A Fitting Procedure for Markov Modulated Poisson Processes with an Adaptive Number of States

Paulo Salvador and Rui Valadas
University of Aveiro / Institute of Telecommunications
Campus de Santiago, Aveiro, Portugal
Tel: +351234377900; Fax: +351234377901
e-mail: salvador@av.it.pt; rv@ua.pt

Abstract

This paper proposes a fitting procedure for Markov Modulated Poisson Processes (MMPPs) that matches both the autocovariance tail and marginal distribution of the counting process. A major feature of the procedure is that the number of states is not fixed a priori. It is an output of the fitting process thus allowing the number of states to be adapted to the particular trace being modeled. The MMPP is constructed as a superposition of one M/2-MMPP and one 2-MMPP. The M/2-MMPP is designed to match the probability function and the 2-MMPP to match the autocovariance tail. The procedure starts by approximating the autocovariance by a weighted sum of two exponential functions. The exponential with slower decay is selected to model the autocovariance of the 2-MMPP. The autocovariance tail can be adjusted to capture the long-range dependence characteristics of the traffic. The procedure then fits the M/2-MMPP parameters in order to match the probability function, within the constraints imposed by the autocovariance matching. The number of states is also determined as part of this step. The final MMPP with M states is obtained by superposing the 2-MMPP and the M/2-MMPP. We apply the inference procedure to IP 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, in particular, with the well-known Bellcore traces.

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

1 Introduction

Traffic modeling plays an increasingly important role in the management and planning of modern telecommunications networks. In order to make an efficient use of resources, network operators are required to perform frequent traffic measurements and to derive traffic models capable of describing rigorously its data. When selecting a stochastic model to describe a traffic source, there is the need to consider the fitting procedures available for parameter estimation. The design of the fitting procedure is a trade-off between computational complexity and accuracy and requires careful consideration of the model parameters that have more impact on the performance metrics of interest. This concern is present in several works [1], [2], [3], [4], [5], [6], [7], [8].

In recent years it has been clearly shown through experimental evidence, that network traffic may exhibit properties of self-similarity and long-range dependence (LRD) [9], [10], [11], [12], [13], [14], [15]. These characteristics have significant impact on network performance. However, as pointed out in [5], matching the LRD is only required within the time scales of interest to the system under study. For example, in order to analyze 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 LRD.

Providing a good match of the LRD behavior is not enough for accurate prediction of the queuing behavior. The first-order statistics need also careful consideration. The work in [16] discusses the limitations of using only the mean and the autocorrelation function, as statistical descriptors of the input process for the purpose

Published in proceedings of the 9th IFIP Working Conference on Performance Modelling and Evaluation of ATM & IP Networks, Budapest, June 2001

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. While this issue is of major importance, we believe it remains largely neglected.

The main goal of the present work is to develop a parameter fitting procedure for Markov Modulated Poisson Processes (MMPPs) that matches closely both the autocovariance tail and the marginal distribution of the counting process. The work is also motivated by the need to keep the number of states of the MMPP at a minimum so as to reduce the complexity associated with the calculation of the performance metrics of interest. The simultaneous matching of these two statistics is a difficult task since every MMPP parameter has an influence on both. With the purpose of achieving some degree of decoupling when matching the two statistics, we construct the MMPP as a superposition of one M/2-MMPP and one 2-MMPP. We will denote the resulting process as (2⊕M/2)-MMPP. We use the 2-MMPP to match the autocovariance tail and the M/2-MMPP to match the probability function. While superposing these two MMPPs will modify both statistics, we devised a procedure where this is done in a controlled way. Specifically, the procedure assures that the M/2-MMPP contributes only with short-range dependence components to the autocovariance function. Moreover, the probability function of the M/2-MMPP is obtained through deconvolution of the 2-MMPP and (2⊕M/2)-MMPP probability functions thus ensuring that the contribution of the 2-MMPP is taken into account. The procedure starts by approximating the autocovariance by a weighted sum of two exponential functions. The exponential with slower decay is selected to model the autocovariance of the 2-MMPP. The autocovariance tail can be adjusted to capture the LRD characteristics of the traffic. We note that this fitting process captures only the decay and amplitude of the autocovariance tail, and neglects any oscillatory behavior. However, our results indicate that the oscillatory behavior has a negligible effect in terms of queuing performance. After this step, the procedure fits the M/2-MMPP parameters in order to match the probability function, within the constraints imposed by the autocovariance matching. The number of states is also determined as part of this step. The final (2⊕M/2)-MMPP is obtained by superposing the 2-MMPP and the M/2-MMPP. This fitting procedure favors matching the LRD and the marginal distribution, as opposed to the short-range dependence. We believe that, in general, this agrees with the relative importance of these statistics in terms of queuing behavior.

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 [17]. 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. In section 2 we give some background on MMPPs and on the superposition of MMPPs. In section 3 we describe the fitting procedure. In section 4 we present numerical results, which include applying the fitting procedure to measured traffic traces. In section 5 we compare our work with previously published ones. Finally, in section 6 we conclude the paper.

