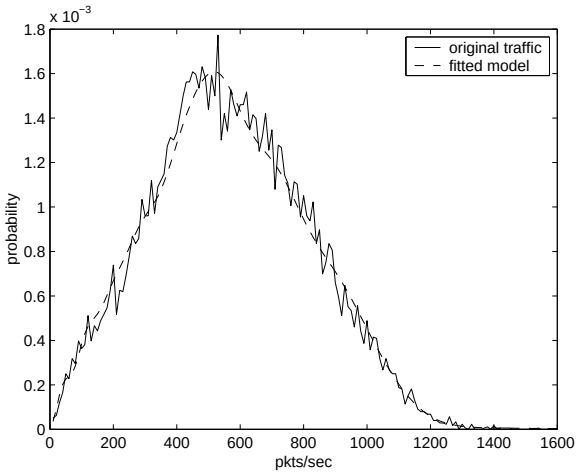

Figure 4. Probability function, pOct.TL.

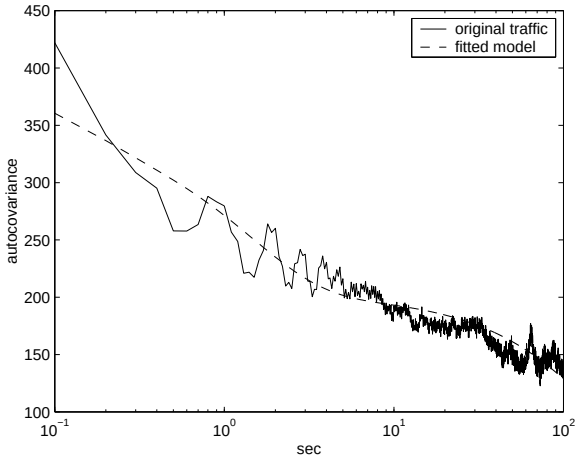

Figure 5. Autocovariance, pOct.TL.

In order to analyze the queuing behavior we considered a queue with two service rates, 517 KBytes/s and 452 KBytes/s, corresponding respectively to link utilizations of $\rho = 0.7$ and $\rho = 0.8$. The buffer size was varied from 1 to 25000 packets. The average packet size for this trace is 638 bytes. Figure 6 shows that the packet loss ratios of original and fitted traces are quite close for all buffer sizes and both link utilizations. This confirms the good matching obtained in both first and second order statistics. It also shows that the oscillatory behavior of the autocovariance has negligible impact on the queuing behavior. Figure 7 shows that the fitted trace exhibits LRD, since the y_j values are aligned between octave 9 and octave 11, the highest octave present in data. The estimated Hurst parameter is $\widehat{H} = 0.838$.

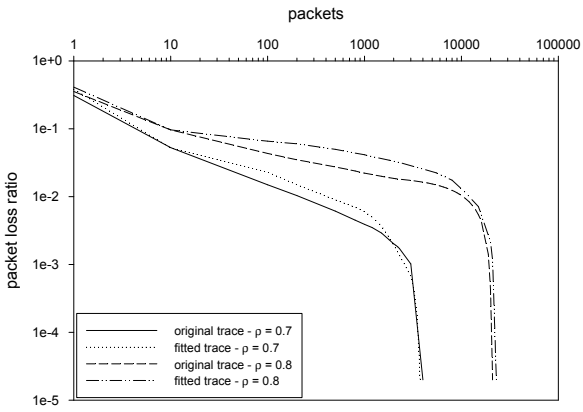

Figure 6. Packet loss ratio, pOct.TL.

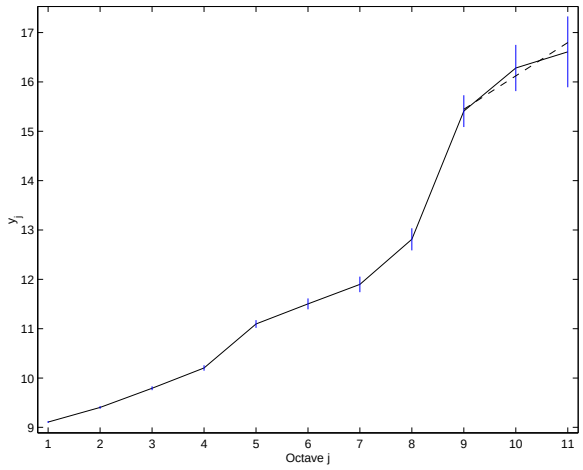

Figure 7. Scaling analysis, pOct.TL fitted.

