Multi-Class Markovian Arrival Processes and Their Parameter Fitting

Peter Buchholz
Department of Computer Science, TU Dortmund, D-44221 Dortmund, Germany
peter.buchholz@tu-dortmund.de

Peter Kemper
Department of Computer Science, College of William and Mary, Williamsburg, VA 23187, USA
kemper@cs.wm.edu

Jan Kriege
Department of Computer Science, TU Dortmund, D-44221 Dortmund, Germany
jan.kriege@tu-dortmund.de

Author prepared version of a paper published in

Performance Evaluation, Volume 67, Issue 11, 2010.

http://dx.doi.org/10.1016/j.peva.2010.08.006

Multi-Class Markovian Arrival Processes and Their Parameter Fitting

Peter Buchholz

peter.buchholz@tu-dortmund.de

Department of Computer Science, TU Dortmund, D-44221 Dortmund, Germany

Peter Kemper

kemper@cs.wm.edu

Department of Computer Science, College of William and Mary, Williamsburg, VA 23187, USA

Jan Kriege

jan.kriege@tu-dortmund.de

Department of Computer Science, TU Dortmund, D-44221 Dortmund, Germany

Abstract

Markovian arrival processes are a powerful class of stochastic processes to represent stochastic workloads that include autocorrelation in performance or dependability modeling. However, fitting the parameters of a Markovian arrival process to given measurement data is non-trivial and most known methods focus on a single class case, where all events are of the same type and only the sequence of interarrival times is of interest. In this paper, we propose a method to fit data to a multi-class Markovian arrival process, where arrivals can be partitioned into a finite set of classes. This allows us to use a Markovian arrival process to represent workloads where interarrival times are correlated across customer classes and to achieve models of greater accuracy. The fitting approach performs in several consecutive steps and applies a single non-linear optimization step and several non-negative least squares computations.

Keywords: Traffic Modeling, Parameter Fitting, Markovian Arrival Process

1 Introduction

An adequate representation of the load of a system is a key aspect of stochastic model building. In many application areas, load modeling is challenging due to significant correlation between interarrival times of load-generating events. This is known for modeling of network traffic [30], disk access patterns [31] and failures in software [11] or hardware systems. In order to include correlations in a workload model, one uses stochastic processes rather than plain distributions. One class of stochastic models that is capable of modeling correlations and complex distributions are Markovian Arrival Processes (MAPs) [17, 27]. In many practical applications, different types of entities arrive and are correlated, e.g., in a network the transfer of a file causes the generation of a sequence of packets of maximal length followed by one short packet; in a software system the occurrence of several soft failures may indicate a hard failure. To model such phenomena, a type of stochastic process is needed that generates different entity classes. A corresponding generalization of MAPs are MAPs with marked arrivals, MMAP[K], where K denotes the number of different types of entities. According to [12], the concept originates from [14, 3]. In [12], He refines the MMAP[K] concept to consider arrivals of batches of customers of K different types and shows that MMAP[K] includes three interesting classes of models. These are 1) the superposition process of K independent Poisson processes, 2) a batch Markovian arrival process (BMAP) where batches of customers of only one class arrive, and 3) a point process with K types of customers where customers arrive individually (batches of size 1). The last case is the one considered in [14, 13]. BMAPs, the second case, have been extensively studied [25]. Batches can be interpreted more generally as different classes. This observation has been formalized also in [23] and has been used for the modeling of IP traffic [32]. In [16], MMAPs are characterized by their conditional moments and are used in a decomposition approach for the analysis of queueing networks.

For better readability, we use MMAP instead of MMAP[K] throughout the rest of the paper. Obviously, a concept like MMAP that allows us to model a workload with multiple types of entities and correlations among entities of the same class as well as correlations among entities of different classes is desirable. In order to make use of it, we must be able to derive an MMAP for given sample data. We need automated methods that perform the necessary parameter fitting. For the case of a simple MAP and some subclasses of processes, several fitting methods have been developed in recent years (e.g., [8, 9]) that perform the parameter fitting of a MAP according to some measured trace. However, even this much simpler case turns out to be a complex optimization problem. Only a very few methods are known to fit the parameters of a BMAP as a specific interpretation of an MMAP. In [6, 22], an Expectation Maximization (EM) algorithm is used with known long runtimes such that it cannot be applied to traffic traces with several million entries. In [32], a two-step approach for discrete time processes is proposed, where first two-state MAPs are fitted to match the interarrival times of the classes which are then superposed and the correlation is added. The problem with the approach is the combinatorial growth of the state space in the number of classes. In [16], a fitting approach for MMAPs is presented to fit MMAPs according to moments and joint moments, like in the approach presented in this paper. However, the approach requires that the moments and joint moments be exactly matched by the MMAP and the fitting results in a representation that is not necessarily a Markov process.

The contribution of this paper is to present an automated fitting method for the parameters of an MMAP that always finds an MMAP that approximates a given set of moments and joint moments. The approach consists of a sequence of steps such that its structure is related to that of [32] for BMAPs, but the individual steps are completely different. The paper is structured as follows. In Section 2, we introduce necessary definitions and notations with respect to MAPs. Section 3 briefly describes some available fitting methods and discusses possible measures for the quality of the fitting. The new MMAP fitting method is presented in Section 4 and Section 5 is devoted to an experimental evaluation. We conclude in Section 6.

2 Basic Definitions and Notations

We begin with some basic definitions and results describing the measured data and the stochastic models, namely, MAPs and MMAPs that are used for modeling data streams.

2.1 Measured Data

We consider traces that consist of a sequence of events where each event happens at a particular point in time and is of a certain type taken from a finite set of types $\mathcal{A}$. For a trace $\mathbf{s}$ of length m, we formally consider two aspects: t_i, the interarrival time, the time distance between events i and $i+1$, and $a_i \in \mathcal{A}$, the type of the i-th entry. Note that interarrival times are measured only between the first and the last event in $\mathbf{s}$, so we restrict our considerations of interarrival times to $t_1, \ldots, t_{m-1}$. We denote by $\mathbf{s}_t$ the trace resulting from $\mathbf{s}$ by neglecting the types. We first define some measures that will be used in the fitting approach. The empirical moments and joint moments of the type independent interarrival time are given by

$$\nu_k = \frac{1}{m-1}\sum_{i=1}^{m-1}(t_i)^k \text{ and } \nu_{k,l} = \frac{1}{m-2}\sum_{i=1}^{m-2}(t_i)^k(t_{i+1})^l \ . \tag{1}$$

Let $X(t)$ be the stochastic process describing the interarrival times. The values ν_k and $\nu_{k,l}$ are estimates for the moments $E(X^k)$ and joint moments $E(X_0^k X_1^l)$ of the interarrival times. Here, a joint moment considers pairs of consecutive times or, in other words, it takes the interarrival time directly before an event (X_0) and directly after that same event (X_1) into account.

We refine the notation to consider only the interarrival times of events that have a particular type a. Thus, let X be equal to $X^{(a)}$ if a type a event occurs. Let $E\left((X^{(a)})^k\right)$ be the kth moment of the interarrival time that follows an event of type a. For joint moments, we are interested in the interarrival times before and after an event of type a. Let $E\left((X_0^{(a)})^k(X_1)^l\right)$ be the joint moment of order (k,l) under the condition that the entity between the two interarrival times was of type a. From trace $\mathbf{s}$ the empirical moments and joint moments of the above type can be computed. Let m_a be the number of events in $\mathbf{s}$ that are of type a.

$$\begin{aligned} \nu_k^a &= \tfrac{1}{m_a - \delta(a_m = a)} \sum_{i=1}^{m-1} \delta(a_i = a)(t_i)^k \text{ and} \\ \nu_{k,l}^a &= \tfrac{1}{m_a - \delta(a_1 = a) - \delta(a_m = a)} \sum_{i=1}^{m-2} \delta(a_{i+1} = a)(t_i)^k(t_{i+1})^l \end{aligned} \tag{2}$$

where $\delta(b)$ is 1 if b is *true* and 0 otherwise.

Observe that all measures we defined describe either a single arrival or two consecutive arrivals and consider the time to the next arrival after observing an arrival of a specific class. These measures result in very efficient fitting algorithms with an appropriate fitting quality as shown below.

2.2 Markovian Arrival Processes

A MAP $X(t)$ of size n is a stochastic process defined by two $n \times n$ matrices $\mathbf{D}_0$ and $\mathbf{D}_1$ such that

- $\mathbf{D}_0(i,j) \geq 0$ for $i \neq j$ and $\mathbf{D}_0(i,i) \leq -\sum_{j=1, i\neq j}^{n} \mathbf{D}_0(i,j)$,
- $\mathbf{D}_1(i,j) \geq 0$ and $\mathbf{D}_0\mathbf{e}^T = -\mathbf{D}_1\mathbf{e}^T$ where $\mathbf{e}$ is the unit row vector and $\mathbf{e}^T$ is its transposition,

- $\mathbf{D}_0$ is non-singular and $\mathbf{D}_0 + \mathbf{D}_1$ is an irreducible generator matrix.

The MAP characterizes a stochastic process where events are generated whenever a transition from $\mathbf{D}_1$ occurs; transitions in $\mathbf{D}_0$ occur as well but are not observable (silent).

Let $\mathbf{M} = (-\mathbf{D}_0)^{-1}$. Matrix $\mathbf{P} = \mathbf{M}\mathbf{D}_1$ is the stochastic transition matrix of the embedded phase process at arrival instants. We assume that $\mathbf{P}$ is irreducible, then the stationary distribution at arrival instants denoted by π is given by $\pi\mathbf{P} = \pi$ and $\pi\mathbf{e}^T = 1.0$. After an event has been generated, the distribution of the MAP is given by π.