2 Background

The MMPP with M states is fully characterized by the infinitesimal generator matrix, $\mathbf{Q}$, and by the diagonal matrix of the Poisson arrival rates, $\mathbf{\Lambda}$,

$$\mathbf{Q} = \begin{bmatrix} -\sigma_1 & \sigma_{12} & ... & \sigma_{1M} \\ \sigma_{21} & -\sigma_2 & ... & \sigma_{2M} \\ ... & ... & ... & ... \\ \sigma_{M1} & \sigma_{M2} & ... & -\sigma_M \end{bmatrix} \qquad \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & ... & 0 \\ 0 & 0 & 0 & \lambda_M \end{bmatrix}$$

where $\sigma_i = -\sum_{i \neq j} \sigma_{ij}$, σ_{ij} is the transition rate from state i to state j, and λ_i is the Poisson rate in state i. We also represent diagonal matrix $\mathbf{\Lambda}$ by vector $\vec{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_M]$. The steady-state probability vector of $\mathbf{Q}$ is the solution of the following system of equations: $\vec{\pi}\mathbf{Q} = 0$ and $\vec{\pi}\vec{e} = 1$, where $\vec{e}$ is a unit column vector.

Consider the superposition of N MMPPs, each characterized by matrices $\mathbf{Q}_j$ and $\mathbf{\Lambda}_j$. The resulting process is also an MMPP, with infinitesimal generator matrix $\mathbf{Q}$ and arrival rate matrix $\mathbf{\Lambda}$,

$$\mathbf{Q} = \mathbf{Q}_1 \oplus \mathbf{Q}_2 \oplus ... \oplus \mathbf{Q}_N \tag{1}$$

$$\mathbf{\Lambda} = \mathbf{\Lambda}_1 \oplus \mathbf{\Lambda}_2 \oplus ... \oplus \mathbf{\Lambda}_N \tag{2}$$

where $\oplus$ denotes the Kronecker sum [18].

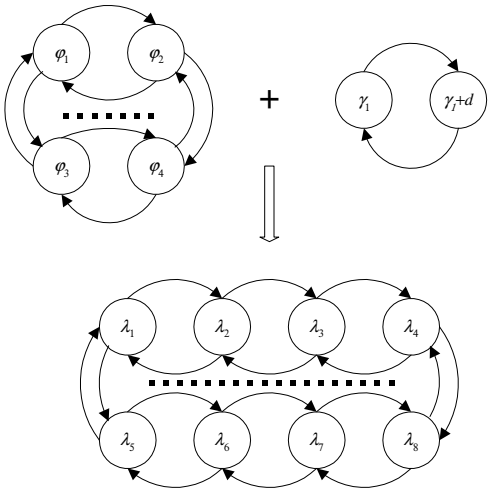

Figure 1: Superposition of M/2-MMPP and 2-MMPP models (case of M = 8).

We will restrict ourselves to the case of the superposition of one M/2-MMPP and one 2-MMPP. The M/2-MMPP will be characterized by the arrival rates vector $\vec{\varphi} = [\varphi_1, \varphi_2, ..., \varphi_{M/2}]$ and the transition rates matrix $\mathbf{Q}_1$. Similarly, the 2-MMPP will be characterized by the arrival rates vector $\vec{\gamma} = [\gamma_1, \gamma_2]$ and the transition rates matrix $\mathbf{Q}_2$

The superposition process is represented in figure 1. The inference procedure works with the counting process, $\{X(k\Delta), k = 1, 2, ...\}$, where $X(k\Delta)$ represents the number of arrivals in the k^{th} sampling interval of duration Δ of the continuous-time MMPP.

The sampling interval Δ should be selected according to the traffic characteristics: a too large interval may smooth away important traffic characteristics (e.g., burstiness); a too short interval may result in a process that has no arrivals in the majority of sampling intervals. Our experiments indicate that a reasonable value for Δ is one that allows the mean number of arrivals per interval to be in the range of $[10, 100]$.

3 Inference Procedure

The inference procedure can be divided in five steps: (i) approximation of the empirical autocovariance by a weighted sum of two exponentials, (ii) inference of the M/2-MMPP probability function and of the 2-MMPP parameters, (iii) inference of the M/2-MMPP Poisson arrival rates, (iv) inference of the M/2-MMPP transition rates and (v) calculation of the final M-MMPP parameters. The flow diagram of this procedure is represented in figure 2. In the following sections we describe these steps in detail. Note that the inference procedure itself is independent of the parameter Δ.

3.1 Autocovariance approximation

The autocovariance function of the 2-MMPP counting process, $X(k\Delta)$, is a single exponential given by:

$$C(\Delta l) = d^2 p_2 (1 - p_2) e^{-(r_{11}+r_{22})\Delta l}, \qquad l = 0, 1, 2, ... \tag{3}$$

where $d = |\gamma_1 - \gamma_2|$ and p_2 is the steady-state probability of state 2. In our approach, this exponential is fitted to the tail of the empirical autocovariance.

As a first step, we approximate the empirical autocovariance by a sum of two exponentials with real positive weights and negative real time constants. This is accomplished through a modified Prony algorithm [19], and the approximation is validated using the methods of [4]. The Prony algorithm returns two vectors,

$$\vec{\alpha} = [\ \alpha_1 \quad \alpha_2\ ] \qquad \vec{\beta} = [\ \beta_1 \quad \beta_2\ ]$$

which correspond to the approximating function

$$C_a(\Delta l) = \alpha_1 e^{-\beta_1 \Delta l} + \alpha_2 e^{-\beta_2 \Delta l}, \qquad l = 0, 1, 2, ... \tag{4}$$

