Interarrival Times Characterization and Fitting for Markovian Traffic Analysis

Giuliano Casale, Eddy Z. Zhang, Evgenia Smirni
College of William and Mary
Computer Science Dept.
23187-8795 Williamsburg, VA, US
{casale, eddy, esmirni}@cs.wm.edu

***Abstract*—We propose a traffic fitting algorithm for Markovian Arrival Processes (MAPs) that can capture statistics of any order of interarrival times. By studying real traffic traces, we show that matching higher order properties, in addition to first and second order descriptors, results in increased queueing prediction accuracy with respect to other algorithms that only match the mean, coefficient of variation, and autocorrelations. The result promotes the idea of modeling traffic traces using the interarrival time process instead of the counting process that is more frequently employed in previous work, but for which higher order moments are difficult to manipulate.**

We proceed by first characterizing the general properties of MAPs using a spectral approach. Based on this characterization, we show how different MAP processes can be combined together using Kronecker products to define a larger MAP with predefined properties of interarrival times. We then devise an algorithm that is based on this Kronecker composition and can accurately fit traffic traces. The algorithm employs nonlinear optimization programs that can be customized to fit an arbitrary number of moments and to meet the desired cost-accuracy tradeoff. Numerical results of the fitting algorithm on real HTTP and TCP traffic data, such as the Bellcore Aug89 trace, indicate that the proposed fitting methods achieve increased prediction accuracy with respect to other state-of-the-art fitting methods.

I. Introduction

Markovian models provide a convenient way of evaluating the performance of network traffic since their queueing analysis enjoys established theoretical results and efficient solution algorithms [27]. Although unable to directly generate traffic with long-range dependent (LRD) behavior, Markovian models can approximate accurately LRD traffic in several ways, e.g., by superposition of flows with short-range dependent (SRD) behavior over many time scales. This is known to be sufficient for the evaluation of real systems since the performance effects of LRD traffic becomes nil beyond a finite number of time scales [12].

One of the main obstacles to the Markovian analysis of network traffic is model parameterization, which often requires to describe the interaction of several tens or hundreds of states. Even for basic Markov Modulated Poisson processes (MMPP) or phase-type renewal processes (PH), few results exist for their exact parameterization and they are restricted to models of two or three states only [5], [11], [16], [17], [23]. Due to the lack of characterization results, it is also impossible to establish general properties of these processes.

In this paper we tackle the above issues by developing characterization and fitting methods for Markovian Arrival Processes (MAPs), a class of Markovian models developed by Neuts [28] that encompasses MMPP and PH processes as special cases. We describe the properties of the interarrival time (IAT) process of a MAP and use these properties to derive accurate fitting algorithms for network traffic.

There are several works in the literature that have focused on fitting Markovian traffic models by exactly parameterizing MAPs/MMPPs with two or three states [8], [11], [16], [17], [23]. The small state space minimizes the costs of queueing analysis, but it places significant assumptions on the form of the autocorrelations. For instance, a MMPP(2) cannot fit negative autocorrelations, while the MAP(2) autocorrelation function is geometrically decreasing with constant decay rate [17].

In [3] Andersen and Nielsen develop a general fitting algorithm to model LRD traffic traces by superposition of several MMPP(2) sources [14]. The algorithm matches first and second order descriptors of the counting process, i.e., the mean traffic rate, the Hurst parameter, and the lag-1 autocorrelation in counts, has low computational costs and captures the properties of the classic Bellcore LRD traces [1], [24].

Following a different approach, Horváth and Telek [20] consider the multifractal traffic model of Riedi et al. [29], and obtain a class of MMPPs which exhibits multifractal behavior [32]. According to this result, one may fit network traffic by first computing an unnormalized Haar wavelet transform of the trace and then by determining the MMPP process which best matches the variance of the wavelet coefficients at different time scales. Simulation results on the Bellcore Aug89 trace show that this algorithm achieves better accuracy than the superposition method in [3], but at the expense of a larger state space.

Recently, several research efforts [16], [17], [21]–[23] are directed toward the accurate fitting of the IAT process instead of the counting process that is considered in [3] and [20]. IATs can be harder to measure than counts [18], but simple analytical expressions are available for their moments and lag correlations [15]. Instead, only the first three moments of a counting process are known and yet only the first two are easy to manipulate [14]. Several authors have shown that fitting the mean, coefficient of variation and autocorrelations

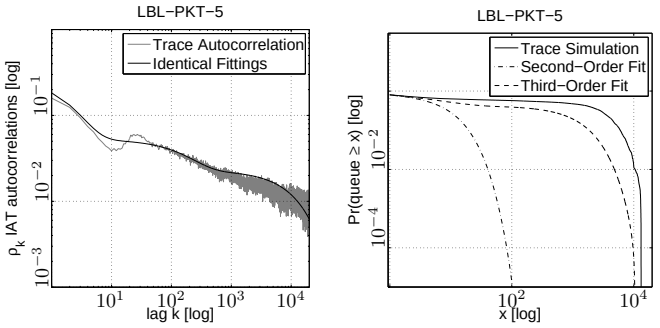

Fig. 1. Autocorrelation and MAP/M/1 queueing behavior (util. 80%) of two MAP(32) fittings of the LBL-PKT-5 trace [1]. The MAPs have identical first and second order properties of the interarrival process, but one has also an accurate fitting of third order properties for which the other is instead loose.

is insufficient to predict queueing behavior [4], [13], [25], therefore fitting the higher order properties of the IAT process seems a natural way to achieve increased prediction accuracy.

To build intuition on the importance of higher order properties we first present an experiment on the LBL-PKT-5 trace of the Internet Traffic Archive [1]. Figure 1 shows two different MAP models we obtained for this traffic trace. The two MAPs have identical first and second order properties of the IAT process, namely same mean, same coefficient of variation (CV), and same autocorrelation function. Mean and CV of IATs are identical to the sample values, the autocorrelation fit is also quite good. However, one model also matches the third order statistics, i.e., the skewness and the bispectrum [9] of the sample IAT process, while the other has a quite loose fit of these descriptors. The strikingly different queueing predictions of the two models, as shown in Figure 1, stress the importance of higher order properties in network traffic.

In this paper, we consequently propose to fit network traffic using higher order properties of the IAT process in addition to the usual first and second order descriptors. Because of the general difficulty in imposing even basic autocorrelations to the IAT process, we first derive characterization results that simplify fitting. Starting from these characterization results, we obtain two MAP fitting algorithms: an efficient hyperexponential moment matching algorithm which applies to traffic that is approximately renewal and a more versatile algorithm based on nonlinear optimization that still matches moments, but can also accurately fit general traffic. The latter algorithm is based on a new MAP definition technique, called Kronecker Product Composition (KPC), which is able to generate MAPs with predefined moments, autocorrelations, and higher order statistics in IATs. Compared to the state-of-the-art fitting methods in [3] and [20], the proposed algorithm shows improved queueing prediction accuracy at similar computational costs. In addition, it does not place limitations on the number and order of statistical properties that can be matched for a traffic trace, thus enabling the selection of the best cost-accuracy tradeoff.

Our detailed contributions can therefore be summarized as follows:

1) MAP(n) Characterization: After reviewing the IAT process in MAPs, in Section III we propose a general spectral characterization of IAT moments, autocorrelations, and higher order moments. These observations clarify the capabilities of general MAPs, provide necessary conditions for fitting, and simplify the analysis of small processes.

2) Hyperexponential Fitting: We develop in Section IV the moment matching algorithm for hyperexponential traffic and illustrate its accuracy using real traffic traces of the Internet Traffic Archive [1]. The algorithm is also instrumental to the general traffic fitting algorithm presented later in the paper.

3) Compositional definition of MAP(n): In Section V we propose the compositional method based on Kronecker products that can easily generate a MAP(n) with predefined properties of IATs from the composition of smaller processes, e.g., MAP(2)s. While traditional superposition is convenient only for imposing first and second order properties of counts, our method is more flexible and gives complete control of the IAT statistics at all orders.

4) MAP Fitting: Exploiting the previous results, we develop in Section VI the general fitting algorithm which first determines the optimal values of IAT moments, autocorrelations, and higher order descriptors using nonlinear optimization, and successively finds the MAP which best matches these target values. The approach is numerically stable and the fitting is usually performed in a few minutes. Comparative analyses in Section VII on the Bellcore Aug89 trace and on Seagate Web trace [30] show that our algorithm achieves increased accuracy with respect to existing methods.

Section VIII draws final conclusions. The final appendix reports the MAPs used to fit the traces discussed in Section VII.

II. IAT Process in MAPs

A MAP(n) is specified by two $n \times n$ matrices: a stable matrix $\boldsymbol{D}_0$ with nonnegative off-diagonal entries and a nonnegative matrix $\boldsymbol{D}_1$ that describe transition rates between n states. Each transition in $\boldsymbol{D}_1$ produces a job arrival; $\boldsymbol{D}_0$ describes instead background transitions not associated with arrivals. The matrix $\boldsymbol{Q} = \boldsymbol{D}_0 + \boldsymbol{D}_1$ is the infinitesimal generator of the underlying Markov process. In the special case where $\boldsymbol{D}_1$ is a diagonal matrix, the process is a MMPP(n).

We henceforth focus on the interval stationary process that describes the IATs. For a MAP(n), this is described by the embedded discrete-time chain with stochastic matrix $\boldsymbol{P} = (-\boldsymbol{D}_0)^{-1}\boldsymbol{D}_1$, with probability vector $\boldsymbol{\pi}_e$, $\boldsymbol{\pi}_e\boldsymbol{e} = 1$, where $\boldsymbol{e}$ is a column vector of 1's of the appropriate dimension. Let $\boldsymbol{P}$ be irreducible with a simple unit eigenvalue $\gamma_1 = 1$. Then, its IAT is phase-type distributed with k-th moment

$$E[X^k] = k!\boldsymbol{\pi}_e(-\boldsymbol{D}_0)^{-k}\boldsymbol{e}, \quad k \geq 0, \tag{1}$$

which implies that $\mathrm{CV}^2 = 2E[X]^{-2}\boldsymbol{\pi}_e(-\boldsymbol{D}_0)^{-2}\boldsymbol{e} - 1$. The lag-$k$ autocorrelation coefficient is

$$\rho_k = (E[X]^{-2}\boldsymbol{\pi}_e(-\boldsymbol{D}_0)^{-1}\boldsymbol{P}^k(-\boldsymbol{D}_0)^{-1}\boldsymbol{e} - 1)/\mathrm{CV}^2. \tag{2}$$

Higher order moments of the IAT process can be described in terms of *joint moments*. Let X_i be the i-th IAT from an arbitrary starting epoch $i_0 = 0$, and consider a sequence $X_{i_1}, X_{i_2}, \ldots, X_{i_L}$, where $0 \leq i_1 < i_2 < \ldots < i_L$. The moments of L consecutive IATs are given by

$$H(\vec{i}, \vec{k}) = E[X_{i_1}^{k_1} X_{i_2}^{k_1} \cdots X_{i_L}^{k_1}],$$

where $\vec{i} = (i_1, i_2, \ldots, i_L)$ and $\vec{k} = (k_1, k_2, \ldots, k_L)$. The moments $H(\vec{i}, \vec{k})$ capture nonlinear temporal relations between samples and are known to completely characterize a MAP [4], [33]. They are computed as [33]

$$H(\vec{i}, \vec{k}) = \boldsymbol{\pi}_e \left(\prod_{l=1}^{L} k_l! (-\boldsymbol{D}_0)^{-k_l} \boldsymbol{P}^{i_l - i_{l-1}} \right) \boldsymbol{e}, \quad (3)$$

where for $l = 1$, i_0 is set to $i_0 = 0$. Noting that it is always $\boldsymbol{P}^{i_L - i_{L-1}} \boldsymbol{e} = \boldsymbol{e}$, (3) reduces in the case $L = 1$ to (1).

In the remaining of this paper and if not otherwise stated, MAP descriptors refer to the IAT process. Further, we use the notation $(\boldsymbol{D}_0, \boldsymbol{D}_1)$ or $((-\boldsymbol{D}_0)^{-1}, \boldsymbol{P})$ to uniquely specify a MAP.

III. Characterization of MAP(n) Processes

We now obtain a spectral characterization of IAT moments and autocorrelations, i.e., a scalar representation of (1)-(2) based on spectral properties of $(-\boldsymbol{D}_0)^{-1}$ and $\boldsymbol{P}$. This simplifies the analysis of MAP moments and autocorrelations, as we show with some examples.