The correlation horizon (CH) for the pOct.TL trace, with a buffer size of 5000 packets and a link utilization of 0.8 (the largest one considered in the queuing behavior study) is 31.1 seconds. As shown in Figure 5, this value is well bellow the maximum autocovariance lag that is matched by our fitting procedure, thus confirming its ability to capture the LRD behavior up to the time-scales of interest to the system under study.

The fitting results for the pAug.TL trace are similar to the ones obtained with the pOct.TL trace, as can be seen in Figure 8 and Figure 9. The fitting of the probability function is slightly better, but the fitting of the autocovariance function is worse. The relatively high fitting error obtained in this case is due to the strong oscillatory behavior of the empirical autocovariance. The packet loss ratios of original and fitted traces, shown in Figure 10, are almost coincident for all buffer sizes and link utilizations. In this case the simula-

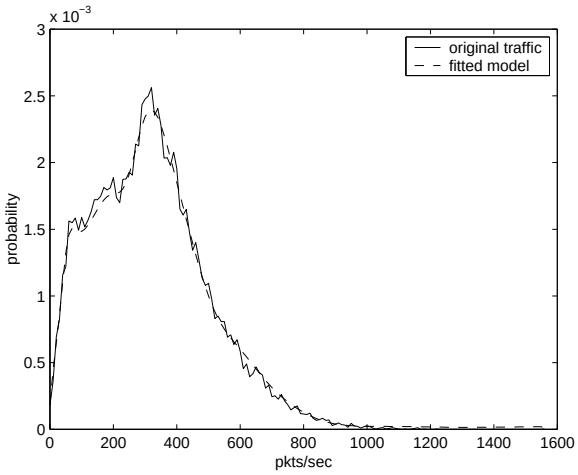

Figure 8. Probability function, pAug.TL.

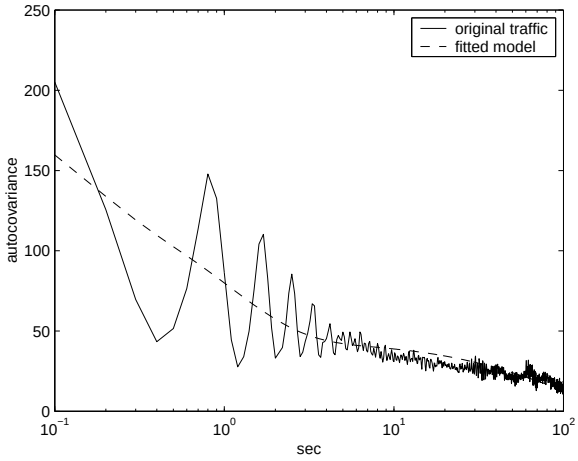

Figure 9. Autocovariance, pAug.TL.

tions were performed with service rates of 197 KBytes/s and 172 KBytes/s, corresponding again to link utilizations of $\rho = 0.7$ and $\rho = 0.8$. The buffer size was varied from 1 to 5000 packets. The average packet size for this trace is 434 bytes. Again, the fitted trace exhibits LRD, since the y_j values are aligned between octave 6 and octave 11, the highest octave present in data, as shown in Figure 11. The estimated Hurst parameter is $\widehat{H} = 0.714$. The CH for the maximum buffer size (5000 packets) and maximum link utilization (0.8) is 15.5 seconds, and Figure 9 shows that, as in the pOct.TL case, this value is bellow the maximum autocovariance lag that is matched by the fitting procedure.

