

# A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation

G. Horváth<sup>1</sup>, P. Buchholz<sup>2</sup>, M. Telek<sup>1</sup>

<sup>1</sup> Dept. of Telecommunications, Technical University of Budapest H-1521 Budapest

<sup>2</sup> Informatik IV, Universität Dortmund D-44221 Dortmund

E-mail: {ghorvath,telek}@hit.bme.hu, peter.buchholz@cs.uni-dortmund.de

## Abstract

*This paper proposes a two-step Markov arrival process (MAP) fitting approach, where the first step is the phase type fitting of the inter-arrival time and the second step is the approximation of the first  $n$  lag correlation values. Depending on the description of the arrival process to approximate various phase type fitting methods can be applied for the first step. In the second step the approximation of the lag correlation values is computed through a non-linear optimization problem.*

*Numerical examples demonstrate the abilities and the limits of the fitting method.*

*Key words: Markov arrival process fitting, inter-arrival time distribution, lag correlation.*

## 1. Introduction

Stochastic modeling of communication and computer systems is usually based on computationally tractable flexible analytical models. With these respects, Markov arrival processes (MAPs) are one of the most attractive candidates to describe traffic processes. They are known to approximate a wide range of processes from renewal ones to long range dependent ones [1, 9], and they allow the use of the computationally effective matrix analytic methods [11]. Unfortunately, the use of MAPs in practical system modeling is still limited by the available MAP fitting methods. There are several effective MAP fitting approaches developed so far.

- Special MAP structures and/or heuristic fitting methods are available for fitting special stochastic processes such as self similar [1], short and long range dependent, multi-fractal [9].
- Explicit expressions are known to fit a second order MAP based on the first 3 moments of the inter-arrival

time and the correlation of the consecutive arrival intervals [8].

- Sample-based Markov modulated Poisson process (MMPP) and MAP fitting methods are proposed based on the expectation maximization (EM) algorithm [15, 4, 14].

But, in spite of these results further research is needed to evaluate the properties of these approaches and to compile a general and robust fitting procedure. The special features of the first and second group of methods compose their limitation as well. The third group is general enough, but the computational complexity and the numerical stability of the sample based MAP fitting methods limit their practical use.

Huge traces are required to describe traffic processes adequately, especially if the processes contain correlation, as it is also demonstrated in an example of this paper. Due to this reason the huge raw traffic traces cannot be directly used for MAP fitting because of the enormous computational complexity. A potential solution is to use a more compact description of the traffic process and to fit a MAP to only this compact description. In this paper we fit a MAP to a traffic process given by the empirical density (or distribution function) of the inter-arrival time distribution and the lag- $k$  ( $k = 1, \dots, K$ ) correlation.

A promising fitting approach was proposed for a wider class of stochastic processes, the matrix exponential arrival process, in [13]. The fitting procedure is composed by two steps. The first step fits an order  $n$  matrix exponential distribution to the first  $2n - 1$  moments of the inter-arrival time distribution and the second step set the first  $2n - 3$  lag correlation values (independent of the first step). This approach is very elegant and effective to fit a matrix exponential arrival process, but the result usually does not have a Markovian representation which seems to reduce its application in practice. To reduce the computational complexity of the sample-based MAP fitting method ([4]) a similar idea is applied in [5], where a two-step EM method is

proposed, such that, an EM method approximates the inter-arrival time distribution with some constraints in the first step and another EM method approximates the lag correlation in the second step. In this paper we present a general framework of two-step (marginal distribution and lag correlation) MAP fitting by relaxing the constraints and the iterative approximation of [5], discuss the problem of different phase type representations of the same distribution, and propose a non-linear optimization based solution method for the second step.

The rest of the paper is organized as follows. Section 2 presents some basic properties of MAPs. The analytical background and the implementation details of the proposed fitting procedure are provided in Section 3 and 4, respectively. A number of numerical examples demonstrate the fitting properties of the proposed approach in Section 5. Finally, Section 6 concludes the paper.

## 2. Basic properties of Markov arrival processes

A Markov arrival process (MAP) is usually defined by two matrices,  $D_0$  and  $D_1$ , such that  $D = D_0 + D_1$  is the generator of the background continuous time Markov chain (CTMC),  $D_0$  contains the transitions of the background CTMC without arrival and  $D_1$  describes the arrival and the associated state transitions. The number of states of the background CTMC,  $m$ , determines the order of the MAP. The row sum of matrices  $D_0$  and  $D_1$  satisfies  $D_0\mathbf{1} = -D_1\mathbf{1}$ , since

$$D_0\mathbf{1} + D_1\mathbf{1} = D\mathbf{1} = 0,$$

where  $\mathbf{1}$  is the column vector of ones, and we utilized the fact that the row sum of a CTMC generator matrix is 0.

The steady state probability vector of the background CTMC,  $\underline{\alpha}$ , is the solution of the linear system  $\underline{\alpha}D = 0, \underline{\alpha}\mathbf{1} = 1$ .

In case of MAPs, the discrete time process embedded at arrival instants plays an important role. The state transition probability matrix of the embedded process is  $P = (-D_0)^{-1}D_1$ . The steady state probability vector of the embedded process,  $\underline{\pi}$ , is the solution of the linear system  $\underline{\pi}P = \underline{\pi}, \underline{\pi}\mathbf{1} = 1$ . The steady state distributions of the original and the embedded processes are related as

$$\underline{\alpha} = \frac{\underline{\pi}(-D_0)^{-1}}{\underline{\pi}(-D_0)^{-1}\mathbf{1}} = \lambda\underline{\pi}(-D_0)^{-1}.$$

In steady state, the inter-arrival time is phase type distributed with initial probability vector  $\underline{\pi}$ , and generator  $D_0$ . Thus, the distribution of the inter-arrival time is

$$P(X < t) = 1 - \underline{\pi}e^{D_0 t}\mathbf{1},$$

and its  $k$ th moment is

$$E(X^k) = k! \underline{\pi}(-D_0)^{-k}\mathbf{1}.$$

The arrival intensity is

$$\lambda = \frac{1}{E(X)} = \frac{1}{\underline{\pi}(-D_0)^{-1}\mathbf{1}} = \underline{\alpha}D_1\mathbf{1}.$$

and the lag- $k$  correlation is computed by

$$\rho_k = \frac{\lambda^2 \underline{\pi}(-D_0)^{-1}P^k(-D_0)^{-1}\mathbf{1} - 1}{2\lambda^2 \underline{\pi}(-D_0)^{-1}(-D_0)^{-1}\mathbf{1} - 1}.$$

## 3. The Fitting Procedure

In this section we summarize the theoretical issues of the fitting procedure. Generally speaking, the main idea of the applied approach is that the  $D_0$  and the  $D_1$  matrices are constructed separately.