A. Characterization of Moments

We begin by describing the moments (1) in terms of the spectrum of $(-\boldsymbol{D}_0)^{-1}$. Recall that the characteristic polynomial of a $n \times n$ matrix $\boldsymbol{A}$ is

$$\phi(\boldsymbol{A}) = s^n + \alpha_1 s^{n-1} + \ldots + \alpha_{n-1} s + \alpha_n, \quad (4)$$

which is a polynomial in s with roots s_i equal to the eigenvalues of $\mathbf{A}$. Consider the Cayley-Hamilton theorem [19], by which the powers of $\boldsymbol{A}$ satisfy

$$\boldsymbol{A}^k = - \sum_{j=1\ldots n} \alpha_j \boldsymbol{A}^{k-j}, \quad k \geq n \quad (5)$$

that is, matrix powers are linearly dependent. Because MAP moments are computed in (1) from powers of $(-\boldsymbol{D}_0)^{-1}$, they are linearly dependent.

Lemma 1. *In a MAP(n), any $n+1$ consecutive moments are linearly dependent according to the relation*

$$E[X^k] = - \sum_{j=1\ldots n} \left(\frac{m_j k!}{(k-j)!} \right) E[X^{k-j}],\ E[X^0] = 1,\ k \geq n, \quad (6)$$

where m_j is the coefficient of s^{n-j} in $\phi((-\boldsymbol{D}_0)^{-1})$.

Proof: Using the Cayley-Hamilton theorem,

$$E[X^k] = -k! \boldsymbol{\pi}_e (\textstyle\sum_{j=1\ldots n} m_j (-\boldsymbol{D}_0)^{-(k-j)}) \boldsymbol{e} \quad (7)$$

which immediately proves the lemma by (1). ∎

Since the coefficients m_j are functions of the eigenvalues of $(-\boldsymbol{D}_0)^{-1}$ we can derive a closed-form formula for $E[X^k]$.

Theorem 1. *Let $(-\boldsymbol{D}_0)^{-1}$ have $m \leq n$ distinct eigenvalues $\theta_t \in \mathbb{C}$, $1 \leq t \leq m$. Let q_t be the algebraic multiplicity of θ_t, $\sum_{t=1\ldots m} q_t = n$. Then the IAT moments are given by*

$$E[X^k] = \sum_{t=1\ldots m} k!\, \theta_t^k \sum_{j=1\ldots q_t} M_{t,j} k^{j-1}, \quad (8)$$

$$E[X^0] = \sum_{t=1\ldots m} M_{t,1} = 1, \quad (9)$$

where the constants $M_{t,j}$'s are independent of k. In particular,

$$M_{t,1} = \boldsymbol{\pi}_e (-\boldsymbol{D}_0)_t^{-1} \boldsymbol{e}, \quad (10)$$

where $(-\boldsymbol{D}_0)_t^{-1}$ is the t-th spectral projector of $(-\boldsymbol{D}_0)^{-1}$, i.e., the product of the right and left eigenvectors for θ_t.

Proof: Denoting by $(-\boldsymbol{D}_0)_t^{-1}$ and $\boldsymbol{M}_t$ the spectral projector and nilpotent matrix of $(-\boldsymbol{D}_0)^{-1}$ associated to the Jordan block for θ_t, the generalized spectral decomposition of $(-\boldsymbol{D}_0)^{-1}$ is [2]

$$(-\boldsymbol{D}_0)^{-1} = \sum_{t=1\ldots m} (\theta_t (-\boldsymbol{D}_0)_t^{-1} + \boldsymbol{M}_t), \quad k \geq 0$$

where it is $\boldsymbol{M}_t^{q_t} = \mathbf{0}$, $\boldsymbol{M}_t (-\boldsymbol{D}_0)_t^{-1} = (-\boldsymbol{D}_0)_t^{-1} \boldsymbol{M}_t$, $\boldsymbol{M}_t (-\boldsymbol{D}_0)_p^{-1} = \mathbf{0}$ and $(-\boldsymbol{D}_0)_t^{-1} (-\boldsymbol{D}_0)_p^{-1} = \mathbf{0}$, $t \neq p$. Therefore, for all $k \geq 0$ we have

$$\begin{aligned}(-\boldsymbol{D}_0)^{-k} &= \Big(\sum_{t=1\ldots m} (\theta_t (-\boldsymbol{D}_0)_t^{-1} + \boldsymbol{M}_t) \Big)^k \\ &= \sum_{t=1\ldots m} (\theta_t (-\boldsymbol{D}_0)_t^{-1} + \boldsymbol{M}_t)^k \\ &= \sum_{t=1\ldots m} \theta_t^k \sum_{i=0}^{\min\{q_t - 1, k\}} \binom{k}{i} (-\boldsymbol{D}_0)_t^{-1} (\theta_t^{-1} \boldsymbol{M}_t)^i,\end{aligned}$$

where we used in the last passage that $\theta_t \neq 0$ being $(-\boldsymbol{D}_0)^{-1}$ an invertible matrix. Inserting the last formula for $(-\boldsymbol{D}_0)^{-k}$ into (8) we get after some manipulations

$$E[X^k] = \sum_{t=1\ldots m} \theta_t^k \sum_{i=1}^{\min\{q_t, k+1\}} \binom{k}{i-1} \widehat{M}_{t,i}, \quad (11)$$

where

$$\widehat{M}_{t,i} = \boldsymbol{\pi}_e ((-\boldsymbol{D}_0)^{-1})_t (\theta_t^{-1} \boldsymbol{M}_t)^{i-1} \boldsymbol{e}. \quad (12)$$

The last expression is equivalent to (1) by expanding the binomials and grouping the coefficients of k^j. This yields the following equivalence

$$M_{t,j} = \sum_{i=j}^{q_t} \frac{s(i-1, j-1)}{(i-1)!} \widehat{M}_{t,i}. \quad (13)$$

where the $s(m, n)$ is the Stirling number of the first kind giving the coefficient of x^n in $x(x-1)(x-2)\cdots(x-m+1)$. Finally, the condition $\sum_t M_{t,1} = 1$ is obtained by evaluating (8) for $k = 0$ and noting that $E[X^0]$ is always $E[X^0] = 1$. ∎

Corollary 1. *If θ_t has algebraic multiplicity $q_t = 1$, then $M_{t,j} = 0$ for $j \geq 2$.*

Proof: In this case the nilpotent $\boldsymbol{M}_t$ of the t-th Jordan block is zero and in (13) the only non-zero projector is $M_{t,1}$. ∎

Note that formula (8) is more general than a standard spectral decomposition of $(-\boldsymbol{D}_0)^{-1}$ since it also holds for defective, i.e., non-diagonalizable, $(-\boldsymbol{D}_0)^{-1}$. This is extremely important, since well-known processes, e.g., the Erlang process, have $\boldsymbol{D}_0$ that is not diagonalizable and thus the usual spectral decomposition based on this assumption cannot be applied.

Example 1. We show how to apply Theorem 1 for the analytical characterization of a MAP. Consider the MAP(3)

$$\boldsymbol{D}_0 = \begin{bmatrix} -2\lambda & \lambda & \lambda \\ 0 & -\lambda & \lambda \\ 0 & 0 & -\lambda \end{bmatrix}, \boldsymbol{D}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \lambda & 0 & 0 \end{bmatrix}, \ \lambda \geq 0.$$

The left eigenvector of $\boldsymbol{P}$ for $\gamma_1 = 1$ is $\boldsymbol{\pi}_e = [1, 0, 0]^T$. Since

$$(-\boldsymbol{D}_0)^{-k} = \begin{bmatrix} 2^{-k}\lambda^{-k} & (1-2^{-k})\lambda^{-k} & k\lambda^{-k} \\ 0 & \lambda^{-k} & k\lambda^{-k} \\ 0 & 0 & \lambda^{-k} \end{bmatrix},$$

from (1) it is $E[X^k] = k!\boldsymbol{\pi}_e(-\boldsymbol{D}_0)^{-k}\boldsymbol{e} = (k+1)!\lambda^{-k}$. However this approach does not generalize easily to larger models, because obtaining a closed-form expression for $(-\boldsymbol{D}_0)^{-k}$ on larger examples can be difficult. We show that the spectral characterization can analyze $E[X^k]$ without the need of a closed-form formula for $(-\boldsymbol{D}_0)^{-k}$. We first compute $E[X] = 2\lambda^{-1}$ and $E[X^2] = 6\lambda^{-2}$, and observe that the eigenvalues of $(-\boldsymbol{D}_0)^{-1}$ are $\theta_1 = (2\lambda)^{-1}$ and $\theta_2 = \lambda^{-1}$ with multiplicity $q_1 = 1$ and $q_2 = 2$. Imposing $E[X]$ and $E[X^2]$ in (8), we find $M_{1,1} = 0, \quad M_{2,1} = 1 - M_{1,1} = 1, \quad M_{2,2} = 1$, and substituting back we finally get $E[X^k] = k!\theta_1^{-k}M_{1,1} + k!\theta_2^{-k}(M_{2,1} + kM_{2,2}) = (k+1)!\lambda^{-k}$.

A consequence of Lemma 1 is that one can study the necessary conditions for the feasibility of $\boldsymbol{D}_0$ without explicit computation of the eigenvalues θ_t. Because stability of $\boldsymbol{D}_0$ is granted if and only if $\mathrm{Re}(-\theta_t^{-1}) < 0, 1 \leq t \leq m$, we can use the Routh-Hurwitz test [**?**] on $\phi((-\boldsymbol{D}_0)^{-1})$ to study when a given set of moments ensures stability. This is shown in the following example.

Example 2. The coefficients m_j's of a MAP(2) are obtained solving (6) for $n = \{2, 3\}$ and are given by

$$m_1 = \frac{E[X^3] - 3E[X]E[X^2]}{3(2E[X]^2 - E[X^2])}, m_2 = \frac{\frac{3}{2}E[X^2]^2 - E[X]E[X^3]}{3(2E[X]^2 - E[X^2])}$$

The Routh-Hurwitz table for the characteristic polynomial of $(-\boldsymbol{D}_0)^{-1}$ has first column $[1, m_1, m_2]^T$. The stability of $\boldsymbol{D}_0$ follows if the coefficients m_j have alternating sign, implying that $\mathrm{Re}(\theta_t) > 0 \Rightarrow \mathrm{Re}(-\theta_t^{-1}) < 0$. From the formulas for m_1 and m_2 we see that in a MAP(2) this requires:

$$\mathrm{CV}^2 > 1 : E[X^3] > \max(3E[X]E[X^2], \tfrac{3}{2}E[X^2]^2E[X]^{-1})$$
$$\mathrm{CV}^2 < 1 : E[X^3] < \min(3E[X]E[X^2], \tfrac{3}{2}E[X^2]^2E[X]^{-1}).$$

Similar conditions can be obtained for larger processes, e.g., the stability of the MAP(3) can be studied as a function of its first $2n - 1 = 5$ moments.

We also observe that if $(-\boldsymbol{D}_0)^{-1}$ is diagonalizable, then $m = n$ and the projectors $M_{t,1}$ are in simple relation to the IAT cdf since

$$F(x) = 1 - \boldsymbol{\pi}_e e^{\boldsymbol{D}_0 x}\boldsymbol{e} = 1 - \sum_{t=1\ldots n} M_{t,1}e^{-x/\theta_t}, \qquad (14)$$

which follows by the relation $e^{\mathrm{diag}(-\theta_1^{-1},\ldots,-\theta_n^{-1})} = \mathrm{diag}(e^{-\theta_1^{-1}}, \ldots, e^{-\theta_n^{-1}})$ and the computational formula for $M_{t,1}$. Note that (14) allows an efficient numerical computation of useful quantities such as the percentiles of the IAT distribution.

B. Characterization of Autocorrelation

The spectral characterization can be extended to autocorrelations using the properties of the powers $\boldsymbol{P}^k$ in (2).

Lemma 2. *In a MAP(n), any $n+1$ consecutive autocorrelations are linearly dependent according to the relation*

$$\rho_k = -\sum_{j=1\ldots n} a_j\rho_{k-j}, \ \rho_0 = \left(1 - 1/\mathrm{CV}^2\right)/2, \ k \geq n, \quad (15)$$

where a_j is the coefficient of s^{n-j} in $\phi(\boldsymbol{P})$ and $\sum_{j=1}^n a_j = 0$.

Proof: We want to prove that $\sum_{j=0\ldots n} a_j\rho_{k-j} = 0$, where $a_0 = 1$. By definition of ρ_k, this is equivalent to prove

$$\sum\nolimits_j a_j(\boldsymbol{\pi}_e(-\boldsymbol{D}_0)^{-1}\boldsymbol{P}^{k-j}(-\boldsymbol{D}_0)^{-1}\boldsymbol{e} - E[X]^2) = 0.$$

The last equation follows if we can show that $\sum_{j=0}^n a_j\boldsymbol{P}^{k-j} = 0$ and $\sum_{j=0}^n a_j = 0$. But the former holds true by the Cayley-Hamilton theorem, while the latter follows by the stochasticity of $\boldsymbol{P}$, since for the unit eigenvalue $\gamma_1 = 1$ it is $\phi(\boldsymbol{P}) = 0 = \sum_{j=0}^n a_j$. This proves $\rho_k = -\sum_{j=1\ldots n} a_j\rho_{k-j}$. The formula for ρ_0 follows by evaluating (2) for $k = 0$, i.e.,

$$\rho_0 = (E[X]^{-2}\boldsymbol{\pi}_e(-\boldsymbol{D}_0)^{-2}\boldsymbol{e} - 1)/\mathrm{CV}^2 = \left(1 - 1/\mathrm{CV}^2\right)/2.$$

since $\boldsymbol{\pi}_e(-\boldsymbol{D}_0)^{-2}\boldsymbol{e} = E[X^2]/2 = (1 + \mathrm{CV}^2)E[X]^2/2$. ■

Similarly to Theorem 1, we can obtain a closed-form expression of ρ_k.