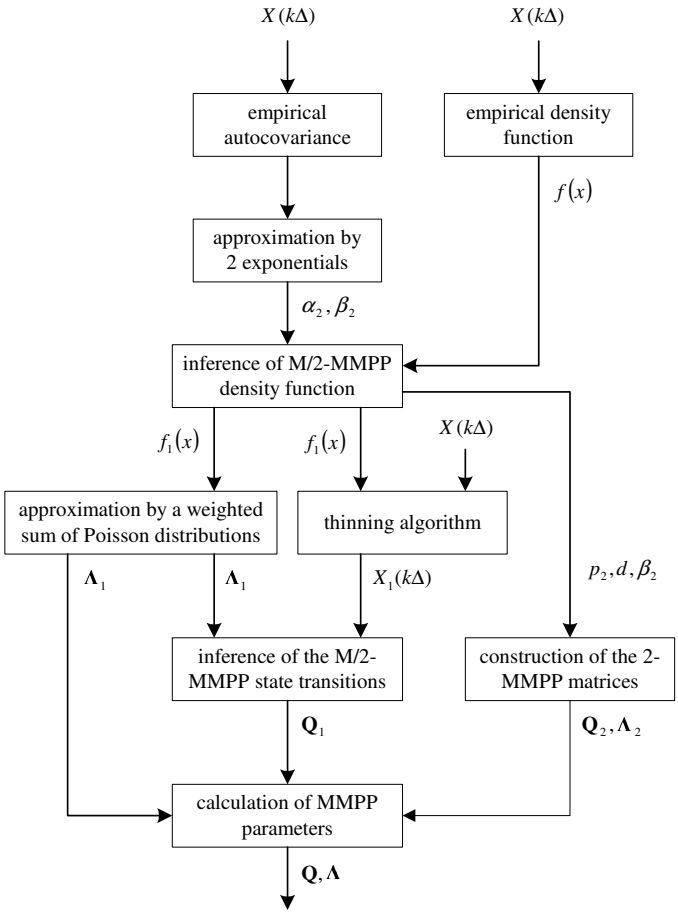

Figure 2: Flow diagram of the inference procedure.

Assuming $\beta_2 \leq \beta_1$, we select the second exponential with parameters β_2 and α_2 as the autocovariance tail approximation. From (3) and (4) results

$$\alpha_2 = d^2 p_2 (1 - p_2) \tag{5}$$

and

$$\beta_2 = r_{11} + r_{22} \tag{6}$$

3.2 Inference of the M/2-MMPP probability function and of the 2-MMPP parameters

The next step is the inference of the M/2-MMPP probability function from the empirical probability function of the original data trace. Let $f(x)$, $f_1(x)$ and $f_2(x)$ denote the probability functions of the (2⊕M/2)-MMPP, M/2-MMPP and 2-MMPP, respectively. As a result of the superposition of the M/2-MMPP and 2-MMPP sources $f(x)$ is given by

$$f(x) = f_1(x) \otimes f_2(x) \tag{7}$$

where $\otimes$ denotes convolution. We consider that the Poisson arrival rate in one state of the 2-MMPP source is zero ($\gamma_1 = 0$) and is constant and equal to d on the other state. In other words, we approximate the probability function of the 2-MMPP by that of an Interrupted Deterministic Poisson (IDP) process. This simplification allows to obtain $f_1(x)$ from $f(x)$ directly, avoiding deconvolution, which results in a faster algorithm. Thus:

$$f_2(x) = (1 - p_2)\,\delta(x) + p_2 \delta(x - d) \tag{8}$$

where $\delta(x)$ is the dirac function. Then, from (7) and (8),

$$f(x) = (1 - p_2) f_1(x) + p_2 f_1(x - d) \tag{9}$$

Since d and p_2 are related by (5) and given that $f_1(x) \geq 0$ for $x \geq 0$ and $f_1(x) = 0$ otherwise, an exact solution for obtaining $f_1(x)$ from (9) may not exist. An approximation to $f_1(x)$ can be obtained through the following minimization process

$$\min_{p_2} \int |f(x) - (1 - p_2) f_1(x) - p_2 f_1(x - d)| dx \tag{10}$$

where $0 < p_2 < 1$. Note that p_2 is not allowed to be 0 or 1 because, in both cases, the 2-MMPP would degenerate into a Poisson process. In practice we restrict p_2 to be in the range $[0.01, 0.09]$.

A numerical solution for (10) can resort to a scalar bounded nonlinear minimization routine [20], which requires the definition of an approximation error on the minimization variable. In order to calculate this error, we define an iterative method for obtaining $f_1(x)$ from $f(x)$ given p_2. Consider that the possible values of the counting process are arranged in vectors $\vec{x}_i = [di, di + 1, ..., d(i + 1) - 1], 0 \leq i \leq \left\lceil \max_k \{X(k\Delta)/d\} \right\rceil$. Each vector $\vec{x}_i$ has d elements and each element represents a number of arrivals in the sampling interval Δ. Since $f_1(x) = 0$ for $x < 0$,

$$f_1(\vec{x}_0) = \frac{f(\vec{x}_0)}{1 - p_2}$$

and, for other $\vec{x}_i$,

$$f_1(\vec{x}_i) = \frac{f(\vec{x}_i)}{1 - p_2} - \frac{p_2}{1 - p_2} f_1(\vec{x}_{i-1})$$

When this iterative process ends, negative values for $f_1(x)$ may result. In this case, we zero all negative values, resulting in the approximation error defined by

$$\varepsilon = \sum_{x,i} |f(\vec{x}_i) - (1 - p_2) f_1(\vec{x}_i) - p_2 f_1(\vec{x}_{i-1})|$$

At this point all parameters of the 2-MMPP have been determined and the 2-MMPP matrices can be constructed. These are

$$\mathbf{\Lambda}_2 = \begin{bmatrix} 0 & 0 \\ 0 & d \end{bmatrix} \qquad \mathbf{Q}_2 = \begin{bmatrix} -p_2\beta_2 & p_2\beta_2 \\ (1 - p_2)\beta_2 & -(1 - p_2)\beta_2 \end{bmatrix}$$

3.3 Inference of the number of states and Poisson arrival rates of the M/2-MMPP

The next step is the inference of the number of states and Poisson arrival rates of the M/2-MMPP from $f_1(x)$. To do this we approximate $f_1(x)$ by a weighted sum of Poisson probability functions, where each Poisson distribution is associated to one state of the M/2-MMPP. In this case, the mean and the weight of the Poisson distribution are the state Poisson arrival rate and steady-state probability, respectively. This is a good approximation whenever the mean sojourn time in each state of the inferred MMPP is several times greater than the sampling interval of the counting process. Our numerical results validate this assumption.