- In the first step, the inter-arrival time distribution is fitted by a phase type distribution, which determines the  $D_0$  matrix (the generator of the PH distribution) and the  $\underline{\pi}$  vector (the initial probability vector of the PH distribution)<sup>1</sup>.
- Then, the  $D_1$  matrix is constructed, such that the inter-arrival time distribution of the resulting MAP remains the same, and its lag correlation function approximates the one of the trace.

### 3.1. Constructing the $D_0$ matrix and the $\underline{\pi}$ vector

The first step of the procedure is a phase type fitting problem for which we refer to [2, 3, 10, 7, 16]. Here we only recall that the various PH fitting methods can handle different input data. The inter-arrival time distribution of the original process can be given with its pdf or cdf, samples or by a given number of moments. The methods in [2, 7, 16] fit a phase type distribution to a set of samples. The methods in [3, 10] allows to fit to both, pdf/cdf and set of samples. Moments based phase type fitting is available up to 3 moments (which is used in [8]).

All fitting methods in [3, 10, 7, 16] provide acyclic phase type distributions, which, in general, have infinitely many identical representations [6]. For a unique representation of acyclic phase type distributions Cumani introduced 3 *canonical forms*. Appendix A presents an equivalent transformation of the  $D_0$  matrix and the  $\underline{\pi}$  vector from canonical form 1 to a modified representation which allows a higher flexibility in lag- $k$  correlation fitting.

<sup>1</sup> This approach is different from the one in [8], where the first 3 moments of the inter-arrival time determine the  $D_0$  matrix, but the  $\underline{\pi}$  vector is a function of the lag-1 correlation.

### 3.2. Constraints of the $D_1$ matrix

The  $D_1$  matrix has to satisfy the following two constraints to maintain the inter-arrival time distribution determined in the first step:

$$\mathbf{C1: } D_1 \mathbf{1} = -D_0 \mathbf{1},$$

$$\mathbf{C2: } \pi(-D_0)^{-1} D_1 = \underline{\pi}.$$

We can formulate these constraints as a linear system of equations. We introduce column vector  $\underline{x}$  (of size  $m^2$ ), which is composed by the columns of matrix  $D_1$ :

$$D_1 = \left[ \begin{array}{c|c|c|c} \{D_1\}_1 & \{D_1\}_2 & \dots & \{D_1\}_m \\ \hline \hline \hline \hline \end{array} \right] \rightarrow \underline{x} = \begin{bmatrix} \{D_1\}_1 \\ \{D_1\}_2 \\ \vdots \\ \{D_1\}_m \end{bmatrix}$$

All possible  $\underline{x}$  vectors (thus,  $D_1$  matrices) satisfying constraints **C1** and **C2** are the solutions of the following system of linear equations with coefficient matrix  $\mathcal{A}$ :

$$\underbrace{\left[ \begin{array}{c|c|c|c} \begin{bmatrix} I_{m \times m} \\ \gamma \end{bmatrix} & \begin{bmatrix} I_{m \times m} \\ \gamma \end{bmatrix} & \dots & \begin{bmatrix} I_{m \times m} \\ \gamma \end{bmatrix} \\ \hline \hline \hline \hline \end{array} \right]}_{\mathcal{A}_{2m \times m^2}} \cdot \underline{x} = \underbrace{\begin{bmatrix} d \\ - \\ \pi \\ \vartheta \end{bmatrix}}_{b_{2m}}, \quad (1)$$

where  $d = -D_0 \mathbf{1}$  and  $\gamma = \pi(-D_0)^{-1}$ . The first  $m$  lines of  $\mathcal{A}$  correspond to constraint **C1**, and the second  $m$  lines are related to **C2**.

A proper  $D_1$  matrix (i.e.,  $\underline{x}$  vector) satisfies the following set of linear equations and inequalities:

$$\begin{aligned} \mathcal{A} \underline{x} &= \underline{b}, \\ \underline{x} &\geq 0. \end{aligned} \quad (2)$$

In general, the  $\mathcal{A} \underline{x} = \underline{b}$  equation is determined for  $m = 2$  and under-determined for  $m \geq 3$ , since we have  $2m$  equations and  $m^2$  unknowns. Consequently the  $D_0$  matrix and the  $\pi$  vector completely determines the  $D_1$  matrix when  $m = 2$ . For  $m \geq 3$ , one can use, for example, the simplex algorithm to find a solution of (2).

### 3.3. Exact Lag-1 Correlation Fitting

The lag-1 correlation can also be expressed as a linear constraint (while the higher lag correlations results in non-linear constraints).

$$\mathbf{C3: } \rho_1 = \frac{\lambda^2 \pi (-D_0)^{-2} D_1 (-D_0)^{-1} \mathbf{1} - 1}{2\lambda^2 \pi (-D_0)^{-2} \mathbf{1} - 1},$$

with  $\underline{f} = (-D_0)^{-1} \mathbf{1}$  that is

$$\lambda^2 \pi (-D_0)^{-2} D_1 \underline{f} = \rho_1 (2\lambda^2 \pi (-D_0)^{-2} \mathbf{1} - 1) + 1,$$

which can be concatenated to matrix  $\mathcal{A}$  and vector  $\underline{b}$  as

$$\underbrace{\left[ \begin{array}{c|c|c|c} \begin{bmatrix} I_{m \times m} \\ \gamma \end{bmatrix} & \begin{bmatrix} I_{m \times m} \\ \gamma \end{bmatrix} & \dots & \begin{bmatrix} I_{m \times m} \\ \gamma \end{bmatrix} \\ \hline \hline \hline \hline \end{array} \right]}_{\mathcal{A}_{(2m+1) \times m^2}} \cdot \underline{x} = \underbrace{\begin{bmatrix} d \\ - \\ \pi \\ \vartheta \end{bmatrix}}_{b_{2m+1}}, \quad (3)$$

where  $\delta = \lambda^2 \pi (-D_0)^{-2}$ ,  $f_i$  is the  $i$ th element of vector  $\underline{f}$  and  $\vartheta = \rho_1 (2\lambda^2 \pi (-D_0)^{-2} \mathbf{1} - 1) + 1$ .