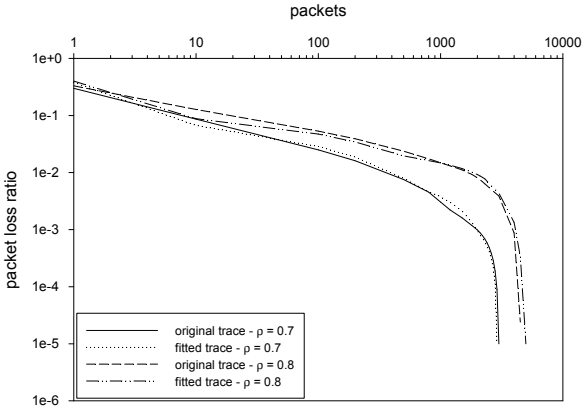

Figure 10. Packet loss ratio, pAug.TL.

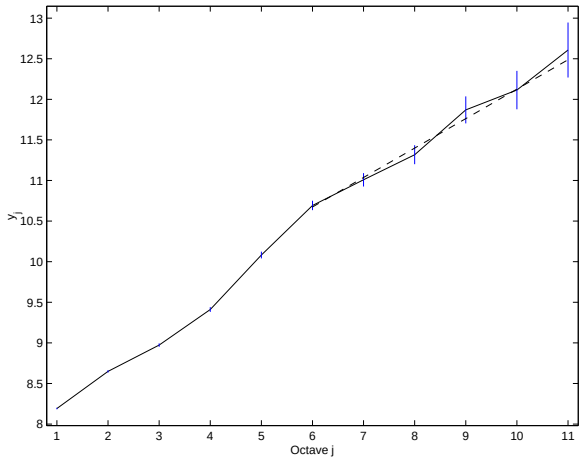

Figure 11. Scaling analysis, pAug.TL fitted.

4.2. INTERNET TRACES MEASURED AT THE UNIVERSITY OF AVEIRO

The inference procedure was also applied to the UA19b trace, a 5 hours trace with approximately 20 million packets (much larger and longer than the Bellcore traces). The procedure fitted the trace to a 12-dMMPP model. Only two relevant time-scales were identified, which explains the smaller number of states of the fitted dMMPP, when comparing with the Bellcore traces. The parameters of the two exponentials are $\vec{\alpha} = [1.00 \times 10^{2} \quad 6.87 \times 10^{1}]$ and $\vec{\beta} = [-6.91 \times 10^{-5} \; -1.28 \times 10^{-2}]$. As in the Bellcore traces, there was a close fitting of both the probability function (Figure 12) and the autocovariance (Figure 13). The packet loss ratio curves (Figure 17) are also very close. The service rates were 685 KBytes/s and 629 KBytes/s, corresponding respectively to link utilizations of $\rho = 0.9$ and $\rho = 0.98$. The reason for

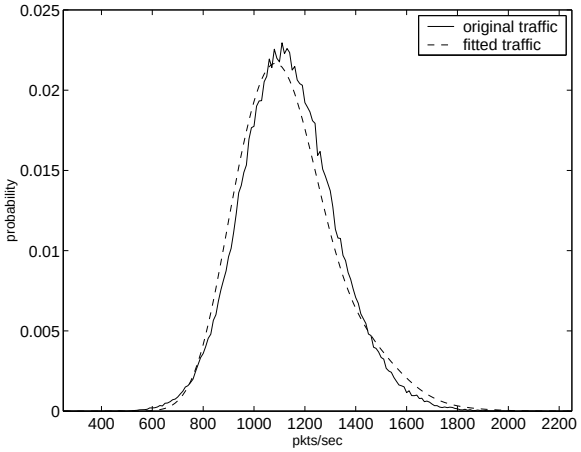

Figure 12. Probability function, UA19b trace.

using link utilizations higher than with the Bellcore traces, is due to the lower burstiness of the UA traffic which leads to lower packet losses for the same link utilization. The buffer size was varied from 1 to 9000 packets. The average packet size of this trace is 557 bytes. The CH is 14.5 seconds for a buffer size of 9000 packets and a link utilization of 0.98, which is again bellow the maximum autocovariance lag matched by the fitting procedure. Figure 14 and Figure 15 show that both traces exhibit LRD, since the y_j values are aligned between a medium octave (7 for the original data trace, 5 for the fitted data trace) and the highest octave present in data. The estimated Hurst parameters are $\widehat{H} = 0.952$ for the original data trace and $\widehat{H} = 0.935$ for the fitted data trace.