The following results can be found in the literature, e.g., in [34, 20]. The moments and joint moments of the MAP are given by

$$\mu_k = E(X^k) = k!\pi\mathbf{M}^k\mathbf{e}^T \text{ and } \mu_{k,l} = E(X_0^k X_1^l) = k!l!\pi\mathbf{M}^k\mathbf{P}\mathbf{M}^l\mathbf{e}^T \ . \tag{3}$$

The lag k autocorrelation is defined as

$$\rho_k = \frac{\mu_1^{-2}\pi\mathbf{M}\mathbf{P}^k\mathbf{M}\mathbf{e}^T - 1}{2\mu_1^{-2}\pi\mathbf{M}^2\mathbf{e}^T - 1} \ . \tag{4}$$

The double transform of the number of arrivals in $(0, t)$ starting with an arrival at time 0 is

$$f(s, z) = \int_{t=0}^{\infty} e^{-st} E(z^{X(t)}) dt = \pi(s\mathbf{I} - \mathbf{D}_0 - z\mathbf{D}_1)^{-1}\mathbf{e}^T \ . \tag{5}$$

According to [34], a MAP is *non-redundant* if the size (i.e., the number of states) equals the order, which is the size of the degree of the denominator of $f(s, z)$. Then we can recognize the following property about the joint density distribution of interarrival times and arriving classes.

Theorem 1. *[34, Theorem 4] The distribution of a non-redundant MAP of size n is determined by at most n^2 parameters, e.g. by $2n - 1$ moments and $(n - 1)^2$ joint moments.*

2.3 Multi-Class Markovian Arrival Processes

The class of MMAPs is essentially the same as what has been defined in [14, 23] and corresponds to BMAPs [25] in continuous time with a different interpretation of the arrivals. An MMAP is a stochastic process where events belong to different classes. Let $\mathcal{A}$ be the finite set of classes and c be the number of classes. Then an MMAP is defined as a Markov process with state space $\{1, \ldots, n\}$, a matrix $\mathbf{D}_0$ and c matrices $\mathbf{D}_1^a$, one for each $a \in \mathcal{A}$. All matrices are $n \times n$ matrices. If we abstract from the classes of arrivals, an MMAP defines a MAP $(\mathbf{D}_0, \mathbf{D}_1)$ with $\mathbf{D}_1 = \sum_{a \in \mathcal{A}} \mathbf{D}_1^a$. Furthermore, we need the condition $\mathbf{D}_1^a \geq \mathbf{0}$.

For a trace we have defined empirical moments and joint moments after an arrival of class a. We now show how to compute the corresponding results for an MMAP (see also [16]). Let π^a be the distribution that follows an arrival of class a. Specifically,

$$\pi^a = \pi\mathbf{M}\mathbf{D}_1^a \ . \tag{6}$$

Obviously $\pi = \sum_{a \in \mathcal{A}} \pi^a$. $p(a) = \pi^a\mathbf{e}^T$ describes the probability that an arbitrary event is of class a such that $\pi^a/p(a)$ yields the normalized version of the distribution describing the conditional distribution under the condition that the event was of class a.

The kth moment of the interarrival time after a class a arrival equals

$$\mu_k^a = E\left((X^{(a)})^k\right) = \frac{k!}{p(a)}\pi \mathbf{M}\mathbf{D}_1^a \mathbf{M}^k \mathbf{e}^T = \frac{k!}{p(a)}\pi^a \mathbf{M}^k \mathbf{e}^T \ . \tag{7}$$

Let $E\left((X_0^{(a)})^k (X_1)^l\right)$ be the joint moment of order k and l under the condition that the arrival between two consecutive times is of class a.

$$\mu_{k,l}^a = E\left((X_0^{(a)})^k (X_1)^l\right) = \frac{k!l!}{p(a)}\pi \mathbf{M}^k \mathbf{M}\mathbf{D}_1^a \mathbf{M}^l \mathbf{e}^T \tag{8}$$

The following theorem extends Theorem 1 to MMAPs (see also [16, Theorem 1]).

Theorem 2. *An MMAP* $(\mathbf{D}_0, \mathbf{D}_1^a \ (a \in \mathcal{A}))$ *of size* n *with* c *classes that describes a non-redundant MAP* $(\mathbf{D}_0, \sum_{a\in\mathcal{A}} \mathbf{D}_1^a)$ *is characterized by at most* cn^2 *parameters.*

Proof: The proof follows the proof of Theorem 4 from [34], extended to the multi-class case. An MMAP is completely characterized by its joint density of the interarrival times and arriving classes.

$$f(t_1, a_1, \ldots, t_m, a_m) = \pi e^{\mathbf{D}_0 t_1} \mathbf{D}_1^{a_1} e^{\mathbf{D}_0 t_2} \mathbf{D}_1^{a_2} \ldots e^{\mathbf{D}_0 t_m} \mathbf{D}_1^{a_m} \mathbf{e}^T$$

where $t_i \geq 0$ and $a_i \in \mathcal{A}$. Since the MMAP is invariant under similarity transformations we can compute the Jordan decomposition $\mathbf{E} = -\mathbf{\Gamma}\mathbf{D}_0\mathbf{\Gamma}^{-1}$ with $\mathbf{\Gamma}\mathbf{e}^T = 1$ and $\mathbf{F}_1^a = \mathbf{\Gamma}\mathbf{D}_1^a\mathbf{\Gamma}^{-1}$ (see [34]) and observe that $\mathbf{\Gamma}^{-1}\mathbf{e}^T = \mathbf{e}^T$ also holds. The joint densities can then be represented by

$$f(t_1, a_1, \ldots, t_m, a_m) = \pi \mathbf{\Gamma}^{-1} \sum_{k_1=0}^{\infty} \tfrac{(\mathbf{E}t_1)^{k_1}}{k_1!} \mathbf{F}_1^a \ldots \sum_{k_m=0}^{\infty} \tfrac{(\mathbf{E}t_m)^{k_m}}{k_m!} \mathbf{F}_1^a \mathbf{e}^T \ .$$

Matrix $\mathbf{E}$ is characterized by the n eigenvalues of $\mathbf{D}_0$ and matrices $\mathbf{F}_1^a$ are determined by $cn^2 - n$ elements since $-\mathbf{E} = \sum_{a\in\mathcal{A}} \mathbf{F}_1^a$. Observe that eigenvalues and matrix elements can be complex. Define $\mathbf{F} = \sum_{a\in\mathcal{A}} \mathbf{F}_1^a$. Vector π is determined by $\mathbf{E}$ and $\mathbf{F}$ since

$$\pi(\mathbf{D}_0)^{-1}\mathbf{D}_1 = \pi\mathbf{\Gamma}^{-1}\mathbf{E}^{-1}\mathbf{\Gamma}\mathbf{\Gamma}^{-1}\mathbf{F}\mathbf{\Gamma} = \pi\mathbf{\Gamma}^{-1}\mathbf{E}^{-1}\mathbf{F}\mathbf{\Gamma} = \pi$$

which implies that $\pi\mathbf{\Gamma}^{-1}$ is the left eigenvector of $\mathbf{E}^{-1}\mathbf{F}$ and $\pi\mathbf{\Gamma}^{-1}\mathbf{e}^T = 1$.

The theorem implies that the representation of an MMAP is redundant since only cn^2 out of $(c+1)n^2$ parameters are required.

3 Related Approaches

For fitting the parameters of a MAP or MMAP according to a measured trace one can use two general approaches [17]. First, it is possible to fit the parameters according to the values in the trace. This is usually done by maximizing the likelihood that the trace has been generated from the fitted MAP. The best method to find parameters that maximize the likelihood are EM algorithms that have been proposed in different variants for parameter fitting of PH distribution or MAPs [6, 22, 4, 21, 35]. However, to the best of our knowledge only two approaches [6, 22] have been developed for BMAPs and can be extended to MMAPs. Likelihood based fitting of the parameters is faced with several problems. First, as pointed out in [34], the representation of a MAP and also of an MMAP (see Theorem 2) is redundant such that different representations of the same process exist making optimization hard. Additionally, empirical investigations indicate that especially for MAPs or MMAPs with more than 2 or 3 states a large number of local maxima of the likelihood function seem to exist such that an EM algorithm, like other local optimization methods, often traps into the next local optimum. However, the most important problem seems to be that the computational effort of EM algorithms is very high since they often show a slow convergence and the time per iteration depends on the length of the trace. This implies that EM algorithms cannot be applied to real traffic traces resulting from computer systems or networks with often more than a million entries. On the other hand, it seems to be necessary to consider such long traces to capture the behavior adequately, in particular, if autocorrelation can be observed [8].

An alternative to the maximization of the likelihood is the approximation of some derived quantities like empirical moments, joint moments or lag k autocorrelations. This approach scales well with the length of a trace since only the estimation of the empirical moments and joint moments requires the detailed trace data. Methods to fit the parameters of a MAP according to moments, joint moments or lag k autocorrelations are given in [8, 20, 15]. For lengthy measurement traces, these methods are much more efficient than EM algorithms and usually need at most a few seconds to fit the parameters of a MAP of moderate size whereas EM algorithms require several hours or days or fail completely.

Unfortunately, fitting MMAPs according to these derived quantities like moments and joint moments is a complex task and we are aware of only one approach [16] that extends the algorithms presented in [34] to MMAPs. We very briefly describe the approach here.

First, notice that the representation of an MMAP is not unique. Consider an MMAP of dimension n with matrices $(\mathbf{D}_0, \mathbf{D}_1^a)$ ($a \in \mathcal{A}$) and let $\mathbf{B}$ be a non-singular $n \times n$ matrix with $\mathbf{B}\mathbf{e}^T = \mathbf{e}^T$. Let $\mathbf{H}_0 = \mathbf{B}^{-1}\mathbf{D}_0\mathbf{B}$ and $\mathbf{H}_1^a = \mathbf{B}^{-1}\mathbf{D}_1^a\mathbf{B}$ ($a \in \mathcal{A}$). If the off-diagonal elements of $\mathbf{H}_0$ are non-negative and $\mathbf{H}_1^a \geq \mathbf{0}$, then these matrices are a different MMAP representation of the same process. This can be easily proven following [34]. If the matrices do not fulfill the non-negative constraints, then the resulting representation is no longer an MMAP but still represents the same stochastic process.