### 3.4. Fitting More Lag Correlations

To fit more lag- $k$  correlation values, we can define an optimization problem with the linear constraints (2) such that a properly chosen goal function ensures the approximation of higher lag correlation values. This way the fitting of a given number of lag- $k$  correlations is a linearly constrained non-linear optimization problem.

We applied the objective function  $c(\underline{x})$ , which is the squared difference between the lag- $k$  correlations of the original process ( $\hat{\rho}_k$ ) and the fitted MAP ( $\rho_k$ ) weighted with  $w_k$ :

$$c(\underline{x}) = \sum_{k=2}^K w_k (\rho_k - \hat{\rho}_k)^2. \quad (4)$$

The largest lag correlation coefficient considered in this objective function is the lag- $K$  correlation coefficient. The weights can be used, e.g., to increase the importance of the accuracy of lower lag- $k$  correlation with respect to higher ones or vice-versa.

### 3.5. Illustration

We illustrate the constraints imposed by  $D_0$  and  $\pi$  and the flexibility of  $D_1$  in lag-1 correlation fitting via small examples.

*Illustration 1:* Some of the phase type fitting tools [3, 10] provide the results in canonical form 1 [6], which has the following structure:

$$\underline{\pi} = [0.3, 0.2, 0.5], \quad D_0 = \begin{bmatrix} -3 & 3 & 0 \\ 0 & -5 & 5 \\ 0 & 0 & -7 \end{bmatrix} \Rightarrow \underline{d} = \begin{bmatrix} 0 \\ 0 \\ 7 \end{bmatrix}.$$

In this case, the first two rows of the  $D_1$  matrix should be zero, according to constraint **C1**. Consequently the **C1**

and **C2** constraints completely determine the  $D_1$  matrix and there is zero degree of freedom to fit the lag- $k$  correlation. Using the equivalent transformation of Appendix A (with  $a_3 = 0.5$  and  $a_2 = 0.6$ ) we represent the same phase type distribution as

$$\underline{\pi}' = [.618, .026, .356], D_0' = \begin{bmatrix} -3 & 1.5 & 0 \\ 0 & -5 & 2 \\ 0 & 0 & -7 \end{bmatrix} \Rightarrow \underline{d}' = \begin{bmatrix} 1.5 \\ 3 \\ 7 \end{bmatrix}.$$

With this representation, all elements of the  $\underline{d}'$  vector differ from zero and the  $\mathcal{A}\underline{x} = \underline{b}$  equation is under-determined, which provides some degrees of freedom to fit the lag- $k$  correlation.

Similar problems arise when the  $\underline{\pi}$  vector contains zero elements. E.g., if  $\pi_i = 0$  then the  $i$ th column of the  $D_1$  matrix should be zero, according to constraint **C2**. This example also indicates that there is a trade off between the  $(\underline{\pi}, D_0)$  and the  $(\underline{\pi}', D_0')$  representations. With the  $(\underline{\pi}, D_0)$  representation the  $\underline{d}$  vector contains only one non-zero element, but the  $\underline{\pi}$  vector contains 3 significant elements, instead with the  $(\underline{\pi}', D_0')$  representation the  $\underline{d}'$  vector contains 3 significant non-zero elements, but the  $\underline{\pi}'$  vector contains only 2 significant elements and  $\pi_2'$  is negligible.

*Illustration 2:* Considering the phase type distribution defined by

$$\underline{\pi} = [0.4995, 0.48918, 0.011324],$$

$$D_0 = \begin{bmatrix} -3.721 & 0.5 & 0.02 \\ 0.1 & -1.206 & 0.005 \\ 0.001 & 0.002 & -0.031 \end{bmatrix}, \Rightarrow \underline{d} = \begin{bmatrix} 3.201 \\ 1.101 \\ 0.028 \end{bmatrix}$$

the minimal and the maximal lag-1 correlations are calculated by the simplex method since the goal function,  $\rho_1$ , is linear. They are obtained by the following  $D_{1min}$  and  $D_{1max}$  matrices

$$\begin{bmatrix} 1.562 & 1.562 & 0.077 \\ 0.55 & 0.551 & 0 \\ 0.0243 & 0.0037 & 0 \end{bmatrix}, \begin{bmatrix} 3.143 & 0.058 & 0 \\ 0.0805 & 1.0205 & 0 \\ 0 & 0.0069 & 0.0211 \end{bmatrix},$$

respectively, and the associated lag- $k$  correlations are:

	min	max
1	-0.0101451904742418	0.326452359342335
2	0.00315598349944327	0.221637940730464
3	-9.52303122836937e-005	0.150462438659736
4	2.0994680540877e-005	0.102133708727262
5	-8.0989922249666e-007	0.0693209104222013

## 4. Implementation of the fitting algorithm

### 4.1. Constructing $D_0$

The inter-arrival distribution of the trace is approximated by a phase type distribution, whose transient generator gives

$D_0$ . However, it does matter what the structure of the applied phase type is. According to our experiments, the structure of the phase type distribution determines the feasible region of lag- $k$  correlation, because constraints **C1** and **C2** provide different "degrees of freedom" for the lag- $k$  correlation fitting with different phase type structures. The "degree of freedom" depends on the number of significant non-zero elements in  $\underline{\pi}$  and  $\underline{d}$ .