Theorem 2. *Let $\gamma_t \in \mathbb{C}$, $1 \leq t \leq m$, be an eigenvalue of $\boldsymbol{P}$ with algebraic multiplicity r_t. If $\gamma_t = 0$ assume that its geometric multiplicity equals its algebraic multiplicity, i.e., the r_t associated Jordan blocks have all order one. Then the autocorrelation function of a MAP is*

$$\rho_k = \sum_{t=2\ldots m} \gamma_t^k \sum_{j=1\ldots r_t} A_{t,j}k^{j-1}, \quad k \geq 1 \qquad (16)$$

$$\rho_0 = \sum_{t=2\ldots m} A_{t,1} = (1 - 1/\mathrm{CV}^2)/2, \qquad (17)$$

where the $A_{t,j}$'s constants are independent of k. In particular,

$$A_{t,1} = E[X]^{-2}\boldsymbol{\pi}_e(-\boldsymbol{D}_0)^{-1}\boldsymbol{P}_t(-\boldsymbol{D}_0)^{-1}\boldsymbol{e}/\mathrm{CV}^2, \qquad (18)$$

in which $\boldsymbol{P}_t$ is the t-th spectral projector of $\boldsymbol{P}$, that is, the product of the right and left eigenvectors associated to γ_t.

Proof: The proof is similar to the proof of Theorem 1. Let us assume first that $\gamma_t \neq 0$ for all t. If γ_t has multiplicity r_t, the generalized spectral decomposition of $\boldsymbol{P}$ gives [2]

$$\boldsymbol{P} = \sum_{t=1\ldots m} (\gamma_t\boldsymbol{P}_t + \boldsymbol{N}_t), \quad k \geq 0$$

where $\boldsymbol{N}_t$ is the nilpotent matrix associated to γ_t, $\boldsymbol{N}_t^{r_t} = \boldsymbol{0}$, $\boldsymbol{N}_t\boldsymbol{P}_t = \boldsymbol{P}_t\boldsymbol{N}_t$, and $\boldsymbol{N}_t\boldsymbol{P}_p = \boldsymbol{0}$, $t \neq p$. Therefore,

$$\boldsymbol{P}^k = \Big(\sum_{t=1\ldots m} (\gamma_t\boldsymbol{P}_t + \boldsymbol{N}_t)\Big)^k = \sum_{t=1\ldots m} (\gamma_t\boldsymbol{P}_t + \boldsymbol{N}_t)^k$$
$$= \sum_{t=1\ldots m} \gamma_t^k \sum_{i=0}^{\min\{r_t-1,k\}} \binom{k}{i} \boldsymbol{P}_t(\gamma_t^{-1}\boldsymbol{N}_t)^i, \quad k \geq 0.$$

Inserting the last formula for $\boldsymbol{P}^k$ into (16) we get after algebraic manipulations

$$\rho_k = \sum_{t=1\dots m} \gamma_t^k \sum_{i=1}^{\min\{r_t,k+1\}} \binom{k}{i-1} \widehat{A}_{t,i},$$

where

$$\widehat{A}_{t,i} = E[X]^{-2}\boldsymbol{\pi}_e(-\boldsymbol{D}_0)^{-1}\boldsymbol{P}_t(\gamma_t^{-1}\boldsymbol{N}_{0,t})^{i-1}(-\boldsymbol{D}_0)^{-1}\boldsymbol{e}/\mathrm{CV}^2.$$

Grouping the coefficients of k^j, we have

$$A_{t,j} = \sum_{i=j}^{r_t} \frac{s(i-1,j-1)}{(i-1)!}\widehat{A}_{t,i}, \tag{19}$$

where the $s(m,n)$'s are the Stirling number of the first kind. Finally, the relation for $A_{t,1}$ follows immediately by (19). ■

From Lemma 2 we see that the function ρ_k when evaluated in $k=0$ assumes the value $\rho_0 = (1-1/\mathrm{CV}^2)/2$. Although this coefficient does not admit any statistical interpretation, since the autocorrelation function is by definition $\rho_k = 1$ for $k=0$, it is useful to consider this limit value since the condition $\rho_0 = \sum_t A_{t,1}$ can simplify the computation of projectors. The value ρ_0 can also help in manipulating the autocorrelation coefficients, since it is often observed that increasing ρ_0 produces a generalized increase of all autocorrelations. For instance, in the special case of a MAP(2) process, it follows from (16) that $\rho_k = \gamma_2^k\rho_0$ and therefore the autocorrelations increases monotonically as a function of ρ_0.

Corollary 2. *If γ_t has algebraic multiplicity $r_t = 1$, then $A_{t,j} = 0$ for $j \geq 2$.*

Proof: If all nilpotents $\boldsymbol{N}_t$ are zero, then the only non-zero projector in (19) is $A_{t,1}$. ■

Without loss of generality, we assume in the rest of the paper that $|\gamma_j| \geq |\gamma_{j+1}|$, $j = 1,\dots,n-1$. According to this ordering, the asymptotic decay of the autocorrelation function is geometric with rate γ_2 (unless $\gamma_2 = -1$ and ρ_k does not converge to zero as $k \to +\infty$). We complete the analysis in Theorem 2 by studying the following degenerate case.

Corollary 3. *If $\boldsymbol{P}$ has r_0 eigenvalues equal to zero and belonging to Jordan blocks of order $l_0^1, l_0^2, \dots, l_0^q$, then*

$$\rho_k = \sum_{j=1\dots r_0} \eta_{k,j} + \sum_{t=2\dots m-1} \gamma_t^k \sum_{j=1\dots r_t} A_{t,j}k^{j-1},$$

where

$$\eta_{k,j} = E[X]^{-2}\boldsymbol{\pi}_e(-\boldsymbol{D}_0)^{-1}(\boldsymbol{N}_j)^k(-\boldsymbol{D}_0)^{-1}\boldsymbol{e}/\mathrm{CV}^2, \tag{20}$$

in which $\boldsymbol{N}_{0,j}$, $\boldsymbol{N}_{0,j}^{l_0^j} = \boldsymbol{0}$, is the nilpotent associated to the Jordan block of order l_0^j, and $\eta_{k,j}$ is equal to zero for $k \geq l_0^j$.

Proof: The generalized spectral decomposition of $\boldsymbol{P}$ is

$$\boldsymbol{P}^k = \Big(\sum_{j=1\dots r_0}\boldsymbol{N}_{0,j} + \sum_{t=1\dots m-1}(\gamma_t\boldsymbol{P}_t + \boldsymbol{N}_t)\Big)^k =$$

$$= \sum_{j=1\dots r_0}\boldsymbol{N}_{0,j}^k + \sum_{t=1\dots m-1}\gamma_t^k \sum_{i=0}^{\min\{r_t-1,k\}}\binom{k}{i}\boldsymbol{P}_t(\gamma_t^{-1}\boldsymbol{N}_t)^i,$$

for $k \geq 0$. The rest of the proof is similar to the proof of Theorem 2. ■

We conclude by remarking that the distinct $A_{t,j}$'s and γ_t's in (16) are no more than $2n-2$. Thus a MAP(n) process can fit up to $2n-2$ independent autocorrelations ρ_k, $k \geq 0$. For a given CV^2, ρ_0 is fixed and the maximum number of independent coefficients becomes $2n-3$.

We now present three examples illustrating respectively:

- the application of the previous results to the characterization of two classes of MAP(3) and MAP(4) that can be employed to match traces with complex eigenvalues in autocorrelations (Example 3);
- the computation of the projectors $A_{t,j}$ in the difficult case of a process with defective $\boldsymbol{P}$ (Example 4);
- A comparison of the non-negligible impact of a defective $\boldsymbol{P}$ on the autocorrelations (Example 5).

Example 3. The Circulant MMPP is proposed in [25] to insert complex eigenvalues in the autocorrelation of counts. According to our results, this approach can be generalized to the IAT process by simply defining a MAP with circulant $\boldsymbol{P}$ and/or $(-\boldsymbol{D}_0)^{-1}$. In particular, if $\boldsymbol{D}_0$ is diagonal, the resulting MAP admits a quite simple characterization. Define $[p_1, p_2, \dots, p_n]$, $p_n = 1 - \sum_{j\neq n} p_j$, to be the first column of the circulant matrix $\boldsymbol{P}$. Since in a circulant $\boldsymbol{P}$ it is $\boldsymbol{\pi}_e = \boldsymbol{e}/n$, from (1) we have $E[X^k] = \left(n^{-1}k!\right)\sum_t \theta_t^k$.

Using Theorem 2 we can also study autocorrelations. For instance, in the case $n=3$ the circulant matrix has two identical or complex conjugate eigenvalues, which implies from the condition on $A_{2,1} + A_{3,1} = \rho_0$ that $A_{2,1} = A_{3,1} = \rho_0/2$. Now letting $\gamma_2 = |\gamma_2|e^{j\omega_2}$,

$$|\gamma_2| = (1/2)\sqrt{(3p_1-1)^2 + 3\Delta_{3,2}^2},\ \omega_2 = \arctan\frac{\sqrt{3}\Delta_{3,2}}{(3p_1-1)},$$

with $\Delta_{i,j} = p_i - p_j$, the autocorrelation is

$$\rho_k = \rho_0|\gamma_2|^k\frac{(e^{jk\omega_2} + e^{-jk\omega_2})}{2} = \rho_0|\gamma_2|^k\cos(k\omega_2).$$

Higher order cases are similar. E.g., for $n=4$ after few manipulations we get

$$\rho_k = A_{a,1}\gamma_a^k + A_{b,1}|\gamma_b|^k\cos(k\omega_b),$$

with

$$\gamma_a = \Delta_{1,2} + \Delta_{3,4},\ |\gamma_b| = \sqrt{\Delta_{4,2}^2 + \Delta_{1,3}^2},\ \omega_b = \arctan\frac{\Delta_{4,2}}{\Delta_{1,3}},$$

$$A_{a,1} = \frac{\mathrm{CV}^2(\theta_4 - \theta_2 + \theta_3 - \theta_1)^2}{(\theta_3+\theta_2+\theta_4+\theta_1)^2},\quad A_{b,1} = \rho_0 - A_{a,1},$$

where the eigenvalues are denoted by the indices a and b since the asymptotic decay rate $|\gamma_2|$ can be either $|\gamma_a|$ or $|\gamma_b|$.

Example 4. Consider the MAP(5) with $(-\boldsymbol{D}_0)^{-1} = \mathrm{diag}(1,2,3,4,5)$ and

$$\boldsymbol{P} = \begin{bmatrix} 0 & \lambda & \lambda & \lambda & \lambda \\ \lambda & 0 & \lambda & \lambda & \lambda \\ 0 & 2\lambda & 0 & \lambda & \lambda \\ 0 & 0 & 3\lambda & 0 & \lambda \\ 0 & 0 & 0 & 4\lambda & 0 \end{bmatrix}, \boldsymbol{J} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix},$$

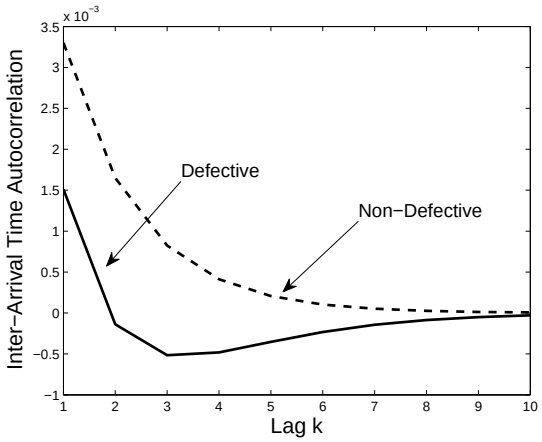

Fig. 2. Autocorrelations for the defective and non-defective $\boldsymbol{P}$ of Example 5

where $\lambda = 0.25$ and $\boldsymbol{J}$ is the Jordan canonical form of $\boldsymbol{P}$. $\boldsymbol{P}$ is defective with $\gamma_2 = \lambda = 0.25$, $r_2 = 4$, and the autocorrelation coefficients are given by

$$\rho_k = \gamma_2^k(A_{2,1} + A_{2,2}k + A_{2,3}k^2 + A_{2,4}k^3), \quad A_{2,1} = \rho_0.$$

The computation of the spectral projectors and nilpotent matrices of $\boldsymbol{P}$ is performed following the same steps of the proof of Theorem 2, see [2] for computational formulas for the projectors $\boldsymbol{P}_t$ and the nilpotents $\boldsymbol{N}_t$. In particular, the $A_{j,t}$ constants are obtained by (19) and are given by $A_{2,1} = 8.160 \cdot 10^{-2}$, $A_{2,2} = -2.843 \cdot 10^{-1}$, $A_{2,3} = 1.565 \cdot 10^{-1}$, $A_{2,4} = -2.076 \cdot 10^{-2}$.