The approximation is carried out through an algorithm that progressively subtracts a Poisson probability function from $f_1(x)$. This algorithm is described in the flowchart of Figure 3. We represent the j^{th} Poisson probability function with mean φ_j by $g_j(x)$. We define $f_d(x)$ as the difference between $f_1(x)$ and the weighted sum of Poisson probability functions. Initially, we set $f_d(x) = f_1(x)$. In each step, we first detect the maximum of $f_d(x)$. The corresponding x-value, $\varphi_j = f_d^{-1}(\max f_d(x))$, will be considered the j^{th} Poisson rate of the M/2-MMPP. We then calculate the weights of each Poisson probability function, $\vec{w} = [w_1, w_2, ..., w_i]$, through the following set of linear equations:

$$f_1(\varphi_l) = \sum_{j=1}^{i} w_j g_j(\varphi_l), \qquad l = 1, ..., i$$

This assures that the fitting between $f_1(x)$ and the weighted sum of Poisson probability functions is exact at φ_l points. The final step in each iteration is the calculation of the new difference function $f_d(x)$. The

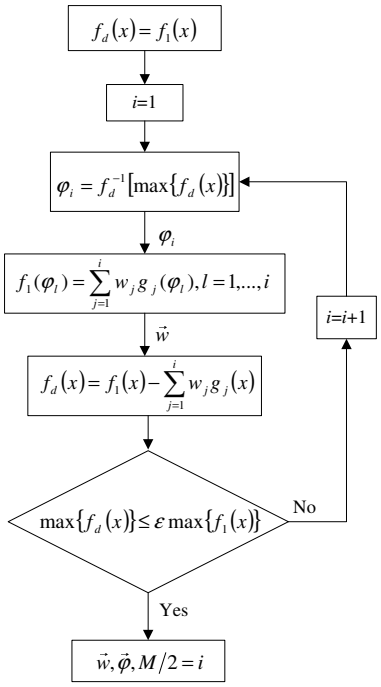

Figure 3: Algorithm for calculation of the number of states and Poisson arrival rates of the M/2-MMPP.

algorithm stops when the maximum of $f_d(x)$ is lower than a pre-defined percentage ε of the maximum of $f_1(x)$. Note that at this point, we set $M/2 = i$. Finally, the matrix of Poisson arrival rates is given by

$$\mathbf{\Lambda}_1 = \begin{bmatrix} \varphi_1 & 0 & ... & 0 \\ 0 & \varphi_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \varphi_{M/2} \end{bmatrix}$$

In the following steps, we assume that the state indexes have been reassigned such that $\varphi_1 < \varphi_2 < ... < \varphi_{M/2}$.

3.4 Inference of the M/2-MMPP transition rates

In order to calculate the M/2-MMPP infinitesimal generator matrix $\mathbf{Q}_1$, we start by extracting, from the original trace, the fraction of data that is associated with the M/2-MMPP from the point of view of the marginal distribution. Note, from (9), that $f(x)$ is a sum of two probability functions, both of them easily related with $f_1(x)$. Thus, a new process $X_1(k\Delta)$ with probability function $f_1(x)$ can be extracted from the original process $X(k\Delta)$ with probability function $f(x)$, using the simple thinning algorithm illustrated in figure 4. In this algorithm, each value of $X(k\Delta)$ will be accepted or rejected as belonging also to $X_1(k\Delta)$ according the following criterion:

1. If $X(k\Delta) < d$ then accept always as belonging to $X_1(k\Delta)$.
2. Otherwise, accept as belonging to $X_1(k\Delta)$ with probability $\frac{(1-p_2) f_1(X(k\Delta))}{p_2 f_1(X(k\Delta)-d)+(1-p_2) f_1(X(k\Delta))}$.

To complete the calculation of the infinitesimal generator matrix $\mathbf{Q}_1$, we first need to assign each value of $X_1(k\Delta)$ to a particular state of the M/2-MMPP and then measure the number of transitions between each pair of states and the average sojourn time in each state. We assume that the process is in state i, $2 \le i \le M/2-1$, in sampling interval $k\Delta$, if

$$\varphi_i - \frac{(\varphi_i - \varphi_{i-1})\, w_i}{w_{i-1}+w_i} \le X_1(k\Delta) < \varphi_i + \frac{(\varphi_{i+1} - \varphi_i)\, w_i}{w_{i+1}+w_i}$$

For state 1, the lower bound is the minimum value of $X_1(k\Delta)$ and, for state M/2, the upper bound is the maximum value of $X_1(k\Delta)$. The state boundaries defined above impose that the steady-state probabilities