The fitting procedure of [16, 34] first generates a matrix vector pair $(\nu, \mathbf{H}_0)$ of dimension n from $2n - 1$ moments using the algorithm of [36]. From this vector matrix pair and the joint moments per class, matrices $\mathbf{H}_1^a$ ($a \in \mathcal{A}$) can be generated (cf. [16, Theorem 1]). The resulting matrices $(\mathbf{H}_0, \mathbf{H}_1^a)$ ($a \in \mathcal{A}$) usually do not define an MMAP since they contain negative elements in the matrices $\mathbf{H}_1^a$ and outside the diagonal of $\mathbf{H}_0$. It is not even clear whether the resulting matrices specify a joint density since moment fitting is done algebraically. We denote the resulting process as an MRAP (*Marked Rational Arrival Process*) [2] which is only correct, if the matrices specify a joint density such that a stochastic process results from the description.

Even if the matrices describe a stochastic process, it is not clear whether they are similar to matrices describing an MMAP of the same size or at least of a finite size. If the matrices are similar to an MMAP of the same size, a non-singular matrix $\mathbf{B}$ exists such that $\mathbf{D}_0 = \mathbf{B}^{-1}\mathbf{H}_0\mathbf{B}$ and $\mathbf{D}_1^a = \mathbf{B}^{-1}\mathbf{H}_1^a\mathbf{B}$ are the matrices of the MMAP. Finding an appropriate matrix $\mathbf{B}$ is a non-linear problem for which only heuristic approaches are known. In [34] $\mathbf{B}$ is generated as a product of elementary transformation matrices based on a non-linear optimization approach using exponential penalty functions. For the examples in Section 5, we developed a modified version of [34] with a different objective function that behaved much better than the original. We describe this variant in the following.

Let $(\mathbf{H}_0, \mathbf{H}_1^a)$ $(a \in \mathcal{A})$ be the current matrix representation and let $\mathbf{B}_{i,j,b}$ $(0 \leq b \leq 1)$ a $n \times n$ matrix with b in position i, j, $1 - b$ in position i, i, 1 in position k, k $(k \neq i)$ and 0 elsewhere (see [34, eq. 25]). Define for some $n \times n$ matrix $\mathbf{H}$,

$$f(\mathbf{H}, i, j, b) = (\mathbf{B}_{i,j,b})^{-1} \mathbf{H} \mathbf{B}_{i,j,b}$$

that is used for the transformation of a current representation $(\mathbf{H}_0, \mathbf{H}_1^a)$ $(a \in \mathcal{A})$ into a new one. We further define several functions that contribute to the objective function, namely

$$g_0(\mathbf{H}) = \sum_{i \neq j} \left((-\mathbf{H}(i,j))^+ \right)^2, \; g_1(\mathbf{H}) = \sum_{i,j} \left((-\mathbf{H}(i,j))^+ \right)^2$$

and

$$h_0(\mathbf{H}) = \sum_{i \neq j} \left((\mathbf{H}(i,j))^+ \right)^2, \; h_1(\mathbf{H}) = \sum_{i,j} \left((\mathbf{H}(i,j))^+ \right)^2$$

where $(a)^+ = \max(0, a)$. Then the following iterative algorithm is applied

1. input$(\mathbf{H}_0, \mathbf{H}_1^a)$
2. $r = 1.0$;
3. repeat
4. $\quad (i, j, b) = \arg\min_{k,l \in \{1,\ldots,n\}, c \in [0,1]} \Big(r \cdot \Big(g_0(f(\mathbf{H}_0, k, l, c)) + \sum_{a \in \mathcal{A}} g_1(f(\mathbf{H}_1^a, k, l, c)) \Big) -$
$$\Big(h_0(f(\mathbf{H}_0, k, l, c)) + \sum_{a \in \mathcal{A}} h_1(f(\mathbf{H}_1^a, k, l, c)) \Big) \Big);$$
5. $\quad r = 1.05 * r$;
6. $\quad$ if $(b > 0)$
7. $\qquad \mathbf{H}_0 = f(\mathbf{H}_0, k, l, b)$ and $\mathbf{H}_1^a = f(\mathbf{H}_0, k, l, b)$ for all $a \in \mathcal{A}$;
8. until $(|b| \leq \epsilon)$

For the optimization step the Nelder Mead algorithm or any other optimization algorithm for non-linear problems can be applied. We will compare this variant of the approach in [34] with the new approach developed in the next section with the help of several examples in Section 5.

4 A New Fitting Approach for MMAPs

According to Theorems 1 and 2, an MMAP is completely characterized by an appropriate choice of moments and joint moments. However, given cn^2 moments and joint moments, it is unclear whether they describe a valid MMAP and, even if they do, no method is yet known to compute the MMAP in all cases. When applied to empirical moments and joint moments of real traces, cn^2 moments and joint moments from the trace usually do not define an MMAP of size n. Thus, we choose a different approach by approximating rather than matching the given set of moments and joint moments. This implies that we can always find an MMAP for a given set of moments and joint moments and no relation between the number of moments and the size of the MMAP is required, although a larger size of the MMAP will usually result in a better approximation. Since empirical (joint) moments are estimates for the (joint) moments of some unknown process, an approximation of reasonable quality will usually suffice for successful stochastic modeling.

The approach we present here is a multi-step approach that fits moments or joint moments in every step. The fitting results of one step put some constraints on the parameters fitted in the next step. At every step, the objective is to minimize the weighted squared difference between empirical moments/joint moments and the corresponding quantities of the resulting MMAP. The approach follows the basic ideas that have been presented in [8] to fit the parameters of a simple MAP.

We assume that empirical moments ν_k ($k \in \mathcal{M}$), ν_k^a ($k \in \mathcal{M}^a$, $a \in \mathcal{A}$) and empirical joint moments $\nu_{k,l}$ ($(k,l) \in \mathcal{J}$), $\nu_{k,l}^a$ ($(k,l) \in \mathcal{J}^a, a \in \mathcal{A}$) are available and an MMAP of size n should be fitted according to these (joint) moments. We do not assume any relation between the number of (joint) moments and n since our goal is the minimization of the weighted squared difference between empirical (joint) moments and (joint) moments of the MMAP which only in some lucky cases results in an exact fit.

In the following subsections we present four different steps for parameter estimation performed consecutively such that the preceding step puts constraints on the optimization problems to be solved in the following steps. The steps are combined to a complete algorithm in Section 4.5.

4.1 Fitting General Moments

In a first step, moments ν_k ($k \in \mathcal{M}$) are fitted. This means that the following function is minimized. For given $\nu_k, \beta_k, k \in \mathcal{M}$, we optimize

$$\min_{(\pi,\mathbf{D}_0)\in PHDist} \left(\sum_{k\in\mathcal{M}} \left(\beta_k \left(\frac{\mu_k}{\nu_k} - 1 \right) \right)^2 \right) \tag{9}$$

In (9), $PHDist$ is the class of PH distributions characterized by matrix $\mathbf{D}_0$ and embedded initial distribution π. The moments of $(\pi, \mathbf{D}_0)$ are μ_k. β_k are non-negative weights which may be used to adjust the importance of individual moments. For example, a selection of $\beta_k > \beta_{k+1}$ emphasizes that a good fit of lower order moments is considered more valuable than that of higher order moments.

Different methods for moment fitting exist, in particular for the class of acyclic PH (APH) distributions where matrix $\mathbf{D}_0$ is acyclic and which are a subclass of the class of PH distributions. The first step in the algorithm from [34], presented in the previous section, generates a PH distribution from a set of moments. An alternative approach to find a PH distribution with a specific structure from a set of moments by expanding the state space can be found in [26] while [28] introduces a method to match three moments with a PH distribution of a specific type. In [19] a method is presented to match $2n-1$ moments with an APH distribution of size n. However, this approach only works if the moments can all be exactly matched by the APH distribution which is not necessarily the case for empirical moments resulting from field data. All the mentioned approaches try to match the moments exactly and do not provide a least squares fitting as defined in (9).

Here we briefly recapture the approach from [8] to solve (9). The approach tries to find an APH distribution that minimizes the least squares problem. An APH distribution has a canonical representation characterized by the vector π and a matrix

$$\mathbf{D}_0 = \begin{pmatrix} -\lambda_1 & \lambda_1 & & & \\ & -\lambda_2 & \lambda_2 & & \\ & & \ddots & \ddots & \\ & & & -\lambda_{n-1} & \lambda_{n-1} \\ & & & & -\lambda_n \end{pmatrix} \tag{10}$$

where $\lambda_{i+1} \geq \lambda_i$ for $1 \leq i < n$. In [10], Cumani defined the canonical representation and described how to transform every APH distribution into its canonical form. Observe that the canonical representation is non-redundant and contains $2n-1$ free parameters. The remaining parameter is defined by the condition $\pi\mathbf{e}^T = 1$.