Example 5. The defective stochastic matrix $\boldsymbol{P}_{def}$ with Jordan canonical form

$$\boldsymbol{J} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1 \\ 0 & 0 & 0 & 1/2 \end{bmatrix}, \boldsymbol{V} = \begin{bmatrix} 1 & -2 & -4 & -6 \\ 1 & 19 & -5 & 5 \\ 1 & -3 & 4 & 24 \\ 1 & 3 & 7 & -23 \end{bmatrix},$$

where $\boldsymbol{J} = \boldsymbol{V}^{-1}\boldsymbol{P}_{def}\boldsymbol{V}$, has two eigenvalues $\gamma_2 = 1/2$, $\gamma_3 = 0$ with multiplicities $r_2 = 2$, $r_3 = 1$. Consider the MAP(4) $(\boldsymbol{D}_0 = \mathrm{diag}(-72, -17, -58, -52), -\boldsymbol{D}_0\boldsymbol{P}_{def})$, imposing ρ_1 and ρ_2 in (16) we get $A_{2,1} = 6.5978 \cdot 10^{-3}$, $A_{2,2} = -3.5763 \cdot 10^{-3}$, $A_{3,1} = \rho_0 - A_{2,1} = 1.9669 \cdot 10^{-1}$, and the autocorrelation is

$$\rho_k^{def} = (A_{2,1} + A_{2,2}k)\gamma_2^k.$$

If we consider a similar non-defective $\boldsymbol{P}_{ndef}$ with same eigenvalues and $\boldsymbol{V}$ matrix, but Jordan canonical form

$$\boldsymbol{J} = \boldsymbol{V}^{-1}\boldsymbol{P}_{ndef}\boldsymbol{V} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{bmatrix},$$

the related MAP $(\boldsymbol{D}_0, -\boldsymbol{D}_0\boldsymbol{P}_{ndef})$ has a different autocorrelation despite the fact that the eigenvalues and their multiplicities are identical. Imposing ρ_1 and ρ_2, we get $A_{2,1} = 6.5978 \cdot 10^{-3}$, $A_{2,2} = 0.0$, $A_{3,1} = \rho_0 - A_{2,1} = 1.9669 \cdot 10^{-1}$, thus

$$\rho_k^{ndef} = A_{2,1}\gamma_2^k.$$

A comparison of ρ_k^{def} and ρ_k^{ndef} is shown in Figure 2.

C. Higher-Order Statistics

We conclude by observing that the characterization given for moments and autocorrelations generalizes in a similar fashion to the joint moments (3), since these functions consist of powers of $(-\boldsymbol{D}_0)^{-1}$ and $\boldsymbol{P}$. For example, in the case where both matrices are diagonalizable and $L = 2$, we have

$$H(\vec{i}, \vec{k}) = E[X_{i_1}^{k_1} X_{i_2}^{k_2}] = \sum_{t=1\ldots,n} \sum_{l=1\ldots,n} H_{t,l}\theta_t^{k_t}\gamma_l^{i_l}, \quad (21)$$

where the joint moment projector $H_{t,l}$ is a constant independent of $\vec{i}$ and $\vec{k}$ and which is computed from the product of the spectral projectors of $(-\boldsymbol{D}_0)^{-1}$ and $\boldsymbol{P}$.

From (3) it can be seen that for general L the joint moment projector is not in simple relation with the projectors $M_{j,t}$ and $A_{j,t}$, since it is obtained by first multiplying several projectors $((-\boldsymbol{D}_0)^{-1})_t$ and $\boldsymbol{P}_t$ and then weighting the result using the $\boldsymbol{\pi}_e$ probabilities. Therefore, moment and autocorrelation fitting algorithms, which impose the eigenvalues θ_t and γ_l and the projectors $M_{t,j}$ and $A_{t,j}$, still leave degrees of freedom to assign the projectors of higher order moments.

IV. Hyperexponential Fitting

The high variability of packet transmission delays makes it easy to find network traffic traces with $\mathrm{CV}^2 > 1$, and most of the traffic traces in the Internet Traffic Archive [1] have this property. For this class of processes, our characterization results immediately suggest an efficient hyperexponential fitting algorithm, see [10] for previous work. This will be important in the general fitting algorithm presented at the end of the paper, but can also be used independently to fit a hyperexponential PH process of a network traffic trace with $\mathrm{CV}^2 > 1$. Note that here we focus only on fitting the sample distribution, since we assume that the process is renewal.

Observing that the $M_{t,j}$'s and θ_t's are $n + m \leq 2n$ parameters subject to $\sum_t M_{t,1} = 1$, we see from (8) that a MAP(n) can fit up to $2n-1$ independent moments. A similar conclusion has been recently obtained by the analysis of the underlying PH-type distribution [33]. Given $2n - 1$ sample moments, it is possible to fit $E[X^k]$ by a hyperexponential PH process using (6)-(8) by the algorithm in Figure 4. According to the computational formulas for $M_{j,t}$ in the proof of Theorem 1, the diagonal form of $\boldsymbol{D}_0$ implies that $\boldsymbol{M} = \boldsymbol{\pi}_e$ and the moments have the form $E[X^k] = k!\sum_t \pi_{e,t}\theta_t^k$. As a result, the PH process $(\boldsymbol{D}_0, -\boldsymbol{D}_0\boldsymbol{e}\boldsymbol{M})$ is renewal since $\boldsymbol{P} = \boldsymbol{e}\boldsymbol{M} = \boldsymbol{e}\boldsymbol{\pi}_e$ has rank one and thus $\gamma_t = 0$, $t \geq 2$.

Example 6. Using the moment matching algorithm in Figure 4, we determine a hyperexponential process PH(3) which fits the first five moments of the LBL-PKT-5 trace. The moment matching algorithm returns the following values:

$$M_{1,1} = 0.0262134, \quad M_{2,1} = 0.3847399, \quad M_{3,1} = 0.5890467,$$
$$\theta_{1,1} = 0.0200142, \quad \theta_{2,1} = 0.0090486, \quad \theta_{3,1} = 0.0022153.$$

The central picture in Figure 3 shows the accurate ccdf fitting obtained by the PH(3) with the above projectors and eigenvalues. The first five moments are matched exactly; higher order

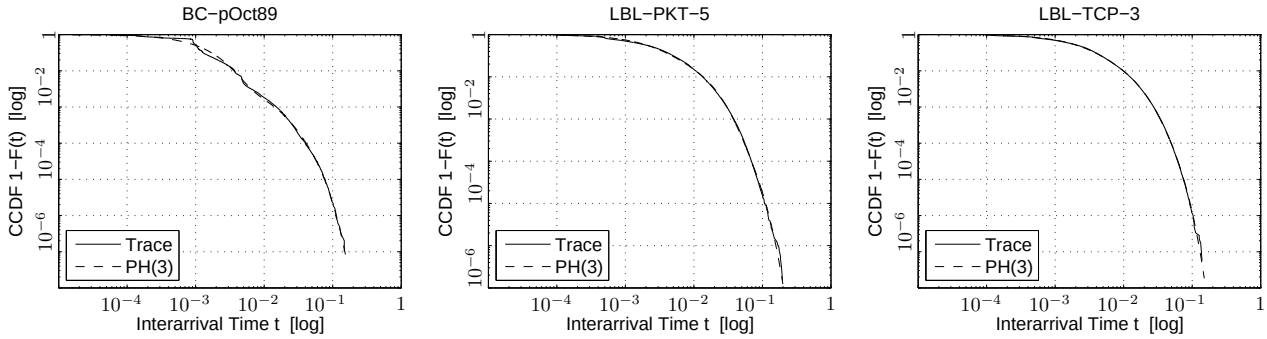

Fig. 3. Distribution fitting of three traces of the Internet Traffic Archive [1] using the algorithm in Figure 4. The computational costs are negligible.

Fig. 4. HYPEREXPONENTIAL MOMENT MATCHING ALGORITHM

Step 1. Use $2n-1$ consecutive sample moments $E[X], E[X^2], \ldots, E[X^{2n-1}]$ to parameterize the linear system composed by $E[X^0] = 1$ and (6) for $k = n, n+1, \ldots, 2n-1$. If the system is singular or $n = 1$ the PH(n) is exponential with mean $E[X]$. Otherwise go to Step 2.

Step 2. Solve for the values m_j and determine the roots of $\phi_{(-\boldsymbol{D}_0)^{-1}} = s^n + m_1\, s^{n-1} + \ldots + m_{n-1}s + m_n$ which are the eigenvalues θ_t of $(-\boldsymbol{D}_0)^{-1}$. If any θ_t is complex or a non-positive real, then drop $E[X^{2n-1}]$ and $E[X^{2n-2}]$ and restart with a $PH(n-1)$.

Step 3. Parameterize the linear system defined by (8) for $k = 0, \ldots, n$ and solve for the $M_{j,t}$'s. If any $M_{j,t}$ is complex or a non-positive real, then drop $E[X^{2n-1}]$ and $E[X^{2n-2}]$ and restart with a $PH(n-1)$.

Step 4. Return the process $(\boldsymbol{D}_0, -\boldsymbol{D}_0\boldsymbol{e}\boldsymbol{M})$ with $\boldsymbol{D}_0 = \mathrm{diag}(-\theta_{1,1}^{-1}, \ldots, -\theta_{n,1}^{-1})$ and $\boldsymbol{M} = [M_{1,1}, \ldots, M_{n,1}]$.

moments are also matched closely, e.g., the sample moment $E[\widehat{X}^8] = 2.7044 \cdot 10^{-11}$ is well approximated by the PH(3) which has $E[X^8] = 2.7908 \cdot 10^{-11}$. Similar quality levels are achieved also for the BC-pOct89 and LBL-TCP-3 traces also shown in Figure 3. Note that, because these traffic traces are all significantly autocorrelated, fitting the underlying distribution, although accurately, is clearly insufficient for predicting their queueing behavior. For this reason, we focus in the rest of the paper on the more difficult problem of fitting the traffic process and its temporal dependence structure. Nevertheless, the algorithm presented in this section can be successfully adopted if the observed network traffic is approximately renewal and is also fundamental to the general fitting algorithm presented in Section VI.

V. COMPOSITIONAL DEFINITION OF LARGE PROCESSES

The accurate fitting of LRD network traffic requires models composed by many states; e.g., the MAP fittings of the Bellcore Aug89 trace in [3] and [20] employ $n = 16$ and $n = 32$ states, respectively. Since traditional superposition is not meant to impose higher order properties of the IAT process, we define a different process composition method which we call Kronecker Product Composition (KPC). Given J MAPs $\{\boldsymbol{D}_0^j, \boldsymbol{D}_1^j\}$, we define the KPC process as the MAP

$$\{\boldsymbol{D}_0^{\mathrm{kpc}}, \boldsymbol{D}_1^{\mathrm{kpc}}\} = \{(-1)^{J-1}\boldsymbol{D}_0^1 \otimes \cdots \otimes \boldsymbol{D}_0^J, \boldsymbol{D}_1^1 \otimes \cdots \otimes \boldsymbol{D}_1^J\}$$

where $\otimes$ is the Kronecker product operator [6]. It can be easily shown by the properties of the Kronecker product that $\boldsymbol{P}^{\mathrm{kpc}} = -(\boldsymbol{D}_0^{\mathrm{kpc}})^{-1}\boldsymbol{D}_1^{\mathrm{kpc}} = \boldsymbol{P}^1 \otimes \cdots \otimes \boldsymbol{P}^J$ and $\boldsymbol{\pi}_e^{kpc} = \boldsymbol{\pi}_e^1 \otimes \cdots \otimes \boldsymbol{\pi}_e^J$, thus our composition generates an embedded process $\boldsymbol{P}^{\mathrm{kpc}}$ with simple structure.

In order to generate a valid MAP, the KPC requires that at least $J-1$ composing processes have diagonal $\boldsymbol{D}_0^j$ as we discuss later in Example 7. Nevertheless, because one MAP can be arbitrary, the KPC does not place modeling restrictions.