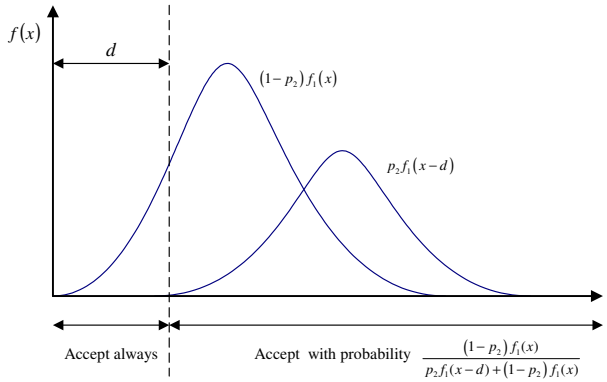

Figure 4: Illustration of the thinning process.

of the M/2-MMPP will be approximately equal to the weights of the Poisson probability functions, w_i, calculated in previous section. This is required in order to assure that the M/2-MMPP probability function is close to $f_1(x)$. Figure 5 illustrates how the weights, w_i, influence the state boundaries defined above: if the weight of state i is higher than the weight of its neighbor k ($i-1$ or $i+1$), the boundary will be closer to φ_k.

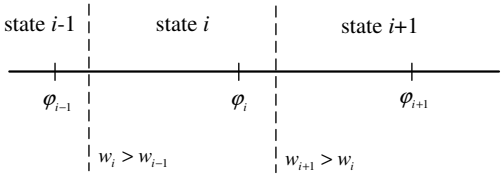

Figure 5: llustration of the boundaries between states.

In order to guarantee that the M/2-MMPP contributes only with short-range dependence components to the autocovariance, the average sojourn time in each state is limited to a maximum that depends on the average sojourn time of the 2-MMPP states. Note that, for state 1 of the 2-MMPP, the average sojourn time is $q_1 = 1/p_2\beta_2$ and for state 2 is $q_2 = 1/(1-p_2)\beta_2$. Specifically, we constrain the sojourn times of the M/2-MMPPs states to be lower than $min(q_1, q_2)/10$. To do this, during the measurement of the state transition probabilities, we force a transition to a randomly selected state (according to the steady-state probabilities $w_i, i = 1, ..., M/2$), whenever the sojourn time exceeds the above-mentioned limit.

Let z_{ij}, denote the measured number of transitions from state i to state j and v_i the measured average sojourn time in state i, obtained using this procedure. The transition rates matrix $\mathbf{Q}_1$ is then given by

$$\mathbf{Q}_1 = \begin{bmatrix} -\frac{1}{v_1} & \frac{1}{v_1}\frac{z_{12}}{\sum_j z_{1j}} & \cdots & \frac{1}{v_1}\frac{z_{1M/2}}{\sum_j z_{1j}} \\ \frac{1}{v_2}\frac{z_{21}}{\sum_j z_{2j}} & -\frac{1}{v_2} & \cdots & \frac{1}{v_2}\frac{z_{2M/2}}{\sum_j z_{2j}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{v_{M/2}}\frac{z_{M/21}}{\sum_j z_{M/2j}} & \cdots & \cdots & -\frac{1}{v_{M/2}} \end{bmatrix}$$

3.5 (2⊕M/2)-MMPP model construction

Finally, the (2⊕M/2)-MMPP process can be constructed from the following equations:

$$\mathbf{\Lambda} = \mathbf{\Lambda}_1 \oplus \mathbf{\Lambda}_2 \qquad \mathbf{Q} = \mathbf{Q}_1 \oplus \mathbf{Q}_2$$

where $\mathbf{\Lambda}_1$, $\mathbf{\Lambda}_2$, $\mathbf{Q}_1$ and $\mathbf{Q}_2$ were calculated in previous sections.

4 Numerical Results

We apply our fitting procedure to several traffic traces: (i) the publicly available Bellcore LAN traces [11] and (ii) traces of IP traffic measured at our institution. The traces measured at our institution are representative of Internet traffic produced within a medium-size research environment with approximately 50 users making random accesses to the Internet. We assess the fitting procedure by comparing the density and the autocovariance functions of the original data traces and of simulated traces obtained from the fitted MMPPs. We analyze the presence of LRD behavior, in both original and fitted data traces, using the method described in [17]. 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.

We also analyze the queuing behavior by comparing the packet loss ratio, obtained through trace-driven simulation, using again the original data traces and the simulated traces obtained from the fitted MMPP. To calculate the packet loss ratio (Bellcore and our institution) we assume a fixed packet size equal to mean packet size. The sampling interval of the counting process was 0.1 seconds for all traces.

4.1 Bellcore traces

The fitting procedure was applied to Bellcore/Telcordia traces pOct.TL and pAug.TL, both with 1 million samples. The pOct.TL trace was fitted to a 16-MMPP and the pAug.TL trace was fitted to a 18-MMPP model.

Figure 6 shows the fitting results for the first-order statistics (pOct.TL trace). In this case, the fitting was performed with a very small approximation error. From figure 7 it can be seen that the autocovariance of the original traffic has an oscillatory behavior and that the fitted traffic captures its average behavior.

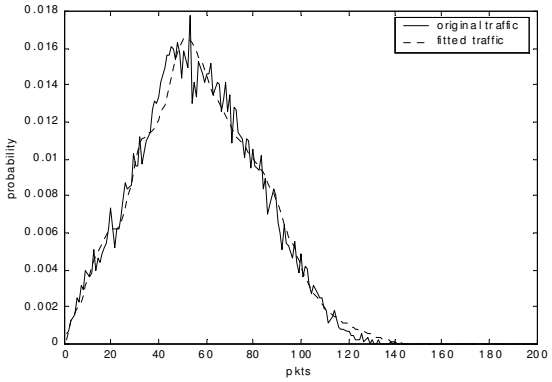

Figure 6: Probability function, pOct.TL.