Unfortunately, most of the general phase type and acyclic phase type structures used in phase type fitting procedures have only few non-zero elements in  $\underline{\pi}$  and  $\underline{d}$ . E.g., *VERPH* [3] and *PhFit* [10] use the canonical representation of acyclic phase type distributions in which  $\underline{d}$  has exactly 1 non-zero element. An exception is the *EM-PHT* tool [2], which allows the user to define the structure (to set the non-zero elements). According to our experience it does not help much to use a phase type structure with a lot of non-zero elements, because the fitting procedure sets most of the non-zero elements very close to 0 and this way we get only a small number of significant non-zero elements in  $\underline{\pi}$  and  $\underline{d}$ .

A possible solution is to apply special phase type structures, e.g., hyper-exponential [7] or mixture of Erlang distributions [16]. For example, in case of a hyper-exponential structure the elements of the  $\underline{\pi}$  and  $\underline{d}$  vectors are all non-zero. A limitation of using hyper-exponential distribution is that the pdf is monotone decreasing and the squared coefficient of variation is greater than one. For those cases in which the marginal distribution has these properties the fitting of a hyper-exponential distribution (e.g., with [2] or [7]) is usually the best solution.

For general marginal distributions we recommend to fit a generalized Erlang distribution or an acyclic phase type distribution (e.g., with [16] and [10]). In case of acyclic phase type distribution fitting the resulted canonical form 1 can be transformed into a more appropriate equivalent representation according to Appendix A. Usually the best representation is the one which results in the highest number of significant non-zero elements in  $\underline{\pi}$  and  $\underline{d}$ .

### 4.2. Constructing $D_1$

The construction of  $D_1$  requires the solution of a linear constrained nonlinear optimization problem. There are lots of general software tools to solve such nonlinear problems, but those, we tried, failed to fit the lag- $k$  correlation reasonable well. It seems the surface is hilly and the methods sticks in local optimum close to the initial point. Based on this assumption we implemented a simple optimization method. The optimization part simple linearizes the objective function and runs the simplex algorithm within a given small radius of the parameters in each step. This method also stops at the closest local optimum, but the crucial feature of our procedure is that we generate a relative large number of initial guesses and start from the best of them. It

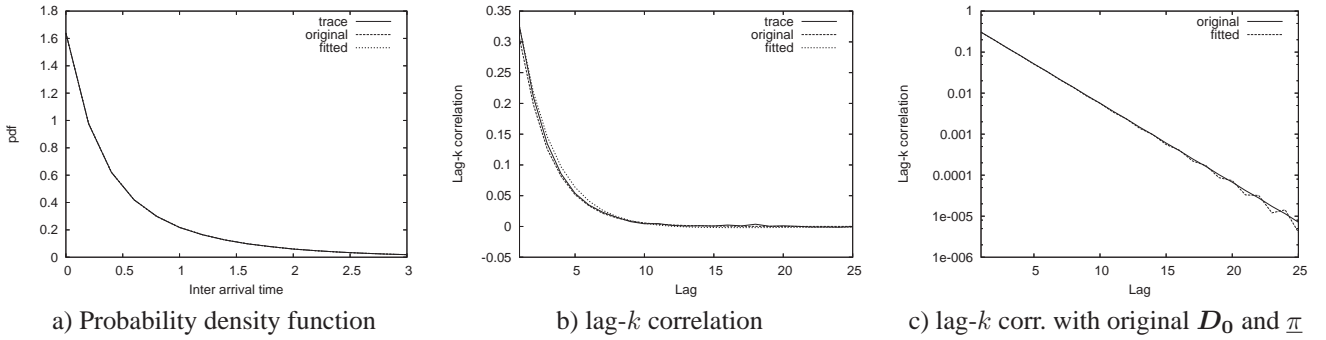


Figure 1. Results of Example 1

seems the optimization phase does not improve the solution significantly, but the starting point determines the goodness of fitting.

Our initial points are the randomly chosen corners of the  $\mathcal{A}\underline{x} = \underline{b}, \underline{x} \geq 0$  surface. We use the simplex method to calculate these corner points with randomly chosen linear goal functions (i.e., with random weights of the parameters).

For the results of this paper we calculated 2500 corner points. On a P4 (2.8GHz, 512MB RAM) machine, the computation time of 2500 corner points were 16 s (PH(3)), 28 s (PH(4)), 56 s (PH(5)), 101 s (PH(6)) and 515 s (PH(8)). We found that the set of solutions reachable from the 2500 corner points can be divided into clearly distinct “good” and “bad” subsets. The majority of the solutions are bad, but some of them ( $\sim 10$ ) are significantly better. These better solutions result in different  $D_1$  matrices with very close optimum and very similar lag- $k$  correlation.

Throughout the numerical examples we used our MATLAB implementation of the proposed method. The default weights of the goal function (4) were  $w_k = (K - k + 1)$  with  $K = 20$ .

## 5. Numerical Experiments

To demonstrate the fitting properties of the proposed approach, we evaluated four numerical examples. The first two ones approximate given MAPs. We check how close the result of the fitting is in terms of inter-arrival time distribution and lag- $k$  correlation. In the third and fourth example real traffic traces are fitted. These traces are the lbl-3 and the bcpAug-89 traces [17], which are commonly used for testing phase type distribution and MAP fitting.

### 5.1. Example 1

In this example we consider a 3-state MAP with the following matrices:

$$D_0 = \begin{bmatrix} -3.721 & 0.5 & 0.02 \\ 0.1 & -1.206 & 0.005 \\ 0.001 & 0.002 & -0.031 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0.2 & 3 & 0.001 \\ 1 & 0.1 & 0.001 \\ 0.005 & 0.003 & 0.02 \end{bmatrix}.$$

First we took the original  $D_0$  matrix and  $\underline{\pi}$  vector, and executed the  $D_1$  fitting method (with exact lag-1 fitting), which gave the following  $D_1$  matrix:

$$D_1 = \begin{bmatrix} 0 & 3.1967 & 0.0043246 \\ 1.0686 & 0.032399 & 0 \\ 0 & 0.0080394 & 0.019961 \end{bmatrix}.$$

As Figure 1c) shows, that the lag- $k$  correlation of the original and the fitting MAPs are the same up to lag-15. After that the lag correlation of the fitting MAP fluctuates. It turns out that the eigenvalues of the  $P$  matrix of the original MAP are 1,  $-0.675519$ ,  $0.641501$  and the ones of the fitting MAP are 1,  $-0.774415$ ,  $0.640013$ . The fluctuation starts when the effect of the 3rd (positive) eigenvalue vanishes with respect to the effect of the 2nd (negative) one [11]. Since the difference of the 2nd and the 3rd eigenvalues of the fitting MAP is larger it starts fluctuating earlier. The original MAP starts fluctuating at lag-77.