The results obtained with the remaining four UA traces are similar. The queuing results are shown in Figures 17, 18, 19 and 20. The link utilizations were again $\rho = 0.9$ and $\rho = 0.98$ for all traces. In all cases the buffer size was varied from 100 packets to 12000 packets.

5. Related work

In this section, we restrict our attention to fitting procedures for MMPPs. We start by noting that most published procedures only apply to 2-MMPPs [24–27]. While 2-MMPPs can capture traffic burstiness, the number of states is in general not enough to provide a good match of the marginal distribution when the traffic shows variability on a wide range of arrival rates.

Skelly *et al.* [28] propose a method for estimating the parameters of a generic MMPP that only matches the first-order statistics: the Poisson arrival rates are inferred from the empirical density function and the state transition rates from a direct measurement of the observed trace. We use a similar

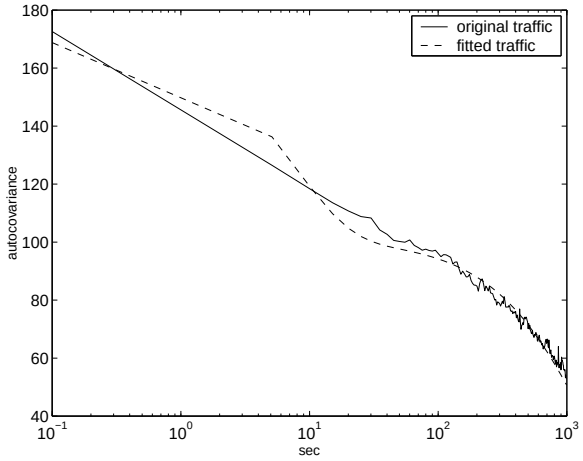

Figure 13. Autocovariance, UA19b trace.

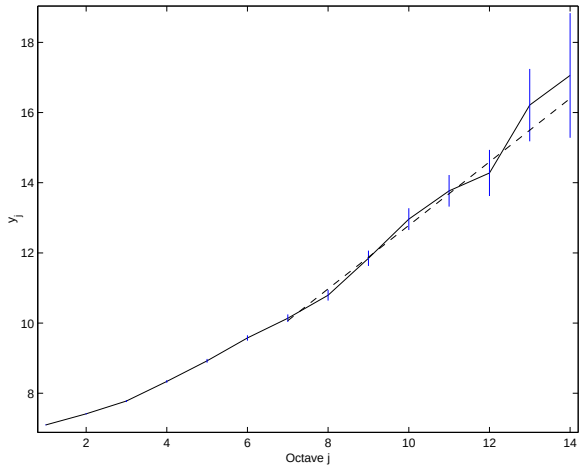

Figure 14. Scaling analysis, UA19b trace original

method in the steps of our procedure described in subsection 3.3. However, a limitation in [28] is that the number of states results from arbitrating the number of bins for constructing the empirical density function and the length of these bins determines the Poisson arrival rates. This results in equally spaced Poisson arrival rates. In our case, we approximate the empirical probability function by a weighted sum of Poisson distributions. This allows for a better selection of the Poisson arrival rates, resulting in a lower number of states, when compared with [28], adapted to the particular characteristics of the trace being modeled.

The work by Li and Hwang [29] is closely related to ours, in that it also matches both the autocovariance and the marginal distribution. The fitting

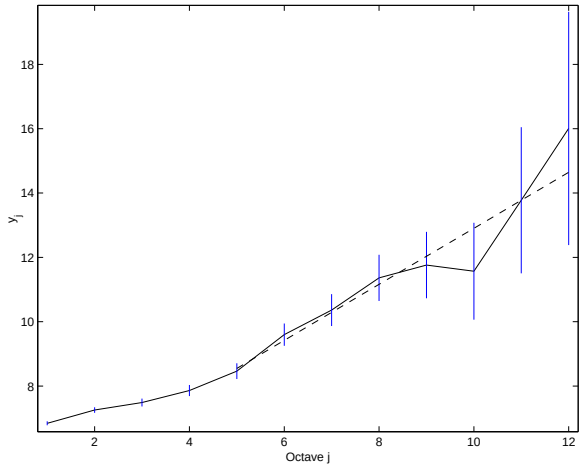

Figure 15. Scaling analysis, UA19b trace fitted.