The basic property of a MAP obtained by KPC is that we can easily impose its eigenvalues and projectors in both moments and autocorrelations as we show later in Theorem 3. Equivalently, one may impose directly moments and autocorrelation values, as described in Theorem 4. This is important because, by the characterization in Section III, the fitting of real traffic is essentially an inverse eigenvalue problem for the eigenvalues of $\boldsymbol{P}$ and $(-\boldsymbol{D}_0)^{-1}$. Inverse eigenvalue problems are notoriously hard, but the KPC method provides an effective solution. A MAP(n) can be defined to assume an arbitrary number of autocorrelation and moment values, with the only practical difficulty of limiting the order of the resulting MAP. In the rest of the section, we show how one can a priori determine moments and autocorrelations of the KPC process given the knowledge of the properties of the composing MAPs.

A. KPC Process Characterization

Without loss of generality we study $\{\boldsymbol{D}_0^{\mathrm{kpc}}, \boldsymbol{D}_1^{\mathrm{kpc}}\}$ for the case $J = 2$. The results presented here recursively characterize also the case $J > 2$.

Theorem 3. *Let MAP$_a$ $= \{\boldsymbol{D}_0^a, \boldsymbol{D}_1^a\}$ and MAP$_b$ $= \{\boldsymbol{D}_0^b, \boldsymbol{D}_1^b\}$ be MAPs of order n_a and n_b, respectively, and assume that $\boldsymbol{D}_0^b$ is a diagonal matrix. Let γ_p^a, θ_p^a, $A_{p,1}^a$ and $M_{q,1}^a$ be the eigenvalues and projectors of MAP$_a$. Let γ_q^b, θ_q^b, $A_{p,1}^b$ and $M_{q,1}^b$ be the equivalent descriptors of MAP$_b$. Then the KPC*

$$\textit{MAP}_a \otimes \textit{MAP}_b = \{-\boldsymbol{D}_0^a \otimes \boldsymbol{D}_0^b, \boldsymbol{D}_1^a \otimes \boldsymbol{D}_1^b\}$$

is a MAP of order $n_a n_b$ with eigenvalues $\gamma_t = \gamma_p^a \gamma_q^b$, $\theta_t = \theta_p^a \theta_q^b$, and projectors

$$M_{t,1} = M_{p,1}^a M_{q,1}^b, \quad A_{t,1} = (A_{p,1}^a \mathrm{CV}_a^2)(A_{q,1}^b \mathrm{CV}_b^2)/\mathrm{CV}^2,$$

for all $1 \le p \le n_a$, $1 \le q \le n_b$.

Proof: The relations for the eigenvalues follow from basic properties of the Kronecker product [6]. The projector associated to $\theta_t = \theta_p^a \theta_q^b$ is

$$\begin{aligned} M_{t,1} =& \boldsymbol{\pi}_e((-\boldsymbol{D}_0)^{-1})_t \boldsymbol{e} \\ =& (\boldsymbol{\pi}_e^a \otimes \boldsymbol{\pi}_e^b)((-\boldsymbol{D}_0)^{-1})_p^a \otimes ((-\boldsymbol{D}_0)^{-1})_q^b (\boldsymbol{e}^a \otimes \boldsymbol{e}^b) \\ =& (\boldsymbol{\pi}_e^a ((-\boldsymbol{D}_0)^{-1})_p^a \boldsymbol{e}^a)(\boldsymbol{\pi}_e^b ((-\boldsymbol{D}_0)^{-1})_q^b \boldsymbol{e}^b) = M_{p,1}^a M_{q,1}^b. \end{aligned}$$

Similarly, the projector of $\gamma_t = \gamma_p^a \gamma_q^b$ is

$$\begin{aligned} A_{t,1} =& E[X]^{-2} \boldsymbol{\pi}_e (-\boldsymbol{D}_0)^{-1} \boldsymbol{P}_t (-\boldsymbol{D}_0)^{-1} \boldsymbol{e}/\mathrm{CV}^2, \\ =& E[X]^{-2} (E[X^a]^2 A_{p,1}^a \mathrm{CV}_a^2)(E[X^b]^2 A_{q,1}^b \mathrm{CV}_b^2)/\mathrm{CV}^2, \\ =& (A_{p,1}^a \mathrm{CV}_a^2)(A_{q,1}^b \mathrm{CV}_b^2)/\mathrm{CV}^2. \end{aligned}$$

■

Theorem 4. *Moments and autocorrelations of the KPC satisfy*

$$E[X^k] = E[X_a^k] E[X_b^k]/k!, \tag{22}$$

$$CV^2 \rho_k = (CV_a^2)\rho_k^a + (CV_b^2)\rho_k^b + (CV_a^2 CV_b^2)\rho_k^a \rho_k^b, \tag{23}$$

where the quantities in the right-hand side refer to MAP^a and MAP^b. In particular the relation for $E[X^k]$ immediately implies

$$1 + CV^2 = (1 + CV_a^2)(1 + CV_b^2)/2. \tag{24}$$

Proof: We begin by proving (22). Using the properties of the Kronecker product [6] we have

$$\begin{aligned} E[X^k] &= k! \boldsymbol{\pi}_e (-\boldsymbol{D}_0)^{-k} \boldsymbol{e} \\ &= k! (\boldsymbol{\pi}_e^a \otimes \boldsymbol{\pi}_e^b)(-((-1)^{2-1} \boldsymbol{D}_0^a \otimes \boldsymbol{D}_0^b))^{-k} (\boldsymbol{e}^a \otimes \boldsymbol{e}^b) \\ &= k! (\boldsymbol{\pi}_e^a \otimes \boldsymbol{\pi}_e^b)((\boldsymbol{D}_0^a)^{-k} \otimes (\boldsymbol{D}_0^b)^{-k})(\boldsymbol{e}^a \otimes \boldsymbol{e}^b), \end{aligned}$$

and multiplying by $(-1)^{-2k}$ which equals one for all $k \in \mathbb{N}$

$$\begin{aligned} E[X^k] &= k! (-1)^{-2k} (\boldsymbol{\pi}_e^a (\boldsymbol{D}_0^a)^{-k} \boldsymbol{e}^a)(\boldsymbol{\pi}_e^b (\boldsymbol{D}_0^b)^{-k} \boldsymbol{e}^b) \\ &= k! (\boldsymbol{\pi}_e^a (-\boldsymbol{D}_0^a)^{-k} \boldsymbol{e}^a)(\boldsymbol{\pi}_e^b (-\boldsymbol{D}_0^b)^{-k} \boldsymbol{e}^b) \\ &= E[X_1^k] E[X_2^k]/k!. \end{aligned}$$

Equation (23) follows the same steps as (22) by considering (2). ■

The two theorems provide a complete characterization of first and second order IAT properties of the KPC process. The KPC also simplifies the definition of higher order statistics. We have this characterization of the joint moments of the composed process.

Theorem 5. *The joint moments of MAP$_a \otimes$ MAP$_b$ satisfy*

$$H(\vec{i}, \vec{k}) = \frac{H^a(\vec{i}, \vec{k}) H^b(\vec{i}, \vec{k})}{k_1! k_2! \cdots k_L!}, \tag{25}$$

being $H^a(\vec{i}, \vec{k})$ and $H^a(\vec{i}, \vec{k})$ the joint moments of $\{\boldsymbol{D}_0^a, \boldsymbol{D}_1^a\}$ and $\{\boldsymbol{D}_0^b, \boldsymbol{D}_1^b\}$, respectively.

Proof: The proof follows similarly to the proof of Theorem 4 by considering (3). ■

Example 7. To motivate the assumption on the diagonal structure of the $\boldsymbol{D}_0^j$'s, consider the KPC of the MAP(2)

$$\boldsymbol{D}_0^1 = \begin{bmatrix} -3\alpha & \alpha \\ 2\alpha & -4\alpha \end{bmatrix}, \boldsymbol{D}_1^1 = \begin{bmatrix} \alpha & \alpha \\ \alpha & \alpha \end{bmatrix} \tag{26}$$

with the MMPP(2)

$$\boldsymbol{D}_0^2 = \begin{bmatrix} -2\beta & \beta \\ \beta & -3\beta \end{bmatrix}, \boldsymbol{D}_1^2 = \begin{bmatrix} \beta & 0 \\ 0 & 2\beta \end{bmatrix}. \tag{27}$$

None of the two processes has diagonal $\boldsymbol{D}_0^j$, thus the KPC is the infeasible MAP(4) with negative off-diagonal entries

$$\boldsymbol{D}_0 = \begin{bmatrix} -6\delta & 3\delta & 2\delta & -\delta \\ 3\delta & -9\delta & -\delta & 3\delta \\ 4\delta & -2\delta & -8\delta & 4\delta \\ -2\delta & 6\delta & 4\delta & -12\delta \end{bmatrix}, \boldsymbol{D}_1 = \begin{bmatrix} \delta & 0 & \delta & 0 \\ 0 & 2\delta & 0 & 2\delta \\ \delta & 0 & \delta & 0 \\ 0 & 2\delta & 0 & 2\delta \end{bmatrix},$$

where $\delta = \alpha\beta$. If any process has diagonal $\boldsymbol{D}_0^j$, the negative anti-diagonal of $\boldsymbol{D}_0$ becomes zero and the result is feasible.

Example 8. It is known that MAP(2) and MMPP(2) processes have IAT autocorrelation ρ_k that cannot be greater than 0.5, see [17]. The fitting of real traces requires to address this problem, but to the best of our knowledge no examples of MAP(n)s with large ρ_1 have been given in the literature. An example can be easily generate by KPC. Consider a process $\{\boldsymbol{D}_0^a, \boldsymbol{D}_1^a\}$ with quite large autocorrelation, e.g., the MAP

$$\boldsymbol{D}_0^a = \begin{bmatrix} -10001 & 0 \\ 0 & -101 \end{bmatrix}, \boldsymbol{D}_1^a = \begin{bmatrix} 10000 & 1 \\ 1 & 100 \end{bmatrix},$$

which has lag-1 autocorrelation $\rho_1 = 0.485$ and $CV_a^2 = 49.5$. We seek for a MAP^b $\{\boldsymbol{D}_0^b, \boldsymbol{D}_1^b\}$ which may produce $\rho_1 \ge 0.5$ by KPC, and we focus on the case where this process is PH-type renewal, thus $\rho_k^a \equiv 0$, for all k. From (23)-(24) we have

$$\rho_k = \left(\frac{CV_a^2}{CV^2}\right) \rho_k^a = \left(\frac{CV_a^2}{(1 + CV_a^2)(1 + CV_b^2)/2 - 1}\right) \rho_k^a,$$

and to increase the autocorrelation it is sufficient to select a process with CV_b^2 such that $CV_a^2 > CV^2$. For instance,

$$\boldsymbol{D}_0^b = \begin{bmatrix} -5 & 5 & 0 \\ 1 & -2.5 & 1 \\ 0 & 0 & -1 \end{bmatrix}, \boldsymbol{D}_1^b = \begin{bmatrix} 0 & 0 & 0 \\ 0.5 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

is PH-type renewal with $\mathrm{CV}_b^2 = 0.584$, and from (24) we have $\mathrm{CV}^2 = 39.0 < \mathrm{CV}_a^2$. As expected the KPC of the two MAPs yields a MAP(6) with $\rho_1 = 0.616$ which addresses the MAP(2) limit $\rho_k \le 0.5$. Processes with even larger autocorrelation can be obtained with a similar approach.

VI. General MAP Fitting Algorithm

Using KPC, we define a general-purpose fitting algorithm for network traffic. We illustrate the algorithm in the case where the J composing MAPs used in the KPC are an arbitrary MAP(2) (index $j = 1$) and $J - 1$ MAP(2)s with diagonal $\boldsymbol{D}_0$, but the method works with minor modifications also with other processes. The algorithm proceeds in three steps.

Fig. 5. AUTOCORRELATION AND CV^2 FITTING [STEP 1]

minimize $\sum_{k \in K} (\rho_k - \widehat{\rho}_k)^2$

subject to

$(\mathrm{CV}^2 - \widehat{\mathrm{CV}}^2)^2 \leq \mathrm{tol}_{\mathrm{CV}^2}$,

$\mathrm{lb}_{\mathrm{CV}^2}(j) \leq \mathrm{CV}^2(j) \leq \mathrm{ub}_{\mathrm{CV}^2}(j)$, $\forall\ j \in \mathbf{J}$;

$\mathrm{lb}_{\gamma_2}(j) \leq \gamma_2(j) \leq \mathrm{ub}_{\gamma_2}(j)$, $\forall\ j \in \mathbf{J}$;

where

$\widehat{\mathrm{CV}}^2 \leftarrow$ sample CV^2,

$\widehat{\rho_k} \leftarrow$ sample autocorrelation, $\forall\ k \in \mathbf{K}$

$\mathrm{CV}^2 \leftarrow (24)$ recursively using $\mathrm{CV}^2(j)$, $\forall\ j \in \mathbf{J}$,

$\rho_k(j) \leftarrow \frac{1}{2}\left(1 - \frac{1}{\mathrm{CV}^2(j)}\right)\gamma_2(j)^k$, $\forall\ k \in \mathbf{K}, \forall\ j \in \mathbf{J}$;

$\rho_k \leftarrow (23)$ recursively using $\rho_k(j)$, $\forall\ k \in \mathbf{K}, \forall\ j \in \mathbf{J}$.