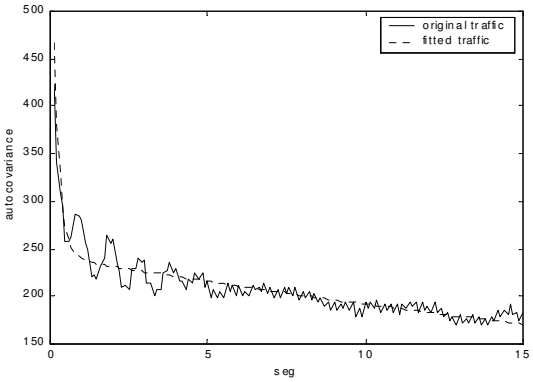

Figure 7: Autocovariance, pOct.TL.

In order to analyze the queuing behavior we considered a queue with a service rate of 750 pkts/s (3.8 Mb/s). The buffer size was varied from 1 to 7000 packets. The average packet size for this trace is 638 bytes. Figure 8 shows that the loss packet ratios of original and fitted traces are almost coincident for all buffer sizes. This confirms the good matching obtained in both first and second order statistics. Figure 9 shows that the fitted trace exhibits LRD, since the y_j values are aligned between octave 6 and octave 11, the highest octave present in data. The estimated Hurst parameter is $\widehat{H} = 0.861$.

The fitting results for pAug.TL trace exhibit similar performance results as the ones obtained with the pOct.TL trace, as can be seen in figure 10 and figure 11. As a result the loss ratios of original and fitted traces, shown in figure 12, are again almost coincident for all buffer sizes. In this case the simulations were performed with a service rate of 470 pkts/s (1.6 Mb/s) and the buffer size was varied from 1 to 2000 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 13. The estimated Hurst parameter is $\widehat{H} = 0.848$.

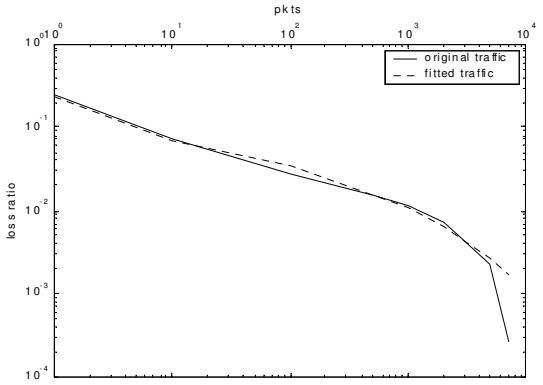

Figure 8: Packet loss ratio, pOct.TL.

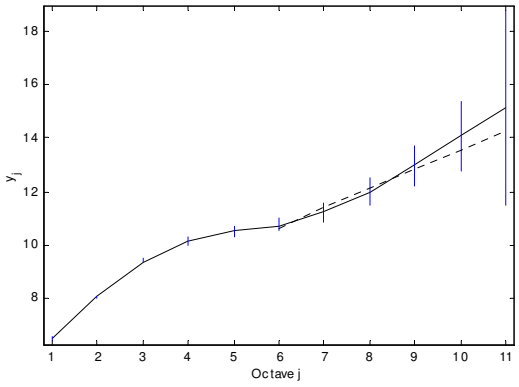

Figure 9: Scaling analysis, pOct.TL fitted.

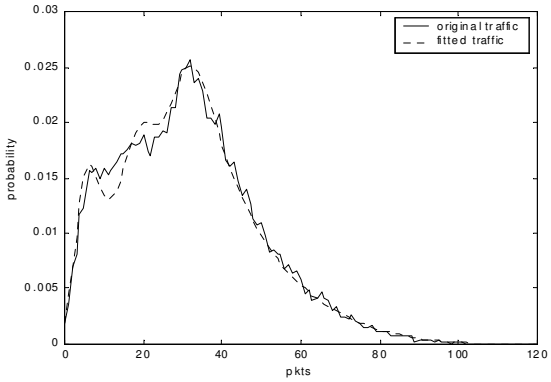

Figure 10: Probability function, pAug.TL.

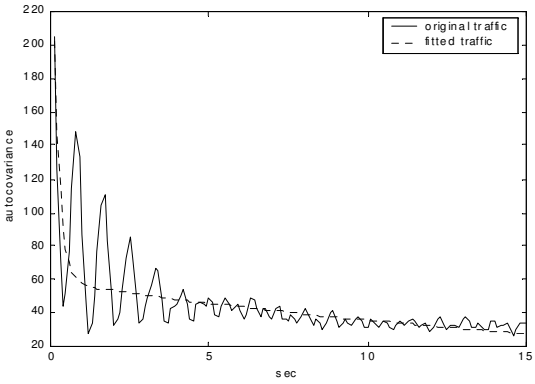

Figure 11: Autocovariance, pAug.TL.

4.2 Internet traces measured at our institution

The inference procedure was applied to traffic measured at our institution (the Institute of Telecommunications - IT), a 40 minutes trace with approximately 500000 packets. We fitted the trace to a 16-MMPP model. There was an almost perfect fitting of the first-order statistics, as can be seen in figure 14. As in previous cases, the matching of the autocovariance for high values of time lag is almost perfect (figure 15), but exhibits some mismatch for lower lag times. This may explain some mismatch in the queuing behavior for small buffer sizes (figure 16). The service rate was 280 pkts/s (1.6 Mb/s) and the buffer size was varied from 1 to 6000 packets. The average packet size of this trace is 714 bytes.