For an APH distribution in canonical form, we have

$$\mathbf{M} = (-\mathbf{D}_0)^{-1} = \begin{pmatrix} (\lambda_1)^{-1} & (\lambda_2)^{-1} & \dots & \dots & (\lambda_n)^{-1} \\ & (\lambda_2)^{-1} & (\lambda_3)^{-1} & \dots & (\lambda_n)^{-1} \\ & & \ddots & \ddots & \\ & & & (\lambda_{n-1})^{-1} & (\lambda_n)^{-1} \\ & & & & (\lambda_n)^{-1} \end{pmatrix}. \tag{11}$$

Now define $\mathbf{m}^k = \mathbf{M}^k\mathbf{e}^T$ such that

$$\mathbf{m}^k(i) = \begin{cases} \sum\limits_{j=i}^{n} (\lambda_j)^{-1}\mathbf{m}^{k-1}(j) & \text{for } k \geq 1 \text{ ,} \\ 1 & \text{otherwise,} \end{cases} \tag{12}$$

and $\mu_k = k! \sum_{i=1}^{n} \pi(i)\mathbf{m}^k(i)$. If we substitute this representation of μ_k into (9) we obtain a non-linear optimization problem which can be solved with standard methods like the Nelder Mead method. Alternatively, one can use an approach that alternates between the computation of π and $\Lambda = (\lambda_1, \dots, \lambda_n)$. Assuming Λ is known, then we have to find the minimal solution of

$$\min_{\pi \geq \mathbf{0}, \pi\mathbf{e}^T=1} (\|\pi\mathbf{A} - \mathbf{b}\|_2) \tag{13}$$

where the $\mathbf{A}$ is an $n \times |\mathcal{M}|$ matrix where the lth column equals $\mathbf{m}^k$ if k is the lth element in $\mathcal{M}$ and $\mathbf{b}(l) = \nu_k$. Observe that (13) is a non-negative least squares problem with one linear constraint that can be solved with the approach proposed in [24, 33]. For a fixed vector π the computation of Λ is a non-linear optimization problem.

State i of an APH distribution is an exit state if $\sum_{j=1}^{n} \mathbf{D}_0(i,j) < 0$, it is an entry state if $\pi(i) > 0$. The number of free parameters for fitting joint moments depends on the product of the number of exit and entry states. Thus we need to transform the canonical representation of an APH distribution with only one exit state into an equivalent one with several exit states. Suitable equivalence transformations are given in [8, 20, 7].

Alternatively, if the trace is not too long or if it can be sufficiently aggregated [29] one may use some maximum likelihood-based method for generating a PH (APH) distribution. Necessary algorithms for distribution fitting are described in [21, 35, 18]. Observe that these algorithm are specifically tailored to the efficient fitting of distributions, they cannot be applied for MAP or MMAP fitting.

In any case, the result of the first step is a PH distribution $(\pi, \mathbf{D}_0)$.

4.2 Fitting General Joint Moments

The solution of (9) provides us with $(\pi, \mathbf{D}_0)$. In the second step, matrix $\mathbf{D}_1$ is computed such that the empirical joint moments $\nu_{k,l}$ ($(k,l) \in \mathcal{J}$) are approximated by minimizing the following function. For given $\nu_{k,l}, \beta_{k,l}, (k,l) \in \mathcal{J}, \pi$, and $\mathbf{D}_0$, we optimize

$$\min_{\mathbf{D}_1:\ \mathbf{D}_1 \geq \mathbf{0}, \mathbf{D}_1\mathbf{e}^T = -\mathbf{D}_0\mathbf{e}^T, \pi\mathbf{M}\mathbf{D}_1 = \pi} \left(\sum_{(k,l)\in\mathcal{M}} \left(\beta_{k,l} \left(\frac{\mu_{k,l}}{\nu_{k,l}} - 1 \right) \right)^2 \right) \tag{14}$$

Since $\mathbf{D}_0$ and $\mathbf{M}$ are fixed, only $\mathbf{D}_1$ can be modified. To yield a valid MAP, the constraints on the row sums of $\mathbf{D}_1$ have to be observed and $\mathbf{D}_1$ has to be chosen in a way that π remains the left eigenvector of $\mathbf{M}\mathbf{D}_1$ according to eigenvalue 1. Altogether this defines $2n-1$ linear constraints. Since the joint moments of the MAP can be represented as

$$\mu_{k,l} = k!l! \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{v}^k(i)\mathbf{D}_1(i,j)\mathbf{w}^l(j) \text{ where } \mathbf{v}^k = \pi\mathbf{M}^{k+1} \text{ and } \mathbf{w}^l = \mathbf{M}^l\mathbf{e}^T \tag{15}$$

a non-negative least squares problem with n^2 variables and $2n-1$ linear constraints has to be solved.

The outcome of this step is a MAP $(\mathbf{D}_0, \mathbf{D}_1)$, which can be refined to an MMAP as shown in the following two steps.

4.3 Fitting Class-Specific Moments

We assume for this step a PH distribution $(\pi, \mathbf{D}_0)$ is available resulting from the first step and fitted according to moments ν_k ($k \in \mathcal{M}$). In this step, vectors π^a ($a \in \mathcal{A}$) are computed such that the moments ν_k^a ($k \in \mathcal{M}^a$, $a \in \mathcal{A}$) are approximated. The following minimization problem has to be solved. For given $\nu_k^a, \beta_k^a, k \in \mathcal{M}^a, \pi$, and $\mathbf{D}_0$, we optimize

$$\min_{\pi^a(a\in\mathcal{A}): \sum_{a\in\mathcal{A}} \pi^a = \pi, \pi^a \geq \mathbf{0}} \left(\sum_{k\in\mathcal{M}^a} \left(\beta_k^a \left(\frac{\mu_k^a}{\nu_k^a} - 1 \right) \right)^2 \right) \tag{16}$$

where β_k^a are non-negative weights.

$$\mu_k^a = \frac{k!}{p(a)} \pi^a \mathbf{M}^k \mathbf{e}^T \tag{17}$$

implies that the problem of finding appropriate vectors π^a in (16) is a non-negative least squares problem with cn variables and n constraints since $\pi = \sum_{a\in\mathcal{A}} \pi^a$. Values for probabilities $p(a)$ are estimated from measurement data as

$$p(a) = \frac{1}{m-1} \sum_{i=1}^{m-1} \delta(a_i = a) \text{ for all } a \in \mathcal{A}\ . \tag{18}$$

One can also perform this step if $(\pi, \mathbf{D}_0)$ is not available; then the only constraint is $\left(\sum_{a\in\mathcal{A}} \pi^a\right) \mathbf{e}^T = 1$. Of course, in this case, matrix $\mathbf{D}_0$ has to be computed as well. This can be done by an iterative approach that performs non-negative least squares to find π^a for fixed $\mathbf{M}$ and a non-linear optimization approach to find $\mathbf{D}_0$ in canonical form for fixed π^a.

4.4 Fitting Class-Specific Joint Moments

We assume that $(\mathbf{D}_0, \mathbf{D}_1)$ and π^a are available and $\mathbf{D}_1^a$ ($a \in \mathcal{A}$) should be computed with respect to the following minimization problem. For given $\nu_{k,l}^a, \beta_{k,l}^a, (k,l) \in \mathcal{J}^a, \pi^a, \mathbf{D}_0$, and $\mathbf{D}_1$, we optimize

$$\min_{\mathbf{D}_1^a:\, \mathbf{D}_1^a \geq \mathbf{0},\, \sum_{a\in\mathcal{A}} \mathbf{D}_1^a = \mathbf{D}_1, \pi\mathbf{M}\mathbf{D}_1^a = \pi^a} \left(\sum_{(k,l)\in\mathcal{J}^a} \left(\beta_{k,l}^a \left(\frac{\mu_{k,l}^a}{\nu_{k,l}^a} - 1 \right) \right)^2 \right) \tag{19}$$

where $\beta_{k,l}^a$ are non-negative weights. $\mu_{k,l}^a$ can be represented as

$$\mu_{k,l}^a = \frac{k!l!}{p(a)} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{v}^k(i) \mathbf{D}_1^a(i,j) \mathbf{w}^l(j) \tag{20}$$

where $\mathbf{v}^k$ and $\mathbf{w}^l$ are as in (15). This defines a non-negative least squares problem with cn^2 variables and $n^2 + (c-1)n$ linear constraints. n^2 constraints result from $\sum_{a\in\mathcal{A}} \mathbf{D}_1^a = \mathbf{D}_1$ and $(c-1)n$ result from $\pi\mathbf{M}\mathbf{D}_1^a = \pi^a$. If for all except one $a \in \mathcal{A}$ the condition holds, then it also holds for the last vector since the choice of $\sum_{a\in\mathcal{A}} \mathbf{D}_1^a = \mathbf{D}_1$ assures $\pi\mathbf{M}\mathbf{D}_1 = \pi$.

If $\mathbf{D}_1$ is not available (i.e., step 2 has not been performed), then only the constraints $\sum_{a\in\mathcal{A}} \mathbf{D}_1^a \mathbf{e}^T = -\mathbf{D}_0\mathbf{e}^T$ apply and a non-negative least squares problem with cn constraints has to be solved.

4.5 The Fitting Algorithm

The complete fitting algorithm consists of the steps described in the previous subsections and is summarized below.

1. Analyze the available trace and determine empirical moments ν_k ($k \in \mathcal{M}$) and ν_k^a ($k \in \mathcal{M}^a$, $a \in \mathcal{A}$) and empirical joint moments $\nu_{k,l}$ ($(k,l) \in \mathcal{J}$) and $\nu_{k,l}^a$ ($(k,l) \in \mathcal{J}^a, a \in \mathcal{A}$).
2. Use the moments ν_k to fit an acyclic or cyclic phase type distribution $(\pi, \mathbf{D}_0)$.
3. Use the joint moments $\nu_{k,l}$ to fit the parameters of a MAP to determine $\mathbf{D}_1$ by solving the non-negative least squares problem defined in (14).
4. Use the class-specific moments ν_k^a to determine the embedded class-specific vectors π^a by solving the non-negative least squares problem defined in (16).
5. Use the class-specific joint moments $\nu_{k,l}^a$ to determine matrices $\mathbf{D}_1^a$ by solving the non-negative least squares problem defined in (19).