Next, we generate a trace of 500000 arrivals by the original MAP and fit [7] a hyper-exponential distribution (HED) to the experimental inter-arrival time distribution to obtain  $D_0$  and  $\underline{\pi}$ . Using these the  $D_1$  fitting method with exact lag-1 fitting provided the following MAP:

$$D_0 = \begin{bmatrix} -1.1728 & 0 & 0 \\ 0 & -3.6435 & 0 \\ 0 & 0 & -0.03132 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 1.1311 & 0.041664 & 0 \\ 0.14273 & 3.4511 & 0.049643 \\ 0.010006 & 0 & 0.021314 \end{bmatrix}.$$

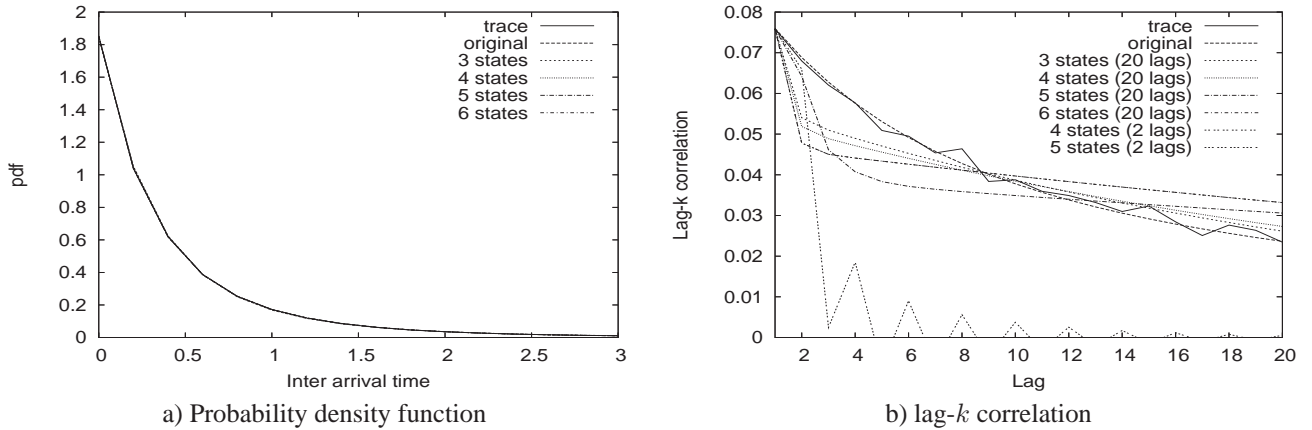


Figure 2. Results of Example 2

Figure 1a) and 1b) depicts the fitting results. The inter-arrival time approximation is very accurate. The difference of the pdf-s is not visible in the figure. The lag- $k$  correlation fitting is reasonably accurate as well.

### 5.2. Example 2

In this example we approximate the following MMPP:

$$D_0 = \begin{bmatrix} -0.1 & 0.1 & 0 & 0 & 0 & 0 \\ 0.2 & -1.3 & 0.1 & 0 & 0 & 0 \\ 0 & 0.2 & -2.3 & 0.1 & 0 & 0 \\ 0 & 0 & 0.2 & -3.3 & 0.1 & 0 \\ 0 & 0 & 0 & 0.2 & -4.3 & 0.1 \\ 0 & 0 & 0 & 0 & 0.2 & -5.2 \end{bmatrix},$$

$$D_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix}.$$

First we kept the original  $D_0$  and  $\pi$  again. Surprisingly the  $D_1$  fitting procedure gave back the  $D_1$  matrix of the original MMPP up to 10 tangible digits.

The considered MMPP has a very special structure which suggests that smaller MAPs can approximate its behaviour. Similar to the previous example, we generated a trace of 500000 arrivals from the MMPP and studied how can we approximate the 6-state MMPP with smaller MAPs.

We approximated the inter-arrival time distribution with a HED [7] again. The resulted 3 and 4-state MAPs are

$$D_0 = \begin{bmatrix} -1.36856 & 0 & 0 \\ 0 & -3.43205 & 0 \\ 0 & 0 & -0.082935 \end{bmatrix},$$

$$D_1 = \begin{bmatrix} 1.1778 & 0.033576 & 0.15722 \\ 0.06028 & 3.3718 & 0 \\ 0.069005 & 0 & 0.01393 \end{bmatrix},$$

$$D_0 = \begin{bmatrix} -1.24582 & 0 & 0 & 0 \\ 0 & -1.3979 & 0 & 0 \\ 0 & 0 & -0.0831938 & 0 \\ 0 & 0 & 0 & -3.49944 \end{bmatrix},$$

$$D_1 = \begin{bmatrix} 0 & 0.034812 & 0.019984 & 1.191 \\ 0.012632 & 1.2142 & 0.15471 & 0.016357 \\ 0 & 0.069023 & 0.014171 & 0 \\ 0 & 0.054796 & 0 & 3.4446 \end{bmatrix}.$$

The corresponding plots, together with the fitting 5 and 6-state MAPs, are shown in Figure 2a) and 2b). The pdf is closely approximated with 3 states already. The lag- $k$  correlation is less accurate, although the higher lags are quite close to the original values. To improve the accuracy of low lag correlation values we set the goal function to optimize only the lag-2 correlation ( $c(\underline{x}) = (\rho_2 - \hat{\rho}_2)^2$ ). This setting resulted in accurate lag-2 correlation on the price of poorer higher lag correlation. We suspect that the poor approximation of the correlation is due to the bad representation of the marginal distribution, since the original  $D_0$  matrix and  $\pi$  vector provided a close correlation fitting.