Figure 17 and figure 18 show that both traces exhibit LRD, since the y_j values are aligned between a medium octave (5 for the original data trace, 7 for the fitted data trace) and the highest octave present in data. The estimated Hurst parameters are $\widehat{H} = 0.847$ for the original data trace and $\widehat{H} = 0.802$ for the fitted data trace.

5 Related work

In this section, we restrict our attention to fitting procedures for MMPPs. We start by noting that most procedures only apply to 2-MMPPs [21], [22], [23], [24]. 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.* [25] 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 probability function and the state transition rates from a direct measurement of the observed trace. We use a similar method in the steps of our procedure described in sections 3.3 and 3.4. However, a limitation in [25] is that the number of states results from arbitrating the number of bins for constructing the empirical probability function and the length of these bins determines the Poisson arrival rates. This results in equally spaced Poisson arrival rates.

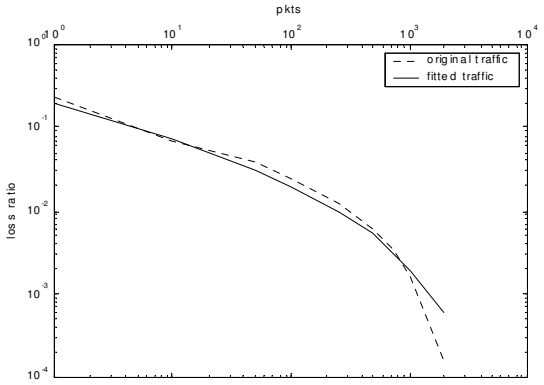

Figure 12: Packet loss ratio, pAug.TL.

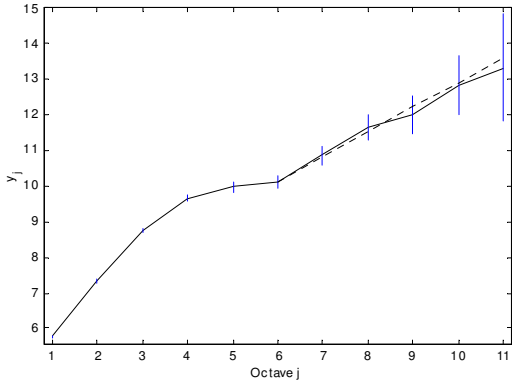

Figure 13: Scaling analysis, pAug.TL fitted.

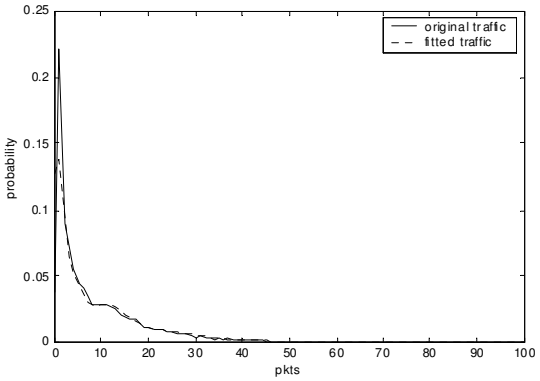

Figure 14: Probability function, IT trace.

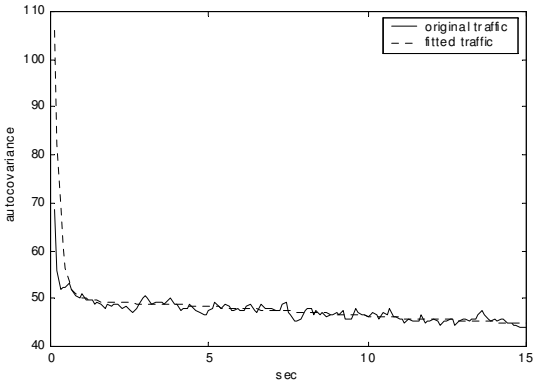

Figure 15: Autocovariance, IT trace.

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 [25], adapted to the particular characteristics of trace being modeled.

The work by Li and Hwang [6] is closely related to ours, in that it also matches both the autocovariance and the marginal distribution. The fitting 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, 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. In [6], there is less flexibility in adjusting the marginal distribution, since 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 rates the queuing behavior can be significantly affected. 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.

Andersen and Nielsen [1] use 2-MMPPs to model several time-scales of the autocovariance function. Each of the time scales is fitted to an exponential function, resulting in a model that corresponds to the superposition of several 2-MMPPs. When compared to ours, this procedure allows a more detailed matching of the autocovariance at low time scales. We only consider the largest time scale where the LRD behavior may be observed. However, the fitting of the first-order statistics is very poor, since only the mean is matched.

Deng and Mark [26] propose a method for estimating the parameters of a MMPP with any number of states, based on the maximum-likelihood principle. The same principle was also used in [23], [24] in the context of 2-MMPPs. As referred in [24], the method of [26] 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 the statistics of observed data. Instead, they are targeted to minimizing the estimation

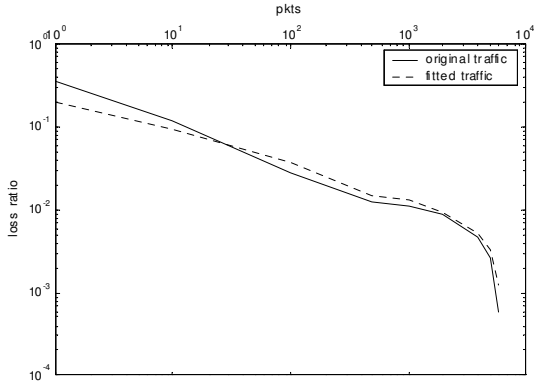

Figure 16: Packet loss ratio, IT trace.