The third step of the algorithm is elective. If the third step is omitted, then the number of constraints in (19) is reduced since $\mathbf{D}_1$ is not available. For the second step of the algorithm, we can choose any method for fitting PH distributions. We are not forced to use moments fitting, one can as well use an EM algorithm to maximize the likelihood function with respect to trace data as in [4, 5]. However, a moment fitting algorithm scales better for large data sets than an EM algorithm.

To apply the complete fitting algorithm, we need to choose values for weights $\beta_{\cdot}$ in the individual optimization problems. From our experience, we can recommend the following. If traces from MMAPs are used where the (joint) moments can be fitted exactly, setting all $\beta_{\cdot}$ to 1 is usually the best choice. However, for empirical (joint) moments that cannot be matched exactly by an MMAP of size n, it is more important to approximate lower order (joint) moments well than higher order ones. Our experiments indicate that 0.1^{i-1} and 0.1^{i+j-2} are good choices for β_i and $\beta_{i,j}$, respectively.

Apart from determining matrix $\mathbf{D}_0$ in the second step, all other steps require only the solution of non-negative least squares problems. The constraints are integrated in the least squares problems using weights [24]. The dimension of the least squares problems equals $n \times |\mathcal{M}| + 1$ for (9), $n^2 \times |\mathcal{J}| + 2n$ for (16), $cn \times (\sum_{a\in\mathcal{A}} |\mathcal{M}^a|) + n$ for (14) and $cn^2 \times (\sum_{a\in\mathcal{A}} |\mathcal{J}^a|) + cn$ or $cn^2 \times (\sum_{a\in\mathcal{A}} |\mathcal{J}^a|) + cn + n^2$ for (19) depending on whether $\mathbf{D}_1$ has been precomputed or not. Solving these problems is computationally inexpensive as long as n or c are not too large (i.e., cn^2 has to be in the range of a few hundred at most). Only the determination of $\mathbf{D}_0$ requires a non-linear optimization step. We used the Nelder Mead algorithm from Gnu-Lib for the non-linear optimization step and the available implementation of the non-negative least square solver [33].

5 Experiments and Results

In this section, we evaluate the performance of the proposed approach with the help of several examples. First, we fit MMAPs with respect to moments and joint moments of some known MMAPs in order to evaluate if our approach yields MMAPs that are in close correspondence to the original MMAPs. Afterwards, we apply our approach to a real network trace.

5.1 Fitting an MMAP to data generated from a known MMAP

We begin with an experiment that uses known MMAPs as input in order to see if the method yields consistent results and a fitting of good quality. Since iterative numerical methods are applied in several steps of the algorithm, e.g., to solve the least squares problems, we can naturally only reach accuracy up to some $\epsilon > 0$. The most crucial point is the non-linear optimization step required for the generation of $\mathbf{D}_0$. In the following, we compute the moments with 8 significant digits and show the matrices of the resulting MMAPs with 3 significant digits. The weights β_k, $\beta_{k,l}$ are set to 1 in all cases.

We begin with a simple MMAP with 2 states and 2 classes.

$$\mathbf{D}_0 = \begin{pmatrix} -2.00 & 0.00 \\ 0.00 & -1.00 \end{pmatrix},\ \mathbf{D}_1^a = \begin{pmatrix} 0.00 & 2.00 \\ 0.00 & 0.00 \end{pmatrix},\ \mathbf{D}_1^b = \begin{pmatrix} 0.00 & 0.00 \\ 1.00 & 0.00 \end{pmatrix}.$$

The MMAP generates an alternating sequence $abab\ldots$, the time after an a has been observed is exponentially distributed with rate 1, the time after a b has been observed is exponentially distributed with rate 2. We compute the moments ν_k, the conditional moments ν_k^x, and the conditional joint moments $\nu_{k,l}^x$ ($k, l \in \{1, 2, 3\}$, $x \in \{a, b\}$) for fitting the parameters of the MMAP. We do not compute the unconditional joint moments (i.e., step 3 of the algorithm is not performed). The algorithm computes the original MMAP from these moments and joint moments in about one second.

Using the approach from [16] an MRAP with negative elements in the matrices is generated. However, our procedure finds an appropriate transformation matrix $\mathbf{B}$ to transform the matrices of the MRAP into $\mathbf{D}_0$, $\mathbf{D}_1^a$, $\mathbf{D}_1^b$.

Equation	$\|\ldots\|_2$	$\min\left(\nu_{\cdot}^{\cdot}/\mu_{\cdot}^{\cdot}\right)$	$\max\left(\nu_{\cdot}^{\cdot}/\mu_{\cdot}^{\cdot}\right)$
Eq. (9)	5.745e-13	1.000	1.000
Eq. (16)	1.429e-10	1.000	1.000
Eq. (19)	1.007e-02	0.990	1.002

Table 1. Difference between original and fitted MMAP.

The next example models IP traffic and is taken from [22]. It is an MMAP of size 3 with 4 classes.

$$\mathbf{D}_0 = \begin{pmatrix} -228.0 & 1.00 & 1.00 \\ 3.00 & -248.00 & 2.00 \\ 1.00 & 2.00 & -288.00 \end{pmatrix}, \ \mathbf{D}_1^a = \begin{pmatrix} 1.00 & 2.00 & 3.00 \\ 20.0 & 10.0 & 30.0 \\ 40.0 & 50.0 & 100.0 \end{pmatrix},$$
$$\mathbf{D}_1^b = \begin{pmatrix} 2.00 & 5.00 & 7.00 \\ 10.0 & 50.0 & 1.0 \\ 1.00 & 20.0 & 7.00 \end{pmatrix}, \ \mathbf{D}_1^c = \begin{pmatrix} 100.0 & 5.0 & 1.0 \\ 0.0 & 20.0 & 1.0 \\ 0.0 & 0.0 & 10.0 \end{pmatrix}, \ \mathbf{D}_1^d = \begin{pmatrix} 100.0 & 0.0 & 0.0 \\ 0.0 & 100.0 & 1.0 \\ 2.0 & 5.0 & 50.0 \end{pmatrix}.$$

Since $\mathbf{D}_0$ is not acyclic and we fit the parameters of an acyclic PH distribution in the first step, our approach cannot generate the same MMAP. Thus, the question is how good the approximation with an acyclic matrix $\mathbf{D}_0$ will be.

From the given MMAP, we compute the values for parameters ν_k, ν_k^x and $\nu_{k,l}^x$ for $k, l \in \{1, \ldots, 5\}$ and $x \in \{a, b, c, d\}$. For these values, our algorithm generates the following MMAP in less than a second.

$$\mathbf{D}_0 = \begin{pmatrix} -227.3 & 0.00 & 0.00 \\ 0.00 & -243.2 & 0.00 \\ 0.00 & 0.00 & -285.8 \end{pmatrix}, \ \mathbf{D}_1^a = \begin{pmatrix} 0.000 & 8.224 & 0.491 \\ 15.82 & 14.50 & 34.46 \\ 75.00 & 1.673 & 114.5 \end{pmatrix},$$
$$\mathbf{D}_1^b = \begin{pmatrix} 0.000 & 10.28 & 1.673 \\ 10.67 & 45.73 & 9.767 \\ 16.81 & 1.267 & 11.39 \end{pmatrix}, \ \mathbf{D}_1^c = \begin{pmatrix} 98.24 & 7.839 & 0.000 \\ 0.000 & 13.89 & 0.000 \\ 0.000 & 0.000 & 6.252 \end{pmatrix},$$
$$\mathbf{D}_1^d = \begin{pmatrix} 99.81 & 0.000 & 0.000 \\ 0.000 & 98.36 & 0.000 \\ 0.000 & 12.14 & 46.78 \end{pmatrix}.$$

This MMAP differs from the original one, but the moments and conditional moments are in very good agreement as shown in Table 1 where we print the residual norm of the least square fitting and the difference of the moments by printing the minimum and maximum values of $\nu_{\cdot}^{\cdot}/\mu_{\cdot}^{\cdot}$ where $\nu_{\cdot}^{\cdot}$ are the (joint) moments of the original MMAP and $\mu_{\cdot}^{\cdot}$ are the corresponding values for the fitted MMAP. For the joint moments small differences between the original and the fitted MMAP appear but the difference is in all cases less than 1% which is an excellent result.

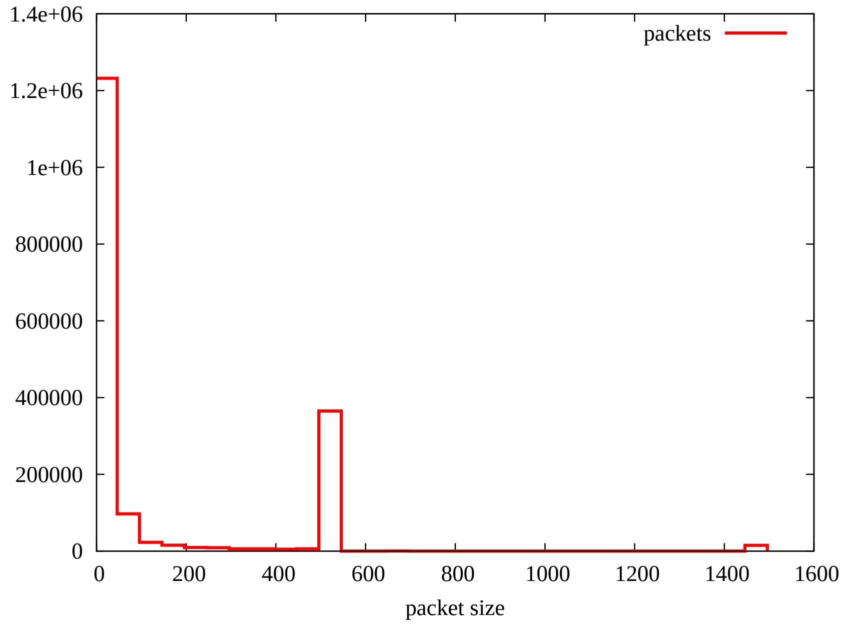

Figure 1. Distribution of the packet sizes for the trace *LBL-TCP-3*

If we apply the approach of [16], an MRAP is generated first which is then transformed into the following MMAP.