Fig. 6. MOMENT AND HIGHER-ORDER FITTING [STEP 2]

minimize $\sum_{(\vec{i},\vec{k}) \in \mathbf{H}} (H(\vec{i},\vec{k}) - \widehat{H}(\vec{i},\vec{k}))^2$

subject to

$(E[X] - \widehat{E}[X])^2 \leq \mathrm{tol}_{\mathrm{E[X]}}$,

$(E[X^3] - \widehat{E}[X^3])^2 \leq \mathrm{tol}_{\mathrm{E[X^3]}}$,

$\mathrm{lb}_{\mathrm{E[X]}}(j) \leq E[X](j) \leq \mathrm{ub}_{\mathrm{E[X]}}(j)$, $\forall\ j \in \mathbf{J}$;

$\mathrm{lb}_{\mathrm{E[X^3]}}(j) \leq E[X^3](j) \leq \mathrm{ub}_{\mathrm{E[X^3]}}(j)$, $\forall\ j \in \mathbf{J}$;

where

$\widehat{E}[X] \leftarrow$ sample E[X],

$\widehat{E}[X^3] \leftarrow$ sample $\mathrm{E[X^3]}$,

$E[X] \leftarrow (22)$ recursively using $E[X](j)$, $\forall\ j \in \mathbf{J}$,

$E[X^3] \leftarrow (22)$ recursively using $E[X^3](j)$, $\forall\ j \in \mathbf{J}$,

$E[X^2](j) \leftarrow (1 + SCV(j))(E[X](j))^2$, $\forall\ j \in \mathbf{J}$.

Step 1 - Autocorrelation and CV^2 *Fitting.* Let $\widehat{\mathrm{CV}}^2$ be the sample CV^2; similarly, let $\widehat{\rho}_k$ be the sample autocorrelation computed on a set of lags $\mathbf{K}$, and let $\mathbf{J} = \{1, 2, \ldots, J\}$. We fit second order IAT properties by the nonlinear optimization program in Figure 5. The fitting algorithm is essentially a least-square algorithm constrained to the properties of the KPC. The result of the optimization are two sets $\mathrm{CV}^2(j)$ and $\gamma_2(j)$ for $j \in \mathbf{J}$ which specify the optimal CV^2 and autocorrelation for each of the J MAPs used in the KPC. For each variable, a set of upper and lower bounds are imposed, e.g., $\mathrm{ub}_{\mathrm{CV}^2}(j)$ and $\mathrm{lb}_{\mathrm{CV}^2}(j)$ are respectively upper and lower bounds on the value $\mathrm{CV}^2(j)$ to be determined by the solver. Since $\mathrm{CV}^2(j)$ and $\gamma_2(j)$ are constrained by proper bounds, they can be always chosen to be feasible for a MAP(2), see [16] for existing bound formulas. If the optimization uses processes other than the MAP(2), the feasibility constraints given below need to be adjusted accordingly. In all experiments we set the upper bound on the CV^2 to be $\mathrm{ub}_{\mathrm{CV}^2}(j) = \infty$, $j \in \mathbf{J}$. Further, for the arbitrary MAP(2) we have

$$\mathrm{lb}_{\mathrm{CV}^2}(1) = 0.5, \quad \mathrm{lb}_{\gamma_2}(1) = -1, \quad \mathrm{ub}_{\gamma_2}(1) = 1 - \epsilon,$$

being ϵ an arbitrarily small positive quantity. The remaining $J - 1$ MAP(2)s with diagonal $\boldsymbol{D}_0$ can be shown to have hyperexponential marginal probabilities, and we set

$$\mathrm{lb}_{\mathrm{CV}^2}(j) = 1 + \epsilon, \quad \mathrm{lb}_{\gamma_2}(j) = 0, \quad \mathrm{ub}_{\gamma_2}(j) = 1 - \epsilon.$$

The value $\mathrm{tol}_{\mathrm{CV}^2}$ is a tolerance on the exact matching of the CV^2. On certain traces where the value of the lag-1 autocorrelation ρ_1 differs significantly from $\rho_0 = (1 - 1/\mathrm{CV}^2)/2$, flexibility on the CV^2 fitting avoids an excessive constraining to impose the passage through ρ_1 which can result in bad fitting of autocorrelation at high lags.

Step 2 - Moment and Higher-Order Fitting. Once that the optimal values of $\mathrm{CV}^2(j)$ and $\gamma_2(j)$ are obtained after one or more runs of the previous algorithm, we search for the missing parameters required to define valid MAP(2)s, namely the means $E[X](j)$ and third moments $E[X^3](j)$ for all $j \in \mathbf{J}$. Indeed, the second moments $E[X^2](j)$ are readily obtained from the $\mathrm{CV}^2(j)$ for given $E[X](j)$. As shown by the motivating example in Figure 1, given fixed autocorrelation and CV^2 there exist many possible valid processes; we thus solve a new nonlinear optimization program to select the one that results in better fitting of higher order properties of IATs on a set of sample joint moments $\widehat{H}(\vec{i},\vec{k})$ for $(\vec{i},\vec{k}) \in \mathbf{H}$. The nonlinear program is given in Figure 6. For MAP(2)s we use the following moment bounds [16]

$$\mathrm{lb}_{E[X]}(j) = \sqrt{2E[X^2]}, \quad \mathrm{ub}_{E[X]}(j) = +\infty,$$

$$\mathrm{lb}_{E[X^3]}(j) = \sqrt{(1.5 + \epsilon)E[X^2]^2/E[X]}, \ \mathrm{ub}_{E[X^3]}(j) = +\infty.$$

Step 3 - MAP(n) *Generation.* Given the target optimal values for the $E[X](j), \mathrm{CV}^2(j), E[X^3](j), \gamma_2(j)$ we generate the J MAPs as follows. The $J-1$ diagonal MAPs are usually feasible since the constraints on moments and autocorrelations are sufficient for feasibility [16]. The related values of $M_{t,j}$ and $\theta_{t,j}$ are computed from the first three moments using the fitting algorithm in Section IV and exploiting that $E[X^2](j) = (1 + \mathrm{CV}^2(j))(E[X](j))^2$. The matrix $\boldsymbol{P}(j)$ is immediately specified by the vector $\boldsymbol{M} = [M_{1,1}, 1 - M_{1,1}] = \boldsymbol{\pi}_e$ and by $\gamma_2(j) = \det(\boldsymbol{P}(j))$. For the arbitrary MAP(2) we use standard fitting algorithms, see e.g., [8], [11]. Whenever the fitting results in infeasible processes (e.g., negative rates in $\boldsymbol{D}_1$ or in the off-diagonal elements of $\boldsymbol{D}_0$), we perform a least square fitting to best match the target $E[X^3](j)$ and $\gamma_2(j)$ while keeping fixed $E[X](j)$ and $E[X^2](j)$. Once that J feasible MAPs are obtained, the final process is immediately computed by KPC.

We conclude the section by remarking that with MMPP(2)s/MAP(2)s, the fitting algorithm cannot include complex eigenvalues in the IAT autocorrelations. These may be included by also using one or more circulant MAP(3)s or MAP(4)s such as those described in Example 3, but this may

easily yield processes with several tens or hundreds of states. This state space explosion associated to the use of circulant matrices has been pointed out also in the fitting of the counting process [7] and remains an open problem. However, we empirically observe that many traffic traces that exhibit multiple complex eigenvalues in the counting process often have IAT autocorrelation that does not require complex eigenvalues, and this makes MAP(2)-based IAT fitting sufficient more frequently than counting process-based methods. For instance, Figure 7 compares the Welch power spectrum density (PSD) estimate of the IAT and counting processes on the Bellcore Aug89 trace. The counting process is obtained by computing the arrivals in 10^5 consecutive times slots of identical duration $\Delta T = 10^{-2} sec$. The figure for the counting process indicates power in the low frequency spectrum, whereas the IAT process does not show any significant complex sinusoid and thus can be approximated effectively by real eigenvalues only.

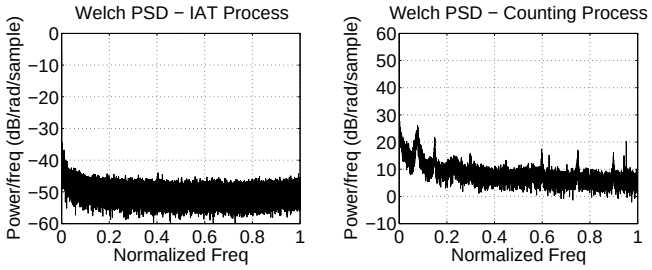

Fig. 7. Comparison of the power spectrum of the IAT process and the counting process for the Bellcore Aug89 trace. The counting process shows power density in the complex spectrum which is instead negligible in IATs.

VII. Experiments

We present a comparison of our algorithm with the best-available algorithms for Markovian analysis of LRD traffic, that is, the method of Andersen and Nielsen (A&N) in [3] and the multifractal approach of Horvath and Telek (H&T) in [20]. We first describe the experimental methodology, later we report fitting results on the Bellcore Aug89 trace [1], [24] and a recently measured Web traffic trace obtained at Seagate and presented in [30].

A. Experimental Methodology

We apply the algorithm described in Section VI as follows. We first fit the autocorrelation on a set of $10^3 - 10^4$ logarithmically-spaced lags ranging in a large interval, e.g., $[1, 10^5]$. Previous work has often limited to match IAT autocorrelations in $[1, 10^4]$, but we have observed that the choice of a larger lag interval can result in increased modeling accuracy at heavy traffic where second order properties are fundamental for queueing prediction [31]. The solution of the least squares program in Figure 5 is usually very efficient (of the order of seconds), and only a few runs are needed for an accurate match. Here we consider four MAPs ($J = 4$); good fitting of the autocorrelation is also possible with only two or three MAPs, but the remaining degrees of freedom are usually insufficient to match accurately higher order properties of IATs.

In the fitting of the joint moments, we have performed several experiments and obtained the best results by matching a set of moments $E[X_{i_1} X_{i_2} X_{i_3}]$, which implicitly define the bispectrum of the IAT process [9]. This seems to be more important than matching moments $E[X_{i_1}^{k_1} X_{i_2}^{k_2}]$ of the IATs, which did not result in improved queueing prediction accuracy with respect to a standard second order fitting. Without loss of generality we set $i_1 = 1$ and fit $E[X_1 X_{i_2} X_{i_3}]$ on a square grid of 10^2 or 25^2 points (i_2, i_3) generated by the Cartesian product of two identical sets of logarithmically-spaced points in $[1, 10^4]$. The point $E[X_1 X_1 X_1] = E[X^3]$ is always included in this grid, thus in Step 2, see Figure 6, we set $\mathrm{tol}_{E[X^3]} = +\infty$ to give more flexibility to the least-squares; in all experiments we instead impose exact matching of $E[X]$ and CV^2, thus $\mathrm{tol}_{E[X]} = 0$ and $\mathrm{tol}_{\mathrm{CV}^2} = 0$. Compared to the autocorrelation, the least square fitting of joint moments seems more difficult and the nonlinear optimizer can occasionally return infeasible solutions. Thus, several runs may be needed to find a good local optimum, which is nevertheless obtained in a few minutes.

The computational costs of the final MAP(n) generation is negligible. We also remark that small corrections of erroneous behaviors are possible without the need of re-running the entire fitting algorithm. For instance, to obtain a slower asymptotic decay rate for the autocorrelations it is possible to increase the value of the largest $\gamma_2(j)$ and regenerate the MAP(n).

Finally, the evaluation of the queueing behavior of the fitted MAP is done with an implementation of the analytical method for the solution of a MAP/D/1 process in [26] and using a numerical tolerance for convergence of $\epsilon = 10^{-10}$. Details on the experimental results are given in the rest of the section.

B. Bellcore Aug89, $-/D/1$ queue

We first compare with the queueing predictions of the models in [3], [20] using the Bellcore Aug89 on a first-come-first-served queue with deterministic service and different utilization levels. This is the standard case for evaluating the quality of LRD trace fitting, e.g., [3], [20], [21]. The traffic trace consists of 10^6 IAT samples collected in 1989 at the Bellcore Morristown Research and Engineering facility and shows a clear LRD behavior, see [24] for details. We run the algorithm described in Section VI to determine a MAP(16) which accurately fits the trace. The size of this MAP is similar to those employed in previous work, which are composed by 16 states (A&N) or 32 states (H&T). Due to the limited length of the trace, we fit all autocorrelations in the interval $[1, 2 \cdot 10^4]$, since at higher lags the sample values are significantly affected by noise. The result of this fit is rather accurate, as shown in Figure 8, and is obtained in less than one minute[1]. In the second phase of the algorithm, the joint moments $E[X_1 X_{i_2} X_{i_3}]$ are matched on a square grid of 25^2 points. On this instance, the computational cost of the program in Figure 6 is low, approximately thirty seconds. The values of the first three moments of the MAP(16) are given in Table

[1] In both figures 8 and 11 we do not report the acf fitting of A&N and H&T since these methods do not match IAT autocorrelations, but autocorrelations in counts.