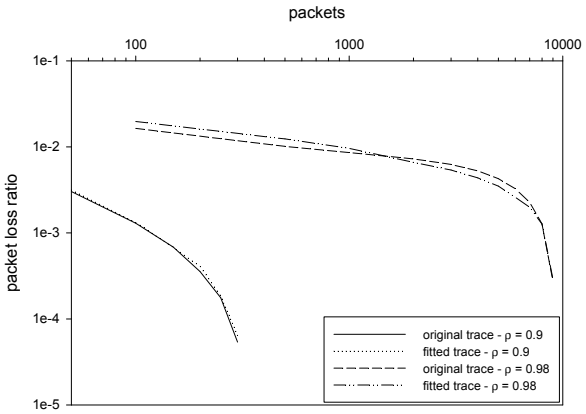

Figure 16. Packet loss ratio, UA19b trace.

procedure applies to CMPPs, which are a special case of MMPPs where the steady-state probabilities are the same for all states. The structure of the CMPP allows circumventing the so-called inverse eigenvalue problem, which is associated with the need of inverting the exponential of the infinitesimal generator matrix to obtain the transition rates from the fitted autocovariance. As opposed to ours, the fitting procedure is able to capture pseudoperiodic components present in data (which shows up in the Bellcore traces), since the infinitesimal generator matrix of a CMPP can have complex eigenvalues. However, our experiments indicate that the pseudoperiodic components have a small importance in what concerns queuing performance. Moreover, in [29] there is less flexibility in adjusting the marginal distribution, since the CMPP states are equiprobable. In particular, with this procedure it is more difficult to detect low probability peaks on the arrival rate. If these peaks occur at high

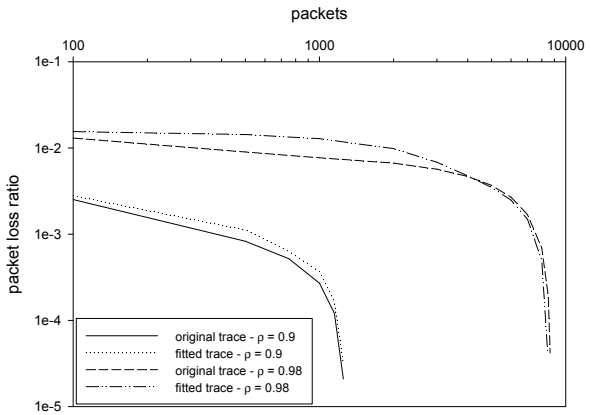

Figure 17. Packet loss ratio, UA17 trace.

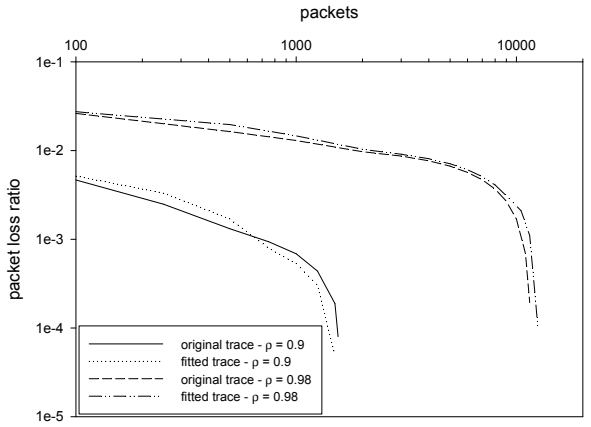

Figure 18. Packet loss ratio, UA1b trace.

arrival rates the queuing behavior can be significantly affected. In [29] the detection resolution can only be improved by increasing the number of states. Our procedure is well adapted to this case, since the method for matching the empirical probability function is specifically based on detection of local peak arrival rates.

As in our approach, Andersen and Nielsen [13] use a superposition of 2-MMPPs to model the different time-scales of the autocovariance. Each of the time-scales in the autocorrelation function is fitted to an exponential function, using a procedure similar to that of [38]. However, the time-scales are defined a priori, whereas in our case a procedure to determine the significant time scales was implemented. Moreover, in [13] the fitting of the first-order statistics is very poor, since only the mean is matched.