$$\mathbf{D}_0 = \begin{pmatrix} -248.1 & 1.76 & 4.06 \\ 5.93 & -287.8 & 2.93 \\ 0.96 & 1.14 & -228.1 \end{pmatrix}, \mathbf{D}_1^a = \begin{pmatrix} 7.32 & 33.07 & 17.23 \\ 37.83 & 102.92 & 30.45 \\ 1.82 & 3.44 & 0.76 \end{pmatrix},$$

$$\mathbf{D}_1^b = \begin{pmatrix} 50.45 & 1.13 & 5.96 \\ 23.37 & 7.15 & 0.05 \\ 4.63 & 7.99 & 1.40 \end{pmatrix}, \mathbf{D}_1^c = \begin{pmatrix} 20.26 & 1.03 & 5.30 \\ 1.19 & 10.14 & 3.19 \\ 5.26 & 1.09 & 99.61 \end{pmatrix},$$

$$\mathbf{D}_1^d = \begin{pmatrix} 100.00 & 0.52 & 0.00 \\ 9.75 & 50.00 & 2.91 \\ 0.00 & 0.00 & 100.00 \end{pmatrix}.$$

This MMAP differs from the original MMAP showing that the representation is not unique.

The two examples give evidence that our fitting approach is able to compute an MMAP that is either identical to or a good approximation of an MMAP used to generate the input data. Similar observations have been made for other example MMAPs, such that we believe this is a common property of the presented fitting algorithm. So far, the algorithm behaves for moments and joint moments resulting from MMAPs of the same size similar to the approach of [16]. The difference is that the approach presented in this paper introduces some approximations but always generates a MMAP whereas the approach of [16] matches moments exactly but the matrix for the similar transformation into an MMAP has to be found which can be

Moments	1	2	3	4	5
Trace	4.022e-3	4.760e-5	1.096e-6	3.946e-8	1.976e-9
MAP 2	4.022e-3	4.760e-5	1.097e-6	3.752e-8	1.659e-9
MAP 4	4.022e-3	4.760e-5	1.096e-6	3.946e-8	1.974e-9
MAP 8	4.022e-3	4.760e-5	1.096e-6	3.946e-8	1.974e-9
RAP 2	4.022e-3	4.760e-5	1.096e-6	3.739e-8	1.648e-9
RAP 3	4.022e-3	4.760e-5	1.096e-6	3.946e-8	1.976e-9

Table 2. Moments of the trace and the fitted MAPs

Joint Moments	1,1	2,2	3,3	4,4	5,5
Trace	2.105e-05	4.559e-09	3.772e-12	7.364e-15	2.375e-17
MAP 2	2.103e-05	4.805e-09	3.265e-12	4.139e-15	8.278e-18
MAP 4	2.105e-05	4.561e-09	3.683e-12	7.043e-15	2.462e-17
MAP 8	2.105e-05	4.560e-09	3.704e-12	7.103e-15	2.493e-17
RAP 2	2.105e-05	4.803e-09	3.249e-12	4.097e-15	8.149e-18
RAP 3	2.105e-05	4.559e-09	3.587e-12	6.639e-15	2.189e-17

Table 3. Joint moments of the trace and the fitted MAPs

cumbersome. Since real-world data sources are not necessarily MMAPs, further empirical evaluations remain necessary.

5.2 Fitting an MMAP to a Network Trace

We present results for MMAP fittings to the trace *LBL-TCP-3* [30] a common benchmark trace taken from the Internet Traffic Archive [1]. It was recorded at the Lawrence Berkeley Laboratory in January 1994 and consists of two hours of TCP traffic resulting in almost 1.79 million observed packets. The trace contains the interarrival times and packet sizes (and some other information like TCP ports that we did not use for our experiments). We partition packets into four classes according to their size because different packet sizes will imply different processing times.

The classes become almost immediately apparent from showing the number of packets for different packet sizes (see Figure 1). The plot contains three peaks. Hence, we choose class b to contain the packet sizes from 1 to 200, class c for the packet sizes between 201 and 600 and class d for packets larger than 600, including the smaller peak at 1450. Finally, we use class a for TCP-Acknowledgement messages that are denoted by a packet size of 0 in the trace *LBL-TCP-3* and are part of the first peak in Figure 1.

We consider the class-independent interarrival times of the trace first and fit the parameters of MAPs with 2, 4 and 8 states to approximate moments ν_i and joint moments $\nu_{i,j}$ with $i, j = 1, \ldots, 5$. We use weights $\beta_i = 10^{-(i-1)}$ for the fitting of the ith moment and $\beta_{i,j} = 10^{-(i+j-2)}$ for the joint moment with index (i, j). Table 2 shows the moments for the trace and the fitted MAPs with 4 digits of accuracy. The MAP with 2 states matches the first three moments and the MAPs with 4 and 8 states match the first four moments and have a small difference in the last digit of the fifth moment. Our results from [8] indicate that it does not make sense to fit higher order moments above order 5 since the estimates of these moments become unreliable even when taken from a trace with nearly two million entries. Table 3 contains the joint moments $\nu_{i,i}$ of the trace and the joint moments $\mu_{i,i}$ of the fitted MAPs. Results for $i \neq j$ are similar. It can be seen that it is not possible to match the higher order joint moments exactly but we obtain a reasonable approximation that becomes better if the number of states of the MAP grows although the difference between 4 and 8 states is marginal.

Now we consider the class-specific moments fitted with an MMAP. Table 4 contains the results. First, one can notice that the moments of the different classes differ such that a fitting with MMAPs has potential benefits. For all three MMAPs the fitting quality is acceptable. The MMAP with 2 states has some problems in fitting the moments of class 4. This becomes better if the number of states is increased to 4 and the MMAP with 8 states provides an almost error-free fitting of the first 5 moments.

The situation is similar for the class-specific joint moments shown in Table 5. The approximation in all cases is reasonable and becomes better with an increasing number of states although this does not hold for all moments and classes. The whole fitting, including the estimation of the moments from the trace, requires between 1 and 2 minutes, where most of the time is used to compute the initial APH distribution.

We also applied the approach of [16] to the moments and joint moments of the *LBL-TCP-3* trace and generated processes with two and three states denoted as (M)RAP 2 and 3, respectively. The (joint) moments of these processes are included in the Tables 2-5. For MRAP 3 it is obvious that the matrices do not describe a valid stochastic process since some of the moments $E((X^{(d)})^k)$ and $E((X_0^{(d)})^k (X_1)^l)$ are negative which is not allowed in a stochastic process with only non-negative values. The situation is not clear for MRAP 2. From moment fitting with the method of [16] we obtain the following matrices.

$$\mathbf{H}_0 = \begin{pmatrix} -516.58 & 182.17 \\ -248.61 & 0.00 \end{pmatrix},$$

$$\mathbf{H}_1^a = \begin{pmatrix} 113.95 & -9.72 \\ 61.63 & 18.22 \end{pmatrix}, \; \mathbf{H}_1^b = \begin{pmatrix} 168.69 & -29.07 \\ 91.20 & 18.93 \end{pmatrix},$$

$$\mathbf{H}_1^c = \begin{pmatrix} 163.15 & -76.20 \\ 92.54 & -36.18 \end{pmatrix}, \; \mathbf{H}_1^d = \begin{pmatrix} 5.14 & -1.53 \\ 3.24 & -0.97 \end{pmatrix}.$$

Moments	1	2	3	4	5
class a					
Trace	4.455e-3	5.619e-5	1.305e-6	4.602e-8	2.215e-9
MMAP 2	4.452e-3	5.742e-5	1.377e-6	4.771e-8	2.117e-9
MMAP 4	4.455e-3	5.614e-5	1.305e-6	4.583e-8	2.205e-9
MMAP 8	4.455e-3	5.619e-5	1.305e-6	4.601e-8	2.233e-9
MRAP 2	4.455e-3	5.746e-5	1.376e-6	4.756e-8	2.105e-9
MRAP 3	4.455e-3	5.619e-5	1.328e-6	4.829e-8	2.430e-9
class b					
Trace	4.348e-3	5.376e-5	1.276e-6	4.743e-8	2.474e-9
MMAP 2	4.345e-3	5.499e-5	1.308e-6	4.519e-8	2.003e-9
MMAP 4	4.348e-3	5.373e-5	1.276e-6	4.734e-8	2.438e-9
MMAP 8	4.348e-3	5.376e-5	1.276e-6	4.674e-8	2.454e-9
MRAP 2	4.348e-3	5.503e-5	1.307e-6	4.505e-8	1.992e-9
MRAP 3	4.348e-3	5.376e-5	1.252e-6	4.500e-8	2.244e-9
class c					
Trace	2.806e-3	2.458e-5	4.841e-7	1.603e-8	7.394e-10
MMAP 2	2.819e-3	2.007e-5	3.131e-7	8.977e-9	3.743e-10
MMAP 4	2.806e-3	2.457e-5	4.836e-7	1.649e-8	8.188e-10
MMAP 8	2.806e-3	2.458e-5	4.841e-7	1.602e-8	7.440e-10
MRAP 2	2.806e-3	1.985e-5	3.075e-7	8.771e-9	3.644e-10
MRAP 3	2.806e-3	2.458e-5	5.132e-7	1.873e-8	9.721e-10
class d					
Trace	3.212e-3	1.808e-5	2.005e-7	4.124e-9	1.275e-10
MMAP 2	3.121e-3	2.698e-5	5.100e-7	1.614e-8	6.967e-10
MMAP 4	3.203e-3	2.153e-5	2.223e-7	3.090e-9	5.396e-11
MMAP 8	3.212e-3	1.820e-5	1.976e-7	4.053e-8	1.311e-10
MRAP 2	3.212e-3	2.912e-5	5.708e-7	1.833e-8	7.933e-10
MRAP 3	3.212e-3	1.808e-5	-1.562e-7	-2.541e-8	-2.065e-09