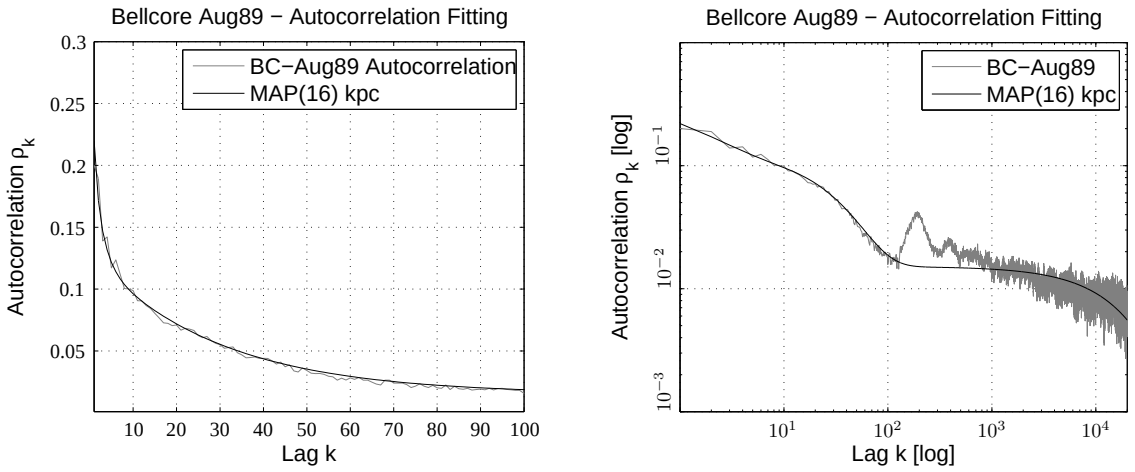

Fig. 8. Fitted autocorrelation for the Bellcore Aug89 trace using the program in Figure 5.

TABLE I
MAP(16) FITTING OF THE BELLCORE AUG89 TRACE USING THE ALGORITHM OF SECTION VI

BC-Aug89	Trace	MAP(16)
$E[X]$	$3.1428 \cdot 10^{-3}$	$3.1428 \cdot 10^{-3}$
CV^2	$3.2236 \cdot 10^{0}$	$3.2235 \cdot 10^{0}$
$E[X^3]$	$2.0104 \cdot 10^{-6}$	$1.1763 \cdot 10^{-5}$
γ_2	n/a	$9.9995 \cdot 10^{-1}$

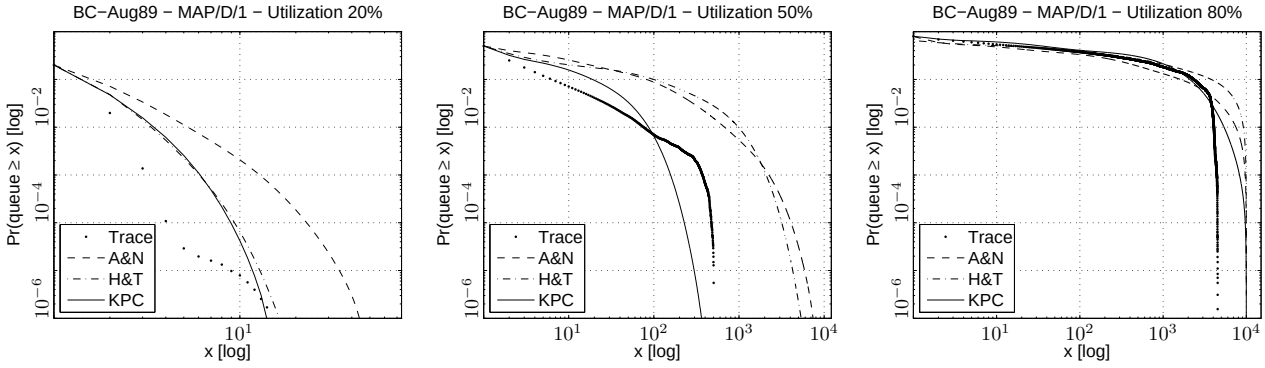

Fig. 9. Queueing predictions for the Bellcore Aug89 trace on a queue with deterministic service.

I; the entries of each composing MMPP(2)/MAP(2) are given in the appendix.

In order to assess the accuracy of the fit, we compare the queueing prediction of our model with the MMPPs obtained in [3], [20] for utilization levels of 20%, 50%, and 80%. All traces have a quite good match of the individual queue probabilities. In Figure 9 we plot the complementary cdf (ccdf) of queue-length probabilities $\Pr(\text{queue} \geq x)$, which accounts also for the residual queueing probability mass and thus shows the impact of the tail probability. At 20% utilization the effects of the long-range dependence seems minimal, and the probability mass is spread over few lags. Our method gives almost the same results of the multifractal technique, while the method of A&N seems to underestimate the queueing probability for the smallest values of x, which also affects the rest of the ccdf.

The intermediate case for 50% utilization is generally difficult to capture, since the network is approaching heavy traffic, but the dependence effects are still not as strong as in slightly higher utilization values, i.e., for $60\%-70\%$ utilization (see, e.g., [3]). All methods initially overestimate the real probability, but for higher values of x our method is closer to the trace values than A&N and H&T which predict a large probability mass also after $x = 10^3$.

Finally, in the case of 80% utilization all three methods perform well, with our algorithm and the H&T being more precise than A&N. The final decay of the curve is again similar, but the KPC method resembles better the simulated trace.

Overall, the result of this trace indicates that the KPC approach seems more effective than both H&T and the A&N methods, while preserving the smallest representation (16 states) of the A&N method. It also interesting to point out that the fitting leaves room for further improvements, especially in the 50% case which is difficult to approximate. This may indicate that significant information about the IAT process may

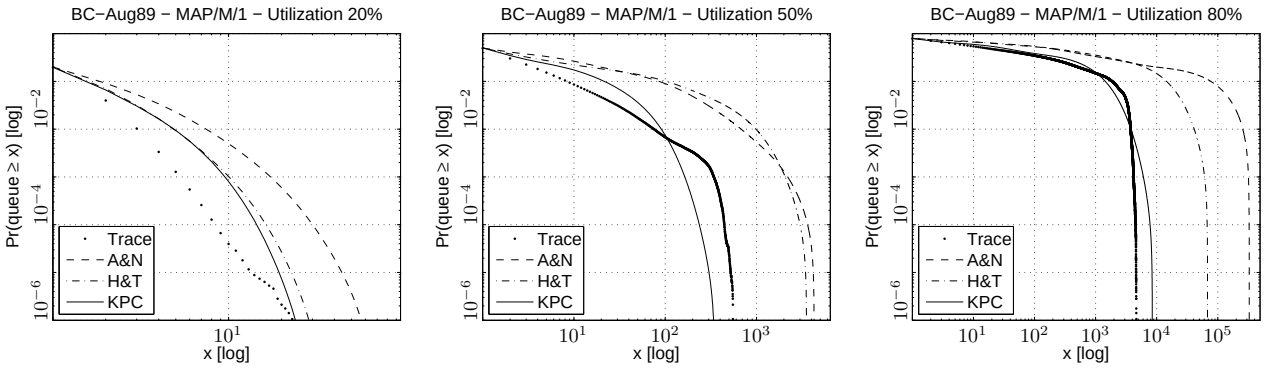

Fig. 10. Queueing predictions for the Bellcore Aug89 trace on a queue with exponential service.

be captured by statistics of higher order than the bispectrum.

C. Bellcore Aug89, $-/M/1$ queue

In the second experiment we evaluate the robustness of the fitting under different variability in the service process. This is important to assess that the fitting captures the essential properties of the traffic process, and thus can provide accurate results regardless of the context in which the fitted MAP is used. In Figure 10 we plot comparative results for a $-/M/1$ queue using in input the same MAPs considered before. As we can see, KPC performs better than in the $-/D/1$ case, and it is now able to capture well the tail decay also for the 80% utilization. A possible explanation of this behavior is that the autocorrelation in the flow becomes more important if the queueing process is more variable, therefore more accurate autocorrelation fitting becomes necessary under such conditions. In comparison, the other methods seem instead to suffer by the increase in variability of the process, as shown by the overestimates which are significantly greater than in the $-/D/1$ case. This indicates that KPC is more robust than counting-process based fittings.

D. Seagate Web Traffic Trace, $-/D/1$ queue

In order to provide a comparison on traffic traces that are representative of other network traffic, we implemented the A&N method[2] and compared its counting process fitting with our method on the HTTP Web traffic trace presented in [30]. The trace is composed by $3.6 \cdot 10^6$ interarrival times of requests at the storage system of a Web server, and has a long-range dependence that is stronger than the BC-Aug89, see [30] for the Hurst coefficient estimates. Thanks to the larger size of the sample, we now fit the autocorrelation in the larger set of lags $[1, 10^5]$ using only 10^3 logarithmically-spaced points since the autocorrelation function is less noisy than for the Bellcore trace, see Figure 11. The joint moments are then fitted on a grid of 10^2 points. The values of the first three moments of the related MAP(16) are reported in Table II; each of the composing MMPP(2)/MAP(2) are given in the appendix.

[2]For this trace we do not report fitting results using the H&T multifractal method because we were unable to implement it. For the Bellcore trace used in the previous subsection, we use the MAP given in [20].

Queueing results for this trace are shown in Figure 12. Here we compare with an implementation of the A&N algorithm [3]. The A&N MAP(16) fitting is obtained by the algorithm parameters $H = 0.85682$, $\rho = 0.74503$, $\lambda^\star = 3.3185$, $n = 5$, $d^\star = 4$.

Although the performance effects of Web traffic on a server is more often modeled by a queue with exponential service, we perform the comparison here assuming a deterministic service time, since the results on the Bellcore trace indicate that this case is more difficult to approximate. Predictions on a $-/D/1$ queue at utilization levels of $20\%, 50\%, 80\%$ are shown in Figure 12. The KPC method is more accurate than the A&N fitting in the cases 50% and 80% while the case 20% is hard to approximate for both methods. This reinforces the validity of the observations on the Bellcore trace: IAT fitting is more effective as soon as the effect of the temporal dependence becomes evident. The 50% and 80% utilization levels for the KPC method are cases of almost perfect fits. In particular, for the 50% case the analytical results indicate that the tail probability is zero with respect to machine accuracy for $x = 61891$, while the simulated queue drops to zero for $x = 61002$.

VIII. Conclusion

We have presented several contributions to the Markovian analysis of network traffic described in terms of packet or request interarrival times. We have obtained a spectral characterization of moments and autocorrelation which simplifies the analysis of MAP processes and proposed a new class of circulant MAPs which exhibit complex eigenvalues in the IAT spectrum. On the basis of this spectral characterization, we have obtained an algorithm for fitting hyperexponential traffic by means of small PH processes. Experimental results on traces of the Internet Traffic Archive indicate high accuracy.

In the second part of the paper, we have studied the definition of large MAPs by Kronecker Product Composition (KPC), and shown that this provides a simple way to create processes with predefined moments and correlations at all orders. A least square fitting procedure based on the properties of these processes has been described. Detailed comparisons with other state-of-the-art fitting methods based on the counting-process show that KPC provides improved fitting of LRD traces that require models that capture their

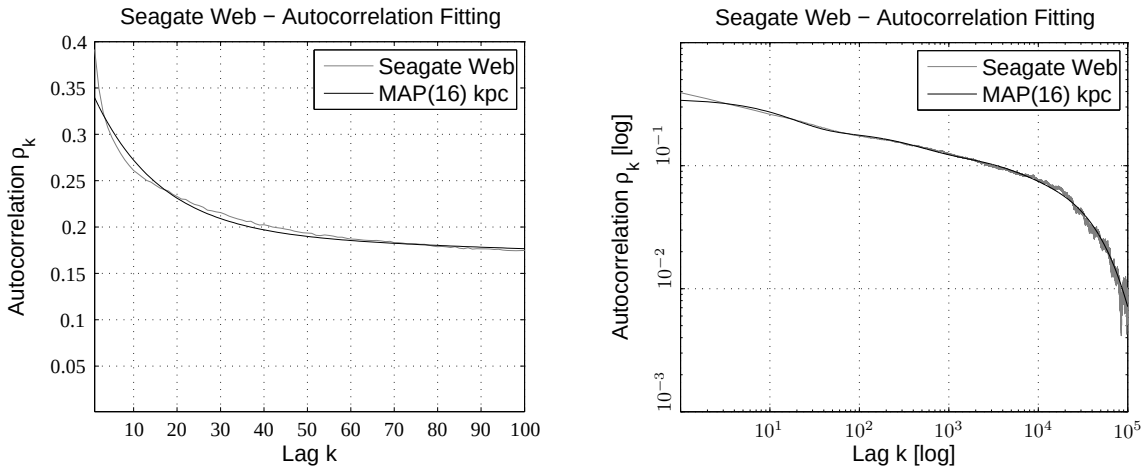

Fig. 11. Fitted autocorrelation for the Seagate Web trace using the program in Figure 5.