Deng and Mark [30] propose a method for estimating the parameters of a MMPP with any number of states, based on the maximum-likelihood princi-

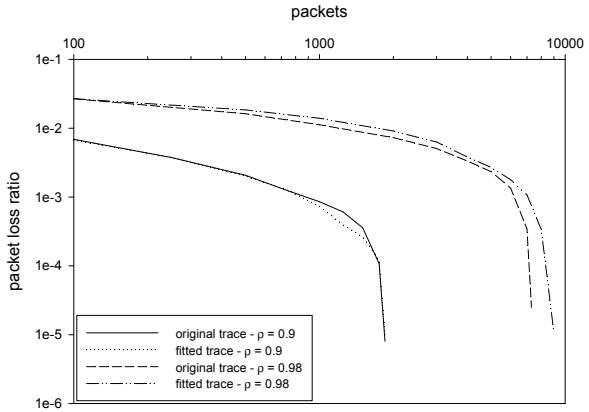

Figure 19. Packet loss ratio, UA6b trace.

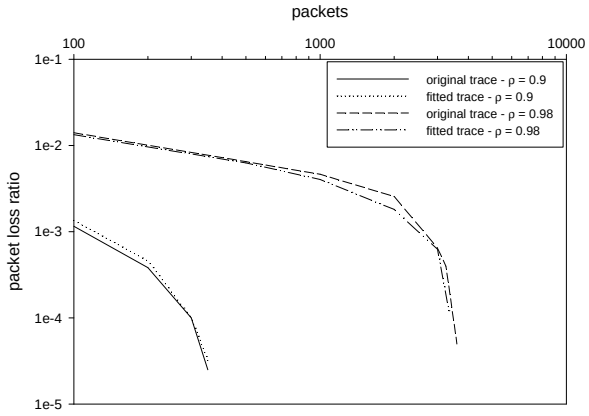

Figure 20. Packet loss ratio, UA10b trace.

ple. The same principle was also used in [26,27] in the context of 2-MMPPs. As referred in [27], the method of [30] is quite sensitive to the choice of the sampling interval and can lead to an exceedingly high number of states. These works do not directly address the issue of matching statistics of observed data. Instead, they are targeted to minimizing the estimation error, under the assumption that the underlying process is an MMPP with a number of states known a priori. These assumptions severely weaken the application of these methods in the area of traffic modeling. In fact, the main challenge in traffic modeling is the discovery of models and associated parameter estimation procedures that provide flexible means of capturing the statistics of observed data that have more impact on network performance. While the MMPP model provides such framework, the performance of the parameter estimation procedure should not be dependent on the assumption that the underlying population is indeed of the MMPP type.

6. Conclusions

This paper proposed a parameter fitting procedure using Markov Modulated Poisson Processes (MMPPs) that leads to accurate estimates of queuing behavior for network traffic exhibiting LRD behavior. Traffic models are a key element in network traffic engineering, since they enable efficient resource allocation (e.g by exploiting potential multiplexing gains through effective bandwidths) and optimal configurations of the network operational parameters (e.g. the queue weights at packet schedulers). The consideration of MMPPs is motivated by the availability of theoretical results for analyzing the queuing behavior and for determining effective bandwidths. The procedure used to fit an MMPP matches both the autocovariance and marginal distribution of the counting process. The MMPP is constructed as superposition of L 2-MMPPs and one M-MMPP. The M-MMPP is designed to match the probability function and the L 2-MMPPs to match the autocovariance function. A major feature is that the number of states is not fixed a priori, and can be adapted to the particular trace being modeled. Our numerical results, which include fitting traffic traces that exhibit long-range dependence, show that the procedure matches closely the autocovariance and probability functions. The queuing behavior, as assessed by the packet loss ratio suffered by the measured and the fitted traces, also shows a very good agreement. Furthermore, the results illustrate that MMPP models, although not being intrinsically long-range dependent, can capture this type of behavior up to the time-scales of interest to the system under study.

Acknowledgements

The authors would like to thank the anonymous referees for their valuables comments and suggestions.