Table 4. Class-specific moments of the trace and the fitted MMAPs

Joint Moments	1,1	2,2	3,3	4,4	5,5
class a					
Trace	2.326e-5	5.274e-9	4.484e-12	8.245e-15	2.296e-17
MMAP 2	2.322e-5	5.573e-9	3.826e-12	4.860e-15	9.726e-18
MMAP 4	2.326e-5	5.275e-9	4.506e-12	9.164e-15	3.366e-17
MMAP 8	2.326e-5	5.275e-9	4.481e-12	8.997e-15	3.264e-17
MRAP 2	2.326e-5	5.293e-9	3.537e-12	4.439e-15	8.819e-18
MRAP 3	2.326e-5	5.275e-9	4.434e-12	8.611e-15	2.910e-17
class b					
Trace	2.435e-5	5.503e-9	4.681e-12	9.929e-15	3.568e-17
MMAP 2	2.431e-5	5.914e-9	4.065e-12	5.163e-15	1.033e-17
MMAP 4	2.435e-5	5.507e-9	4.447e-12	8.401e-15	2.908e-17
MMAP 8	2.435e-5	5.506e-9	4.512e-12	8.739e-15	3.096e-17
MRAP 2	2.435e-5	5.761e-9	3.897e-12	4.909e-15	9.762e-18
MRAP 3	2.435e-5	5.503e-9	4.302e-12	7.767e-15	2.505e-17
class c					
Trace	1.200e-5	1.865e-9	1.133e-12	1.401e-15	2.500e-18
MMAP 2	1.200e-5	1.719e-9	1.033e-12	1.273e-15	2.529e-18
MMAP 4	1.200e-5	1.864e-9	1.171e-12	1.774e-15	5.024e-18
MMAP 8	1.200e-5	1.865e-9	1.142e-12	1.593e-15	4.121e-18
MRAP 2	1.200e-5	2.410e-9	1.700e-12	2.185e-15	4.372e-18
MRAP 3	1.200e-5	1.865e-9	1.130e-12	1.917e-15	6.443e-18
class d					
Trace	8.587e-6	4.586e-10	1.092e-13	4.468e-17	2.213e-20
MMAP 2	8.560e-6	4.250e-10	6.639e-14	2.177e-17	1.181e-20
MMAP 4	8.587e-6	3.826e-10	4.919e-14	1.631e-17	1.271e-20
MMAP 8	8.585e-6	4.473e-10	1.183e-13	9.044e-17	1.270e-19
MRAP 2	8.587e-6	4.798e-10	1.167e-13	9.008e-17	1.506e-19
MRAP 3	8.587e-6	4.586e-10	5.046e-14	-2.097e-16	-1.312e-18

Table 5. Class-specific joint moments of the trace and the fitted MMAPs

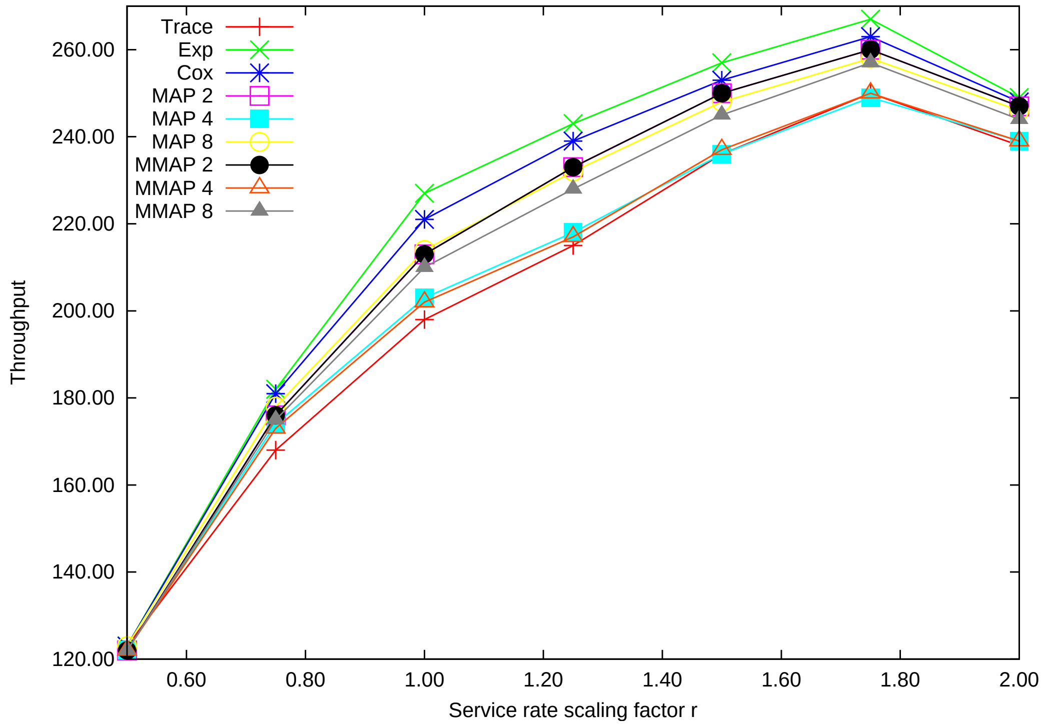

Figure 2. Average throughput of the system under different models of the arrival process

The procedure to transform the matrices into matrices of an MMAP stopped with the following set of matrices.

$$\mathbf{H}_0 = \begin{pmatrix} -111.92 & 0.000708 \\ 0.000446 & -404.66 \end{pmatrix}, \; \mathbf{H}_1^a = \begin{pmatrix} 28.23 & 12.78 \\ 20.24 & 103.94 \end{pmatrix},$$
$$\mathbf{H}_1^b = \begin{pmatrix} 43.12 & 20.03 \\ 19.26 & 144.51 \end{pmatrix}, \; \mathbf{H}_1^c = \begin{pmatrix} 9.906 & -2.293 \\ -5.059 & 117.06 \end{pmatrix}, \; \mathbf{H}_1^d = \begin{pmatrix} 0.013 & 0.127 \\ 0.542 & 4.155 \end{pmatrix}.$$

It can be seen that the matrix $\mathbf{H}_1^c$ still contains two negative values that we were not able to eliminate. Consequently, it remains open whether the matrices describe a valid stochastic process and, even if they do, whether an equivalent MMAP of size 2 exists. However, it can be seen that there is only a marginal difference between the moments and joint moments of (M)MAP 2 and (M)RAP 2 such that the exact moment fitting has no real advantage.

As a final step we consider queueing behavior and analyze a finite capacity queue with capacity 10 and class-specific exponentially distributed service time with rates 300, 250, 200, and 100 for the classes a through d. The rates are multiplied with a common scaling factor $r = 0.5, \ldots, 2.0$ to analyze the system under different load conditions such that a higher value of r yields faster service. Packets are served with class-specific non-preemptive priorities that favor smaller packet sizes. We analyzed the system using a trace driven analysis with the original trace, with Poisson arrivals for all classes *(Exp)*, with class-specific Coxian distributions that match the first 3 moments of the classes *(Cox)*, with the three MAPs *(MAP 2, 4, 8)* and with the MMAPs *(MMAP 2, 4, 8)*.

Results for the average overall throughput and population in the interval $[0, t_{\max}]$, where $t_{\max}$ is the length of the trace, are

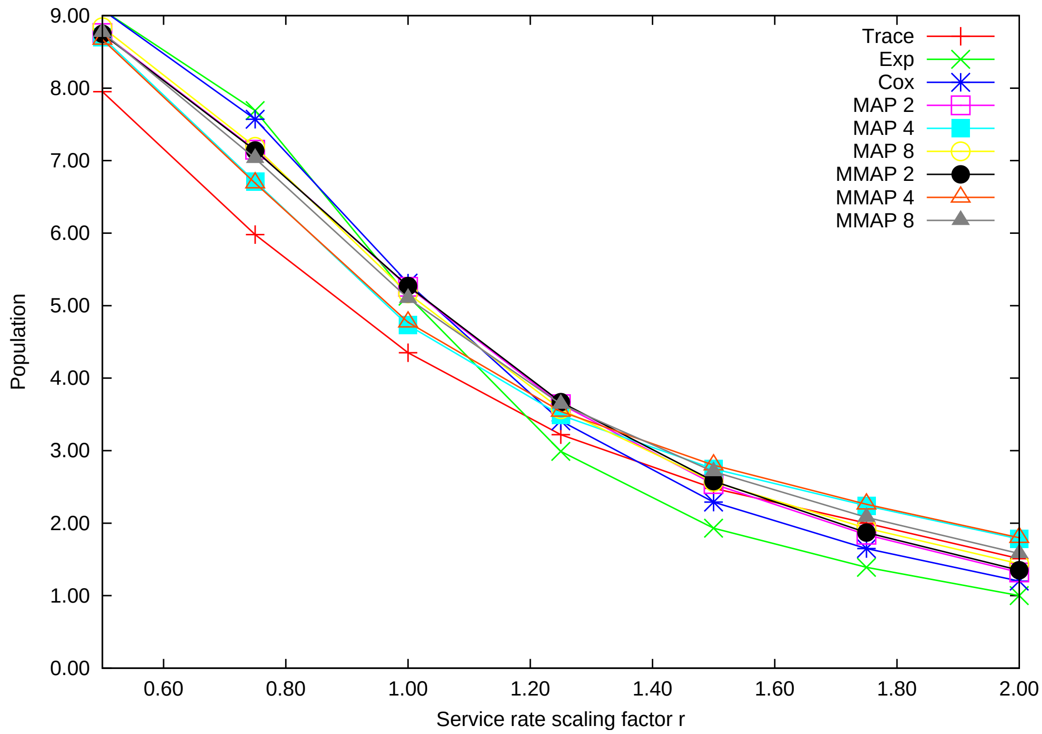

Figure 3. Average population of the system under different models of the arrival process

shown in the Figures 2 and 3. Table 6 includes a summary of the relative error for the different input processes. The average error is presented per class and overall classes for population and throughput. The results give a mixed picture with no clear tendency. The MAPs and MMAPs provide a better approximation than the exponential or Coxian arrivals. The MMAPs are only slightly better than the MAPs and the MAPs with a larger state space are not necessarily better even if the approximation of conditional moments and joint moments is better. The errors for the classes 3 and 4 are large for all input processes which shows that queue conditional moments and joint moments are not the most significant quantities to determine the behavior in these cases.