The difference of the “original” and the “trace” curves indicates that the 500000 samples of the trace were not enough for an accurate empirical lag- $k$  correlation. The empirical lag- $k$  correlation is calculated as

$$\hat{\rho}_k = \frac{1}{(N - k - 1)\hat{\sigma}^2} \sum_{i=1}^{N-k} (t_i - \hat{\mu})(t_{i+k} - \hat{\mu}),$$

where  $N$  is the number of samples,  $t_i$  is the  $i$ th sample and  $\hat{\mu}$  and  $\hat{\sigma}^2$  are the sample mean and variance, respectively.

### 5.3. Example 3

A more challenging problem is to characterize real-life traces with MAPs. In this example we approximate the LBL-TCP-3 trace, which contains two hours’ worth of all wide-area TCP traffic ( $1.8 \cdot 10^6$  TCP packets) between the Lawrence Berkeley Laboratory and the rest of the world



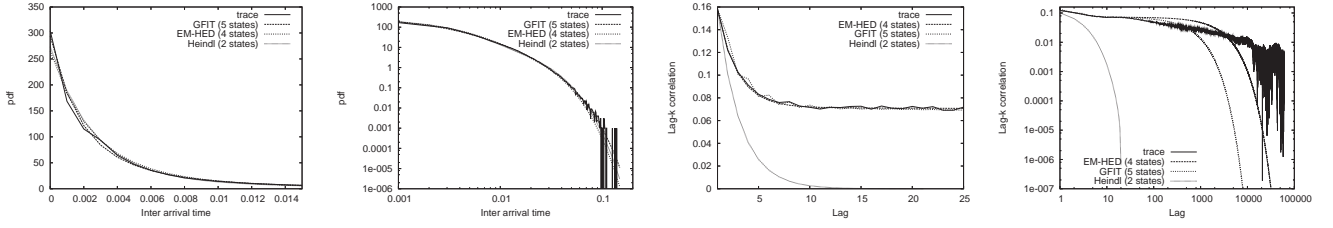


Figure 3. Density function and lag- $k$  correlation of Example 3

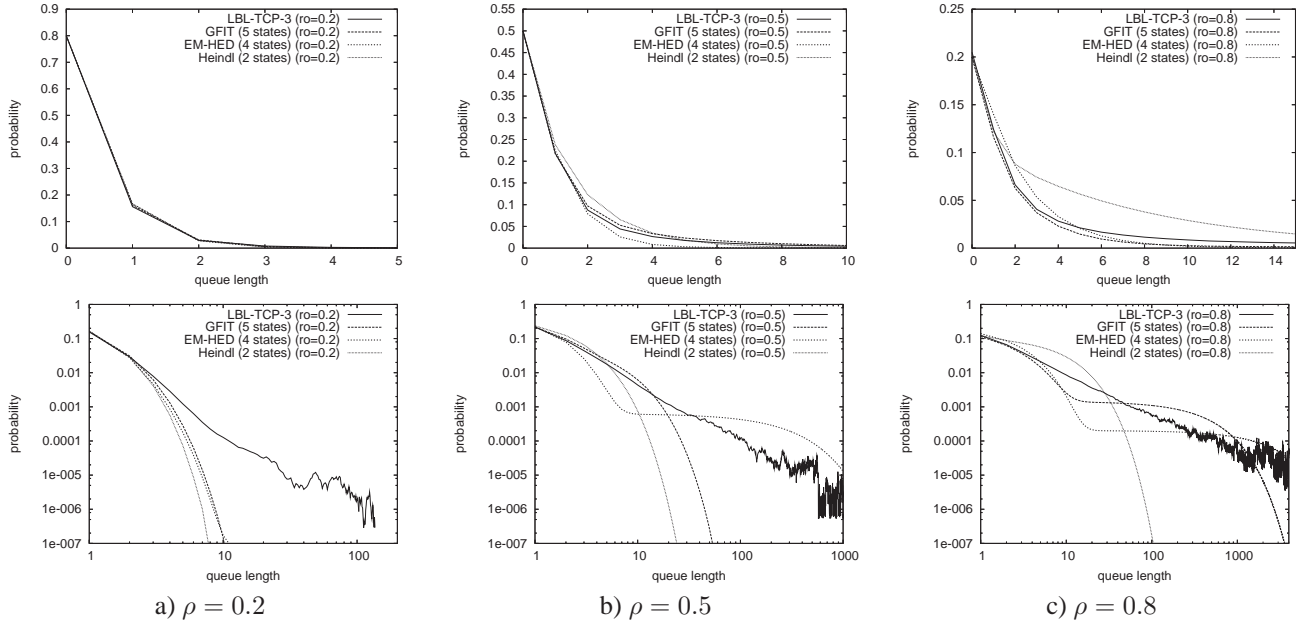


Figure 4. Queue length distributions of Example 3

[17]. We applied two PH fitting tools to fit the inter-arrival time distribution: EM-HED with 4 states [7] and GFIT with 5 states [16] (because we got close marginal fitting with 4 and 5 states, respectively). The result of MAP fitting based on EM-HED is:

$$D_0 = \begin{bmatrix} -508.11 & 0 & 0 & 0 \\ 0 & -526.82 & 0 & 0 \\ 0 & 0 & -112.88 & 0 \\ 0 & 0 & 0 & -292.87 \end{bmatrix},$$

$$D_1 = \begin{bmatrix} 281.9 & 226.06 & 0 & 0.15872 \\ 526.66 & 0.024505 & 0 & 0.13422 \\ 0 & 0 & 82.094 & 30.79 \\ 0.056728 & 0 & 38.799 & 254.01 \end{bmatrix}.$$

GFIT uses a mixture of Erlang distributions instead of HEDs, but in this particular case GFIT found the hyper-exponential structure to be the best. Using the result of

GFIT, the MAP fitting algorithm provided:

$$D_0 = \begin{bmatrix} -90.779 & 0 & 0 & 0 & 0 \\ 0 & -137.54 & 0 & 0 & 0 \\ 0 & 0 & -213.51 & 0 & 0 \\ 0 & 0 & 0 & -338.05 & 0 \\ 0 & 0 & 0 & 0 & -679.54 \end{bmatrix},$$

$$D_1 = \begin{bmatrix} 0.47845 & 90.283 & 0.018233 & 0 & 0 \\ 90.045 & 0.082362 & 0.064649 & 47.035 & 0.31731 \\ 0.023153 & 0.27187 & 119.66 & 0 & 93.553 \\ 0 & 69.036 & 0 & 269.018 & 0 \\ 0 & 0.30414 & 173.73 & 0.030199 & 505.47 \end{bmatrix}.$$

In the comparison of fitting MAPs, we also considered the result of the explicit order 2 MAP fitting method of Heindl [8]. The inter-arrival time distributions and the lag- $k$  correlations of the trace and the fitting MAPs are depicted in Figure 3.