This research was supported in part by Fundação para a Ciência e a Tecnologia, the project POSI/42069/CPS/2001, and the grants BD/19781/99 and SFRH/BSAB/251/01.

References

1. 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, no. 1, pp. 1–15, Feb. 1994.
2. J. Beran, R. Sherman, M. Taqqu, and W. Willinger, "Long-range dependence in variable-bit rate video traffic," *IEEE Transactions on Communications*, vol. 43, no. 2/3/4, pp. 1566–1579, 1995.

3. M. Crovella and A. Bestavros, "Self-similarity in World Wide Web traffic: Evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835–846, Dec. 1997.
4. V. Paxson and S. Floyd, "Wide-area traffic: The failure of Poisson modeling," *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, pp. 226–244, Jun. 1995.
5. B. Ryu and A. Elwalid, "The importance of long-range dependence of VBR video traffic in ATM traffic engineering: Myths and realities," *ACM Computer Communication Review*, vol. 26, pp. 3–14, Oct. 1996.
6. W. Willinger, M. Taqqu, R. Sherman, and D. Wilson, "Self-similarity through high-variability: Statistical analysis of ethernet LAN traffic at the source level," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 71–86, Feb. 1997.
7. W. Willinger, V. Paxson, and M. Taqqu, *Self-similarity and Heavy Tails: Structural Modeling of Network Traffic*, A Practical Guide to Heavy Tails: Statistical Techniques and Applications. Birkhauser, 1998.
8. M. Grossglauser and J. C. Bolot, "On the relevance of long-range dependence in network traffic," *IEEE/ACM Transactions on Networking*, vol. 7, no. 5, pp. 629–640, Oct. 1999.
9. A. Nogueira and R. Valadas, "Analyzing the relevant time scales in a network of queues," in *Proceedings of Internet Performance and Control of Network Systems II, ITCom 2001, SPIE vol. 4523*, Aug. 2001, pp. 243–252.
10. T. Yoshihara, S. Kasahara, and Y. Takahashi, "Practical time-scale fitting of self-similar traffic with Markov-modulated Poisson process," *Telecommunication Systems*, vol. 17, no. 1-2, pp. 185–211, May/Jun 2001.
11. P. Salvador and R.Valadas, "Framework based on Markov modulated Poisson processes for modeling traffic with long-range dependence," in *Proceedings of Internet Performance and Control of Network Systems II, ITCom 2001, SPIE vol. 4523*, Aug. 2001, pp. 221–232.
12. P. Salvador and R. Valadas, "A fitting procedure for Markov modulated Poisson processes with an adaptive number of states," in *Proceedings of the 9th IFIP Working Conference on Performance Modelling and Evaluation of ATM & IP Networks*, Jun. 2001.
13. A. Andersen and B. Nielsen, "A Markovian approach for modeling packet traffic with long-range dependence," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 5, pp. 719–732, Jun. 1998.
14. B. Hajek and L. He, "On variations of queue response for inputs with the same mean and autocorrelation function," *IEEE/ACM Transactions on Networking*, vol. 6, no. 5, pp. 588–598, 1998.
15. A. Feldmann, A.C. Gilbert, and W. Willinger, "Data networks as cascades: Investigating the multifractal nature of internet WAN traffic," in *Proceedings of SIGCOMM 98*, 1998, pp. 42–55.
16. A. Feldmann, A.C. Gilbert, P. Huang, and W. Willinger, "Dynamics of IP traffic: A study of the role of variability and the impact of control," in *Proceedings of SIGCOMM 99*, 1999, pp. 301–313.
17. R. Riedi and J. L. Véhel, "Multifractal properties of TCP traffic: a numerical study," *Technical Report No 3129, INRIA Rocquencourt, France*, Feb. 1997, Available at www.dsp.rice.edu/∼riedi.
18. A. Erramilli, O. Narayan, A. Neidhardt, and I. Saniee, "Performance impacts of multi-scaling in wide area TCP/IP traffic," in *Proceedings of INFOCOM 2000*, 2000.
19. D.M. Lucantoni, "The BMAP/G/1 queue: A tutorial," in *Models and Techniques for Performance Evaluation of Computer and Communication Systems*, L. Donatiello and R. Nelson, Eds., pp. 330–358. Springer Verlag, 1993.