TABLE II
MAP(16) FITTING OF THE SEAGATE WEB TRAFFIC TRACE USING THE ALGORITHM OF SECTION VI

Seagate	Trace	MAP(16)
$E[X]$	$3.0134 \cdot 10^0$	$3.0134 \cdot 10^0$
CV^2	$3.3285 \cdot 10^0$	$3.3285 \cdot 10^0$
$E[X^3]$	$1.5986 \cdot 10^3$	$1.1414 \cdot 10^3$
γ_2	n/a	$9.9997 \cdot 10^{-1}$

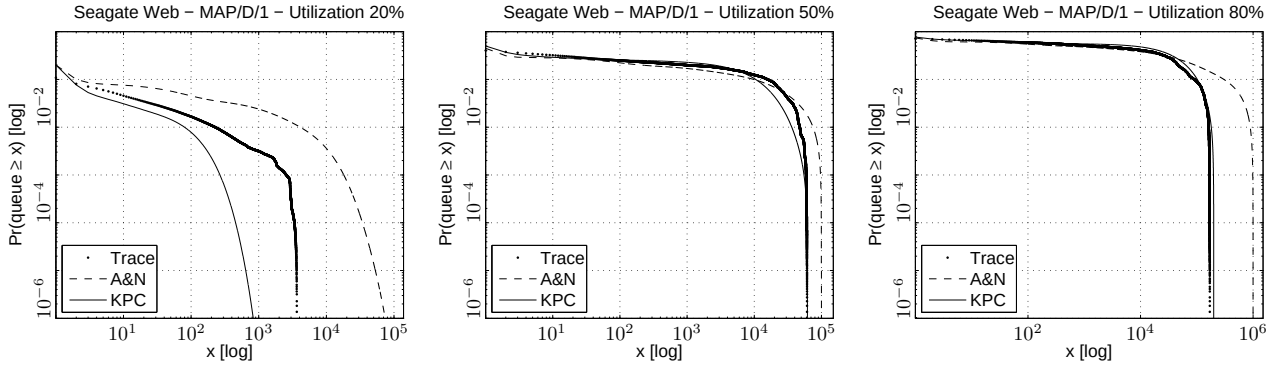

Fig. 12. Queueing predictions for the Seagate Web trace on a queue with deterministic service.

higher-order properties, including the challenging BC-Aug89 trace of the Internet Traffic Archive.

ACKNOWLEDGMENT

A preliminary version of this paper appeared at the Globecom 2007 Workshop on Future Service-oriented Networks, Washington DC. This work was supported by the National Science Foundation under grants ITR-0428330 and CNS-0720699.

The authors thank M. Telek for his comments on an earlier version of this manuscript; B. Van Houdt for his help for the analytical solution of the MAP/D/1 queue; A. Riska for providing the Seagate Web traffic trace; A. Horváth for comments and material related to the multifractal method.

REFERENCES

[1] The internet traffic archive. `http://ita.ee.lbl.gov/`.

[2] *Numerical Methods for Large Eigenvalue Problems*. Manchester University Press, 1992.

[3] A. T. Andersen and B. F. Nielsen. A Markovian approach for modeling packet traffic with long-range dependence. *IEEE JSAC*, 16(5):719–732, 1998.

[4] A. T. Andersen and B. F. Nielsen. On the use of second-order descriptors to predict queueing behavior of MAPs. *Naval Res. Logistics*, 49(4):391–409, 2002.

[5] A.W. Berger. On the index of dispersion for counts for user demand modeling. In *ITU, Madrid, Spain, AT&T Study Group 2, Question 17/2*, June 27-29, 1994.

[6] J.W. Brewer. Kronecker products and matrix calculus in system theory. *IEEE T. Circuits and Sys.*, 25(9), 1978.

[7] H. Che and S. Li. Fast algorithms for measurement-based traffic modeling. In *Proc. of INFOCOM Conf.*, pages 177–186, 1997.

[8] J.E. Diamond and A.S. Alfa. On approximating higher-order MAPs with MAPs of order two. *Queueing Systems*, 34:269–288, 2000.

[9] J. Fan and Q. Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag, New York, USA, 2003.

[10] A. Feldmann and W. Whitt. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. In *Proc. of INFOCOM Conf.*, pages 1096–1104, 1997.

[11] H.W. Ferng and J.F. Chang. Connection-wise end-to-end performance

analysis of queueing networks with MMPP inputs. *Perf. Eval.*, 43(1):39–62, 2001.
[12] M. Grossglauser and J.C. Bolot. On the relevance of long-range dependence in network traffic. In *Proc. of SIGCOMM Conf.*, pages 15–24, 1996.
[13] B. Hajek and L. He. On variations of queue response for inputs with the same mean and autocorrelation function. *IEEE/ACM T. Networking*, 6(5):588–598, 1998.
[14] H. Heffes and D. M. Lucantoni. A markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE JSAC*, 4:856–868, 1986.
[15] A. Heindl. *Traffic-Based Decomposition of General Queueing Networks with Correlated Input Processes*. Ph.D. Thesis, Shaker Verlag, Aachen, 2001.
[16] A. Heindl. Analytic moment and correlation matching for MAP(2)s. In *Proc. of PMCCS Workshop*, pages 39–42, 2003.
[17] A. Heindl, K. Mitchell, and A. van de Liefvoort. Correlation bounds for second-order MAPs with application to queueing network decomposition. *Perf. Eval.*, 63(6):553–577, 2006.
[18] D. P. Heyman and D. Lucantoni. Modeling multiple IP traffic streams with rate limits. *IEEE/ACM T. Networking*, 11(6):948–958, 2003.
[19] F. E. Hohn. *Elementary Linear Algebra*. Dover, 1973.
[20] A. Horváth and M. Telek. Markovian modeling of real data traffic: Heuristic phase type and MAP fitting of heavy tailed and fractal like samples. In *Performance Evaluation of Complex Systems: Techniques and Tools, IFIP Performance 2002, LNCS Tutorial Series Vol 2459*, pages 405–434, 2002.
[21] G. Horváth, P. Buchholz, and M. Telek. A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation. In *Proc. of 2nd Conf. on Quantitative Evaluation of Systems (QEST)*, pages 124–133, 2005.
[22] G. Horváth and M. Telek. A canonical representation of order 3 phase type distributions. *Perf. Eval.*, pages 405–434, 2007.
[23] S.H. Kang, H.K. Yong, D.K. Sung, and B.D. Choi. An application of markovian arrival process (MAP) to modeling superposed ATM cell streams. *IEEE T. Comm.*, 50(4):633–642, 2002.
[24] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of ethernet traffic. *IEEE/ACM T. Networking*, 2(1):1–15, 1994.
[25] S. Li and C. Hwang. Queue response to input correlation functions: continuous spectral analysis. *IEEE/ACM T. Networking*, 1(6):678–692, 1993.
[26] D. M. Lucantoni. New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*, 7:1–46, 1991.
[27] R. D. Nelson. *Probability, Stochastic Processes and Queueing Theory*. Springer-Verlag, 1995.
[28] M. F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. Marcel Dekker, New York, 1989.
[29] R. H. Riedi, M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk. A multifractal wavelet model with application to network traffic. *IEEE T. Inf. Theory*, 45(3):992–1018, 1999.
[30] A. Riska and E. Riedel. Long-range dependence at the disk drive level. In *QEST 2006 Conference*, pages 41–50. IEEE Press, 2006.
[31] K. Sriram and W. Whitt. Characterizing superposition arrival processes in packet multiplexers for voice and data. *IEEE JSAC*, 4(6):833–846, 1986.
[32] M.S. Taqqu, V. Teverovsky, and W. Willinger. Is network traffic self-similar or multifractal? *Fractals*, 63(5):833–846, 1997.
[33] M. Telek and G. Horváth. A minimal representation of Markov arrival processes and a moments matching method. *Perf. Eval.*, 64(9–12):1153–1168, 2007.

APPENDIX

KPC Fitting - Bellcore Aug89 Trace. The Bellcore Aug89 trace is fitted by the KPC process $(\boldsymbol{D}_0^{kpc}, \boldsymbol{D}_1^{kpc})$ defined by

$$\begin{aligned}\boldsymbol{D}_0^{kpc} =& -\boldsymbol{D}_0^a \otimes \boldsymbol{D}_0^b \otimes \boldsymbol{D}_0^c \otimes \boldsymbol{D}_0^d \\ \boldsymbol{D}_1^{kpc} =& \boldsymbol{D}_1^a \otimes \boldsymbol{D}_1^b \otimes \boldsymbol{D}_1^c \otimes \boldsymbol{D}_1^d.\end{aligned}$$

The four composing processes have

$$\boldsymbol{D}_1^a = \begin{bmatrix} 2.5582 \cdot 10^0 & 4.3951 \cdot 10^{-2} \\ 1.1368 \cdot 10^{-2} & 6.6173 \cdot 10^{-1} \end{bmatrix},$$

$$\boldsymbol{D}_1^b = \begin{bmatrix} 2.6769 \cdot 10^0 & 6.6924 \cdot 10^{-5} \\ 4.2706 \cdot 10^{-5} & 1.7082 \cdot 10^0 \end{bmatrix},$$

$$\boldsymbol{D}_1^c = \begin{bmatrix} 4.3309 \cdot 10^0 & 2.7061 \cdot 10^{-4} \\ 6.7564 \cdot 10^{-2} & 2.2578 \cdot 10^{-2} \end{bmatrix},$$

$$\boldsymbol{D}_1^d = \begin{bmatrix} 3.5552 \cdot 10^1 & 2.9355 \cdot 10^{-1} \\ 2.6962 \cdot 10^0 & 4.8230 \cdot 10^0 \end{bmatrix}.$$

and the corresponding $\boldsymbol{D}_0$ are diagonal with i-th element equal in modulus to the sum of the i-th row of the associated $\boldsymbol{D}_1$ matrix, e.g., $\boldsymbol{D}_1^a = \mathrm{diag}(-(2.6769 \cdot 10^0 + 6.6924 \cdot 10^{-5}), -(4.2706 \cdot 10^{-5} + 1.7082 \cdot 10^0))$.

KPC Fitting - Seagate Web Trace. The MAP(16) fitting the Seagate Web trace is the process $(\boldsymbol{D}_0^{kpc}, \boldsymbol{D}_1^{kpc})$ where

$$\begin{aligned}\boldsymbol{D}_0^{kpc} =& -\boldsymbol{D}_0^a \otimes \boldsymbol{D}_0^b \otimes \boldsymbol{D}_0^c \otimes \boldsymbol{D}_0^d \\ \boldsymbol{D}_1^{kpc} =& \boldsymbol{D}_1^a \otimes \boldsymbol{D}_1^b \otimes \boldsymbol{D}_1^c \otimes \boldsymbol{D}_1^d\end{aligned}$$

in which the composing processes have

$$\boldsymbol{D}_1^a = \begin{bmatrix} 6.0174 \cdot 10^{-4} & 1.9726 \cdot 10^{-5} \\ 5.4983 \cdot 10^{-6} & 1.6772 \cdot 10^{-4} \end{bmatrix},$$

$$\boldsymbol{D}_1^b = \begin{bmatrix} 4.7919 \cdot 10^1 & 6.4534 \cdot 10^{-2} \\ 2.8556 \cdot 10^{-2} & 2.1204 \cdot 10^1 \end{bmatrix},$$

$$\boldsymbol{D}_1^c = \begin{bmatrix} 4.4827 \cdot 10^0 & 5.7367 \cdot 10^{-5} \\ 1.6440 \cdot 10^{-5} & 1.2846 \cdot 10^0 \end{bmatrix},$$

$$\boldsymbol{D}_1^d = \begin{bmatrix} 2.9941 \cdot 10^1 & 3.6688 \cdot 10^{-3} \\ 1.9573 \cdot 10^{-3} & 1.5974 \cdot 10^1 \end{bmatrix}.$$

and the corresponding $\boldsymbol{D}_0$ are diagonal with i-th element equal in modulus to the sum of the i-th row of the associated $\boldsymbol{D}_1$ matrix.