We also studied the queueing behaviour resulted by the fitting MAPs, such that we compared the simulated queue

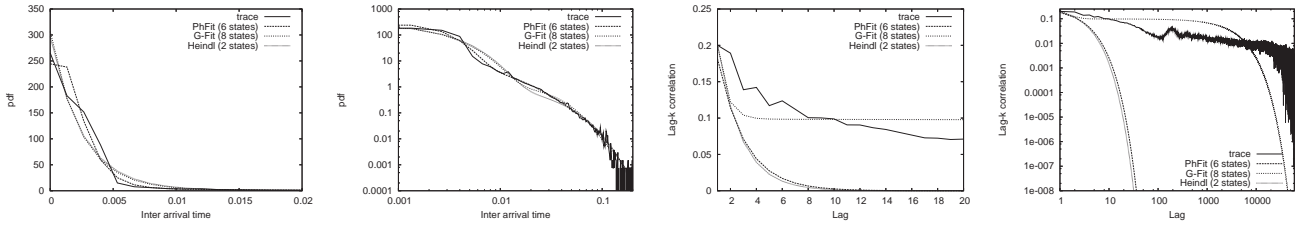


Figure 5. Density function and lag- $k$  correlation of Example 4

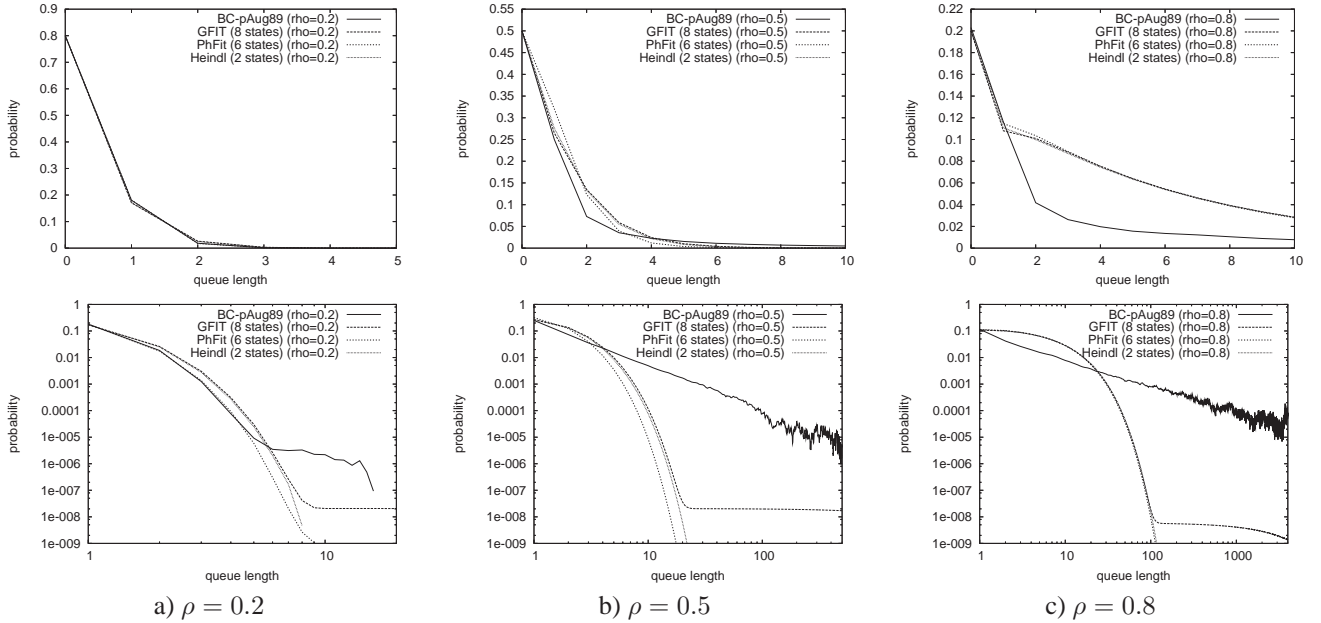


Figure 6. Queue length distributions of Example 4

length distribution of the TRACE/D/1 queue (formed by fixed size TCP packets) with the numerical analysis of the fitting MAP/D/1 queues [12] at different utilization levels. We set the utilization level with appropriate mean service time (i.e. link capacity). The results are depicted in Figures 4. It seems that the close lag- $k$  correlation fitting up to lag-100 allows to capture the slowly decaying tendency of the queue length distribution at high utilization.

#### 5.4. Example 4

Another often used benchmark in traffic modeling is the BC-pAug89 trace, which contains  $10^6$  packet arrivals seen on an Ethernet at the Bellcore Morristown Research and Engineering facility [17]. The density function of this trace has a more complex shape compared to the one in the previous example. Due to this shape we failed to fit the inter-arrival time distribution with HED and with a small number of phases. Finally, we fitted the inter-arrival time distribution using the GFIT [16] and the PhFit [10] tools. GFIT provided a reasonable good solution with 8 states (3 Erlang

alternatives). With PhFit, we could achieve reasonable good fitting with 6 states (3 states to fit the body and 3 states to fit the tail), but only after a longer search of optimal user defined fitting parameters (like the limits of the body and tail fitting). Figure 5 shows the result of the inter-arrival time distribution fitting with linear and logarithmic scales.