20. A. Elwalid and D. Mitra, "Effective bandwidth of general Markovian traffic sources and admission control of high speed networks," *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pp. 329–343, Jun. 1993.
21. G. Kesidis, J. Walrand, and C. Chang, "Effective bandwidths for multiclass Markov fluids and other ATM sources," *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 424–428, Aug. 1993.
22. W. Fischer and K. Meier-Hellstern, "The Markov-modulated Poisson process (MMPP) cookbook," *Performance Evaluation*, vol. 18, no. 2, pp. 149–171, 1993.
23. S. Shah-Heydari and T. Le-Ngoc, "MMPP models for multimedia traffic," *Telecommunication Systems*, vol. 15, no. 3-4, pp. 273–293, Jan. 2000.
24. R. Grünenfelder and S. Robert, "Which arrival law parameters are decisive for queueing system performance," in *Proceedings of ITC 14*, 1994.
25. S. Kang and D Sung, "Two-state MMPP modelling of ATM superposed traffic streams based on the characterisation of correlated interarrival times," in *Proceedings of IEEE GLOBECOM 95*, Nov. 1995, pp. 1422–1426.
26. K. Meier-Hellstern, "A fitting algorithm for Markov-modulated Poisson process having two arrival rates," *European Journal of Operational Research*, vol. 29, 1987.
27. C. Nunes and A. Pacheco, "Parametric estimation in MMPP(2) using time discretization," in *Proceedings of the 2nd Internation Symposium on Semi-Markov Models: Theory and Applications*, Dec. 1998.
28. P. Skelly, M. Schwartz, and S. Dixit, "A histogram-based model for video traffic behaviour in an ATM multiplexer," *IEEE/ACM Transactions on Networking*, pp. 446–458, Aug. 1993.
29. S. Li and C. Hwang, "On the convergence of traffic measurement and queuing analysis: A statistical-match and queuing (SMAQ) tool," *IEEE/ACM Transactions on Networking*, pp. 95–110, Feb. 1997.
30. L. Deng and J. Mark, "Parameter estimation for Markov modulated Poisson processes via the EM algorithm with time discretization," *Telecommunication Systems*, vol. 1, pp. 321–338, 1993.
31. S. Robert and J. Le Boudec, "New models for self-similar traffic," *Performance Evaluation*, vol. 30, no. 1-2, Jul. 1997.
32. D. Veitch and P. Abry, "A wavelet based joint estimator for the parameters of LRD," *IEEE Transactions on Information Theory*, vol. 45, no. 3, Apr. 1999.
33. C. Blondia and O. Casals, "Statistical multiplexing of VBR sources: A matrix-analytic approach," *Performance Evaluation*, vol. 16, no. 1-3, pp. 5–20, 1992.
34. C. Blondia, "A discrete-time batch Markovian arrival process as B-ISDN traffic model," *Belgian Journal of Operations Research, Statistics and Computer Science*, vol. 32, pp. 3–23, 1993.
35. N. Rananand, "Markov approximations to D-BMAPs: information-theoretic bounds on queueing performance," *Stochastic Models*, vol. 11, no. 4, pp. 713–734, 1995.
36. F. Geerts and C. Blondia, "Superposition of Markov sources and long range dependence," in *Proceedings of the IFIP/ICCC International Conference on Information Network and Data Communication*, 1996.
37. Ad Ridder, "Fast simulation of discrete time queues with Markov modulated batch arrivals and batch departures," *AEU International Journal of Electronics and Communications*, vol. 52, pp. 127–132, 1998.
38. A. Feldmann and W. Whitt, "Fitting mixtures of exponentials to long-tail distributions to analyze network performance models," *Performance Evaluation*, vol. 31, no. 3-4, pp. 245–279, 1997.
39. M. Osborne and G. Smyth, "A modified Prony algorithm for fitting sums of exponential functions," *SIAM Journal of Scientific Computing*, vol. 16, pp. 119–138, 1995.

40. D. Karlis and E. Xekalaki, "Robust inference for finite Poisson mixtures," *Journal of Statistical Planning and Inference*, vol. 93, pp. 93–115, 2001.