6 Conclusions

In this paper, we present a parameter fitting method for MMAPs to accurately model arrival streams with multiple classes of arrivals and correlation among those classes. The approach is based entirely on estimated moments and joint moments of an arrival stream, which can be efficiently obtained even for massive amounts of measurement data, meaning that our method scales well with real world applications. For given estimates of moments and joint moments, the overall approach consists of multiple steps with all but one requiring us to solve only least squares minimization problems. The only one that requires non-linear optimization is the fitting of $\mathbf{D}_0$ to the moments of the arrival stream.

We evaluate our method in two ways: we perform a fitting for an MMAP where the given moments are obtained from

Input	Population				
	cl a	cl b	cl c	cl d	overall
Exp	7.78	12.34	31.95	253.4	21.59
Cox	9.00	11.62	22.10	195.1	16.07
MAP 2	7.56	14.17	15.81	135.2	12.10
MAP 4	0.53	13.79	16.33	177.1	12.07
MAP 8	7.58	17.01	12.40	151.5	10.14
MMAP 2	7.56	10.17	15.94	134.2	11.75
MMAP 4	0.80	7.88	11.33	115.7	7.57
MMAP 8	4.78	7.27	14.17	109.1	10.28
	Throughput				
Exp	8.00	5.25	19.93	63.90	8.08
Cox	5.83	4.54	18.79	58.14	7.39
MAP 2	4.39	3.74	16.97	55.63	5.01
MAP 4	1.17	3.90	12.32	50.46	0.95
MAP 8	4.07	3.32	16.77	55.58	4.83
MMAP 2	4.16	3.66	16.34	55.41	4.91
MMAP 4	0.87	2.75	12.23	50.33	0.99
MMAP 8	3.21	1.76	15.34	39.45	3.68

Table 6. Average absolute relative error in percent for the different input processes

a known MMAP so we can evaluate the accuracy of our fitting method. In these cases, it is usually possible to generate an MMAP with a very similar behavior. The second set of experiments fits the moments of a real network trace. The experiences in this case are mixed. It is possible to accurately approximate the moments and joint moments of the trace with an MMAP of low order. However, this does not imply that the performance measures of a queue that is fed with the trace are also accurately approximated when the trace is substituted by the fitted MMAP. It seems that although moments and joint moments completely characterize an MMAP, queueing performance is influenced by several other quantities such as higher order joint moments not considered in our current approach. Nevertheless, this paper presents one of the first fitting methods for MMAPs and can serve as a starting point for the development of methods that consider additional quantities for parameter fitting.

References

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

[2] S. Asmussen and M. Bladt. Point processes with finite-dimensional conditional probabilities. *Stochastic Processes and their Application*, 82:127–142, 1999.

[3] S. Asmussen and G. Koole. Marked point processes as limits of Markovian arrival streams. *Journal of Applied Probability*, 30:365–372, 1993.

[4] S. Asmussen, O. Nerman, and M. Olsson. Fitting phase-type distributions via the EM-algorithm. *Scand. J. Stat.*, 23(4):419–441, 1996.

[5] A. Bobbio and A. Cumani. ML estimation of the parameters of a PH distribution in triangular canonical form. In G. Balbo and G. Serazzi, editors, *Computer Performance Evaluation*, pages 33–46. Elsevier, 1992.

[6] L. Breuer. An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure. *Annals OR*, 112(1-4):123–138, 2002.

[7] P. Buchholz and J. Kriege. Equivalence transformations for acyclic phase type distributions. Technical Report 827, Dep. of Informatics, TU Dortmund, 2009. http://www.cs.uni-dortmund.de/nps/de/Forschung/Publikationen/Graue_Reihe1/Ver__ffentlichungen_2009/827.pdf.

[8] P. Buchholz and J. Kriege. A heuristic approach for fitting MAPs to moments and joint moments. In *QEST*, pages 53–62, 2009.

[9] Giuliano Casale, Eddy Z. Zhang, and Evgenia Smirni. Trace data characterization and fitting for Markov modeling. *Perform. Eval.*, 67(2):61–79, 2010.

[10] A. Cumani. On the canonical representation of homogeneous Markov processes modeling failure-time distributions. *Micorelectronics and Reliability*, 22(3):583–602, 1982.

[11] K. Goseva-Popstojanova and K. S. Trivedi. Effects of failure correlation on software in operation. In *PRDC*, pages 69–76. IEEE Computer Society, 2000.

[12] Qi-Ming He. The versatility of MMAP[K] and the MMAP[K]/G[K]/1 queue. *Queueing Syst.*, 38(4):397–418, 2001.

[13] Qi-Ming He and Attahiru Sule Alfa. The MMAP[K]/PH[K]/1 queues with a last-come-first-served preemptive service discipline. *Queueing Syst.*, 29(2-4):269–291, 1998.

[14] Qi-Ming He and Marcel Neuts. Markov arrival processes with marked transitions. *Stochastic Processes and their Applications*, 74:37–52, 1998.

[15] A. Heindl, G. Horváth, and K. Gross. Explicit inverse characterizations of acyclic MAPs of second order. In A. Horváth and M. Telek, editors, *EPEW*, volume 4054 of *Lecture Notes in Computer Science*, pages 108–122. Springer, 2006.

[16] A. Horváth, G. Horváth, and M. Telek. A traffic based decomposition of two-class queueing network with priority service. *Computer Networks*, 53(8):1235–1248, 2009.

[17] 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 M. Calzarossa and S. Tucci, editors, *Performance*, volume 2459 of *Lecture Notes in Computer Science*, pages 405–434. Springer, 2002.

[18] A. Horváth and M. Telek. Phfit: A general purpose phase type fitting tool. In *Performance Tools 2002*, volume 2324 of *LNCS*, pages 82–91. Springer, 2002.

[19] A. Horváth and M. Telek. Matching more than three moments with acyclic phase type distributions. *Stochastic Models*, 23:167–194, 2007.

[20] G. Horváth, M. Telek, and P. Buchholz. A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag-correlation. In *QEST*, pages 124–133. IEEE CS Press, 2005.

[21] R. El Abdouni Khayari, R. Sadre, and B. Haverkort. Fitting world-wide web request traces with the EM-algorithm. *Performance Evaluation*, 52:175–191, 2003.

[22] A. Klemm, C. Lindemann, and M. Lohmann. Modeling IP traffic using the batch Markovian arrival process. *Perform. Eval.*, 54(2):149–173, 2003.

[23] J. Lambert, B. Van Houdt, and C. Blondia. Queues in docsis cable modem networks. *Computers & OR*, 35(8):2482–2496, 2008.

[24] C. L. Lawson and B. J. Hanson. *Solving Least Squares Problems*. Prentice-Hall, 1974.

[25] D. M. Lucantoni. The BMAP/G/1 queue: A tutorial. In L. Donatiello and R. D. Nelson, editors, *Performance/SIGMETRICS Tutorials*, volume 729 of *Lecture Notes in Computer Science*, pages 330–358. Springer, 1993.

[26] S. Mocanu and C. Commault. Sparse representations of phase-type distributions. *Stochastic Models*, 15:759–778, 1999.

[27] M. F. Neuts. A versatile Markovian point process. *Journal of Applied Probability*, 16:764–779, 1979.

[28] T. Osogami and M. Harchol-Balter. Closed form solutions for mapping general distributions to quasi-minimal PH distributions. *Perform. Eval.*, 63(6):524–552, 2006.

[29] Andriy Panchenko and Axel Thümmler. Efficient phase-type fitting with aggregated traffic traces. *Perform. Eval.*, 64(7-8):629–645, 2007.

[30] V. Paxson and S. Floyd. Wide-area traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, 3:226–244, 1995.

[31] Alma Riska and Erik Riedel. Long-range dependence at the disk drive level. In *QEST*, pages 41–50. IEEE Computer Society, 2006.

[32] P. Salvador, A. Pacheco, and R. Valadas. Modeling IP traffic: joint characterization of packet arrivals and packet sizes using BMAPs. *Computer Networks*, 44(3):335–352, 2004.

[33] S. Suvrit. Source of the non negative least squares algorithm from [24]. http://www.kyb.tuebingen.mpg.de/bs/people/suvrit/ work/progs/nnls.html.

[34] M. Telek and G. Horváth. A minimal representation of Markov arrival processes and a moments matching method. *Perform. Eval.*, 64(9-12):1153–1168, 2007.

[35] A. Thümmler, P. Buchholz, and M. Telek. A novel approach for phase-type fitting with the EM algorithm. *IEEE Trans. Dep. Sec. Comput.*, 3(3):245–258, 2006.

[36] A. van de Liefvoort. The moment problem for continuous distributions. Tech. Rep. WP-CM-1990-02, University of Missouri, 1990.