Unfortunately the obtained PH distributions did not leave too many degrees of freedom for  $D_1$  fitting. In case of GFIT  $\underline{\pi}$  and  $\underline{d}$  contained only 3 non-zero elements, while in case of PhFit all elements of  $\underline{\pi}$  were positive, but  $\underline{d}$  contained only 4 non-zero elements. Interestingly, with the PH distribution given by GFIT we could fit a MAP with slower decaying lag- $k$  correlation compared to the one based on PhFit, which already missed to fit the first lag exactly. The results are depicted in Figure 5. Note that we only took the first 20 lags into account during the optimization. The log-log plot confirms that the lag- $k$  correlation of MAPs always has exponential tail decay, which makes hard to capture the slowly decaying correlation behaviour of the BC-pAug89





The new initial probabilities are the solutions of the following set of linear equations:

$$p_k = \sum_{n=k}^m x_n \sum_{i=0}^{k-1} \prod_{j=0}^{i-1} (1 - a_{n-j}) a_{n-i} q_n(i, 0, k), \quad (5)$$

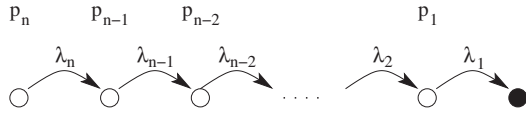
where the  $q_n(i, \ell, k)$  coefficients are defined by the following recursion:

$$q_n(i, \ell, k) = \frac{\lambda_{n-i}}{\lambda_{\ell+1}} q_n(i-1, \ell+1, k) + \left(1 - \frac{\lambda_{n-i}}{\lambda_{\ell+1}}\right) q_n(i, \ell+1, k),$$

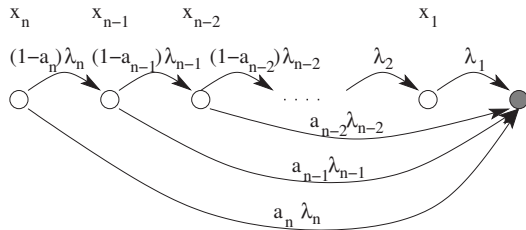
$$q_n(i, n-i, k) = \begin{cases} 1, & k = n, \\ 0, & k \neq n, \end{cases}$$

$$q_n(0, \ell, k) = \begin{cases} \frac{\lambda_n}{\lambda_{\ell+1}}, & k = \ell+1, \\ \left(1 - \frac{\lambda_n}{\lambda_{\ell+1}}\right) q_n(0, \ell+1, k), & k \neq \ell+1. \end{cases}$$

**Proof**  $\pi(sI - D_0)^{-1} \mathbf{1} = \pi'(sI - D'_0)^{-1} \mathbf{1}$ .



**Figure 7. Canonical form 1 of APH distributions**



**Figure 8. The APH distribution after the transformation**

## References

[1] A. T. Andersen and B. F. Nielsen. A markovian approach for modeling packet traffic with long-range dependence. *IEEE Journal on Selected Areas in Communications*, 16(5):719–732, 1998.

[2] S. Asmussen and O. Nerman. Fitting Phase-type distributions via the EM algorithm. In *Proceedings: "Symposium i Advent Statistik"*, pages 335–346, Copenhagen, 1991.

[3] 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 Science Publishers, 1992.

[4] P. Buchholz. An em-algorithm for map fitting from real traffic data. In *Computer Performance Evaluation Modelling Techniques and Tools (TOOLS)*, LNCS 2794, pages 218–236, Urbane, IL, USA, 2003. Springer.

[5] P. Buchholz and A. Panchenko. A two-step em-algorithm for map fitting. In *19th Int. Symp. on Computer and Information Sciences (ISCIS'2004)*, LNCS 3280, pages 217–227, Antalya, Turkey, 2004. Springer.

[6] A. Cumani. On the canonical representation of homogeneous Markov processes modelling failure-time distributions. *Microelectronics and Reliability*, 22:583–602, 1982.

[7] R. El Abdouni Khayari, R. Sadre, and B. Haverkort. Fitting world-wide web request traces with the em-algorithm. In *Proc. of SPIE*, volume 4523, pages 211–220, Denver, USA, 2001.

[8] A. Heindl. Inverse characterization of hyperexponential map(2)s. In *11th Int. Conf. on Analytical and Stochastic Modelling Techniques and Applications (ASMTA)*, pages 183–189, Magdeburg, Germany, 2004.

[9] A. Horváth and M. Telek. A Markovian point process exhibiting multifractal behaviour and its application to traffic modeling. In G. Latouche and P. Taylor, editors, *Matrix analytic methods: Theory and applications, MAM4*, pages 183–208. World Scientific, July 2002.

[10] A. Horváth and M. Telek. PhFit: A general purpose phase type fitting tool. In *Tools 2002*, pages 82–91, London, England, April 2002. Springer, LNCS 2324.

[11] G. Latouche and V. Ramaswami. *Introduction to matrix analytic methods in stochastic modeling*. SIAM, 1999.

[12] D. M. Lucantonio. New results on the single server queue with a batch markovian arrival process. *Stochastic models*, 7(1):1–46, 1991.

[13] K. Mitchell and Appie van de Liefvoort. Approximation models of feed-forward G/G/1/N queueing networks with correlated arrivals. *Performance Evaluation*, 52(2-4):137–152, 2003.

[14] Alma Riska, Mark S. Squillante, Shun-Zheng Yu, Zhen Liu, and Li Zhang. Matrix-analytic analysis of a MAP/PH/1 queue fitted to web server data. In *Matrix-Analytic Methods, MAM4*, pages 335–356. World Scientific, 2002.

[15] T. Ryden. An em algorithm for estimation in markov-modulated poisson processes. *Computational Statistics and Data Analysis*, 21:431–447, 1996.

[16] A. Thümmel, P. Buchholz, and M. Telek. A novel approach for fitting probability distributions to trace data with the EM algorithm. In *Int. Conf. on Dependable Systems and Networks (DSN)*, Yokohama, Japan, 2005. IEEE CS.

[17] The internet traffic archive. <http://ita.ee.lbl.gov/index.html>.