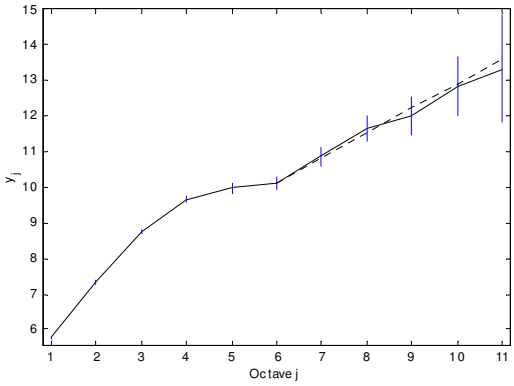

Figure 17: Scaling analysis, IT trace original

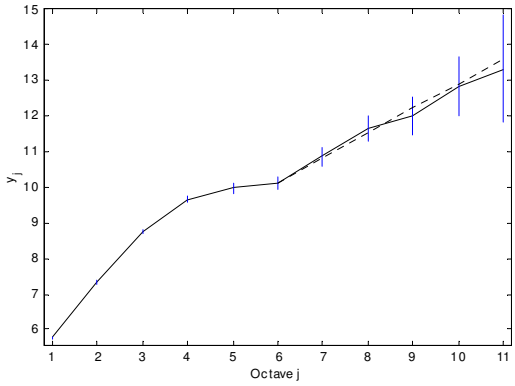

Figure 18: Scaling analysis, IT trace fitted.

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 a 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 fitting procedure for Markov Modulated Poisson Processes (MMPPs) that matches both the autocovariance tail and marginal distribution of the counting process. A major feature of the procedure is that the number of states is not fixed a priori. It is an output of the fitting process thus allowing the number of states to be adapted to the particular trace being modeled. The MMPP is constructed as superposition of one M/2-MMPP and one 2-MMPP. The M/2-MMPP is designed to match the probability function and the 2-MMPP to match the autocovariance tail. Our numerical results, which include fitting traffic traces that exhibit LRD, show that the procedure matches closely the autocovariance tail and the probability function and that the number of states of each MMPP model is relatively small. The queuing behavior as assessed by the packet loss ratio suffered by measured and 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 for limited time scales.

Acknowledgments

This work was partially funded by Portugal Telecom Inovação. P. Salvador wishes to thank Fundação para a Ciência e a Tecnologia, Portugal, for support under grant BD/19781/99. Finally, the authors would like to

acknowledge the valuable comments and suggestions made by anonymous reviewers.

References

[1] 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, pp. 719–732, Jun 1998.

[2] H. Che and S. Li, "Fast algorithms for measurement-based traffic modeling," *IEEE Journal on Selected Areas in Communications*, vol. 16, Jun 1998.

[3] 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.

[4] A. Feldmann and W. Whitt, "Fitting mixtures of exponentials to long-tail distributions to analyze network," *Performance Evaluation*, vol. 31, no. 3-4, pp. 245–279, 1998.

[5] M. Grossglauser and J. Bolot, "On the relevance of long-range dependence in network traffic," *IEEE/ACM Transactions on Networking*, vol. 7, pp. 629–640, Oct 1999.

[6] 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, Fev 1997.

[7] S. Robert and J. L. Boudec, "A modulated markov process for self-similar traffic," *Proceedings of Saarbrucken Schloss Dagstuhl, Germany*, Jan 1995.

[8] S. Robert and J. L. Boudec, "New models for self-similar traffic," *Performance Evaluation*, vol. 30, Jul 1997.

[9] 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.

[10] M. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: Evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 835–846, Dec 1997.

[11] 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, Feb 1994.

[12] V. Paxson and S. Floyd, "Wide-area traffic: The failure of poisson modeling," *IEEE/ACM Transactions on Networking*, vol. 3, pp. 226–244, Jun 1995.

[13] 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.

[14] 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, pp. 71–86, Feb 1997.

[15] 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.

[16] 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.

[17] D. Veitch and P. Abry, "A wavelet based joint estimator for the parameters of LRD," *"Special issue on Multiscale Statistical Signal Analysis and its Applications", IEEE Transactions on Information Theory.*

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

[19] M. Osborne and G. Smyth, "A modified prony algorithm for fitting sums of exponential functions," *SIAM J. Sci. Statist. Comput.*, vol. 16, pp. 119–138, 1995.

[20] G. E. Forsythe, M. A. Malcolm, and C. B. Moler, *Computer Methods for Mathematical Computations.* Prentice Hall, 1976.

[21] R. Grünenfelder and S. Robert, "Which arrival law parameters are decisive for queueing system performance," in *ITC 14*, 1994.

[22] S. Kang and D. Sung, "Two-state MMPP modelling of ATM superposed traffic streams based on the characterisation of correlated interarrival times," *IEEE GLOBECOM'95*, pp. 1422–1426, Nov 1995.

[23] K. Meier-Hellstern, "A fitting algorithm for markov-modulated poisson process having two arrival rates," *European Journal of Operational Research*, vol. 29, 1987.

[24] C. Nunes and A. Pacheco, "Parametric estimation in MMPP(2) using time discretization," *Proceedings of the 2nd Internation Symposium on Semi-Markov Models: Theory and Applications*, Dec 1998.

[25] 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.

[26] L. Deng and J. Mark, "Parameter estimation for markov modulated poisson processes via the EM algorithm with time discretization," *Tel. Systems*, no. 1, pp. 321–338, 1993.