Matching marginal moments and lag autocorrelations with MAPs

Gábor Horváth
Budapest University of Technology and Economics, Department of Networked Systems and Services
MTA-BME Information systems research group
Inter-University Center of Telecommunications and Informatics, Debrecen
Magyar Tudósok krt. 2, 1117 Budapest, Hungary
ghorvath@hit.bme.hu

ABSTRACT

This paper presents a procedure that constructs a Markovian Arrival Process (MAP) based on the mean, the squared coefficient of variation and the lag-1 autocorrelation of the inter-arrival times. This method always provides a valid MAP without posing any restrictions on the three input parameters. Besides matching these three parameters, it is possible to match the third moment of the inter-arrival times and the decay of the autocorrelation function as well, if they fall into the given (very wide) bounds.

1. INTRODUCTION

Markovian Arrival Processes (MAPs) are versatile tools for modeling correlated workloads. They can be efficiently applied in several fields of performance evaluation of various systems. Once the traffic is represented by a MAP, it is easy to generate random samples from it ([11], [13]), which is beneficial in simulations and measurement based performance analysis. It is also well known that queueing models where the arrival and/or service process are given by MAPs can be analyzed in a numerically efficient way by matrix geometric methods [14].

However, the accuracy of the MAP based models depends on how well the traffic is represented by the MAP. There are several fitting and matching procedures published in the literature to obtain a MAP based on various properties of the workload. A popular fitting approach is to apply an EM algorithm to find the MAP that maximizes the likelihood of a measurement trace ([5], [15]). These EM based fitting methods are perhaps the best choices once there is a measurement trace available, and once the inherent computational effort is tolerable.

In several situations our aim is not to fit trace data, but to capture some statistical parameters of the arrival process with a MAP. This approach is typically used in order to approximate the behavior of a complex process (e.g. the departure process of a queue) by a much more tractable

MAP, or to construct a MAP based on a limited amount of information (e.g. we need a MAP with given "burstiness" and lag-1 autocorrelation, but nothing else is given). Our proposed MAP matching procedure belongs to this class of fitting methods as well.

Among the statistical parameter based fitting methods by far the most popular solutions intend to capture the marginal moments and the joint moments or the lag autocorrelation function. Some of these procedures intend to approximate them (like [6] and [7]), but a more difficult problem is capturing them in an exact way, thus performing MAP *matching*.

MAP matching methods can be classified further based on the structure they use. Procedures using a *rigid* structure keep the number of states fixed and can not adapt to the input parameters. Such methods are [3], [8] and [9], where 2-state MAPs are used to fit three marginal moments and the lag-1 autocorrelation. However, if these parameters do not fall into the bounds (which are provided in [3]), these method fails. The moment matching method in [17] operates on N states instead of only 2, but again, if the set of moments are not feasible with N states, there is no valid result (moreover, the bounds of feasible moments are not even known for $N > 2$). At the other hand, methods using a *flexible* structure are able to adopt to the input parameters, meaning that they can grow in size if necessary. Semi-flexible methods which are quite close to this paper in topic are described in [10] and [12], where two or three moments and the lag-1 autocorrelation are matched. There is some flexibility in these procedure in the sense that they build upon Erlang distributed components with the shape parameters providing a degree of freedom. The procedure of [10] is not able to realize all possible input parameter combinations, nor does it explicitly provide the bounds for it. In [12] the Erlang distributed components have a common shape parameter, which may lead to a large state space, furthermore there is no proof that the full range of the autocorrelation can be matched.

Our procedure presented in this paper uses a flexible structure. It is able to construct a MAP having *any* mean, squared coefficient of variation (SCV) and lag-1 autocorrelation parameters that a valid arrival process can have. Furthermore, we can match the third moment as well, except the case when the $SCV < 1$ and the correlation is negative. In this case we have (quite loose) constraints for the third moment, but in all other cases we can match the third moment without any restrictions as well. Additionally,

it is possible to set the decay of the autocorrelation function in a given range.

The rest of the paper is organized as follows. Section 2 re-introduces Markovian Arrival Processes with the notations used throughout the paper. The natural bounds of the parameters we intend to match are discussed in Section 3. Section 4 details the proposed matching procedure. The features of the procedure as well as some possible applications are demonstrated in Section 5. Finally, Section 6 concludes the paper.

2. MARKOVIAN ARRIVAL PROCESSES

A Markovian Arrival Process (MAP, [14]) of order N is given by two $N \times N$ matrices, $\boldsymbol{D_0}$ and $\boldsymbol{D_1}$. The sum of these matrices $\boldsymbol{D} = \boldsymbol{D_0} + \boldsymbol{D_1}$ is the generator of an irreducible continuous time Markov chain (CTMC) with N states, which is the background process of the MAP. Matrix $\boldsymbol{D_1}$ contains the rates of those state transitions which are accompanied by an arrival, and the off-diagonal entries of $\boldsymbol{D_0}$ are the rates of internal state transitions.

To express various properties of the MAP, we have to introduce a discrete time Markov chain (DTMC) representing the background process at arrival instants, given by transition probability matrix $\boldsymbol{P} = (-\boldsymbol{D_0})^{-1}\boldsymbol{D_1}$. The steady state probability vector of the background process embedded at arrival instants, denoted by π is the unique solution of $\pi\boldsymbol{P} = \pi, \pi\mathbb{1} = 1$, where $\mathbb{1}$ is the column vector of ones of appropriate size.

The steady state distribution of the inter-arrival times $\mathcal{X}$ is phase-type (PH) distributed with initial vector π and transient generator $\boldsymbol{D_0}$, thus

$$P(\mathcal{X} < t) = 1 - \pi e^{\boldsymbol{D_0} t}\mathbb{1}. \qquad (1)$$

The marginal moments of the inter-arrival times are then ([14])

$$m_k = E(\mathcal{X}^k) = k!\pi(-\boldsymbol{D_0})^{-k}\mathbb{1}. \qquad (2)$$

With the joint moments of the inter-arrival times it is possible to characterize the correlation structure of the MAP. By denoting the ℓth inter-arrival time by $\mathcal{X}_\ell$, the joint moments are calculated as

$$E(\mathcal{X}_0^i \mathcal{X}_\ell^j) = i!j!\pi(-\boldsymbol{D_0})^{-i}\boldsymbol{P}^\ell(-\boldsymbol{D_0})^{-j}\mathbb{1}. \qquad (3)$$

A popular correlation measure is the lag autocorrelation function (ACF), which provides the correlations between $\mathcal{X}_0$ and $\mathcal{X}_\ell$

$$\rho_\ell = \frac{E(\mathcal{X}_0 \mathcal{X}_\ell) - m_1^2}{m_2 - m_1^2}. \qquad (4)$$

3. NATURAL BOUNDS OF THE MOMENTS AND THE LAG-1 AUTOCORRELATION

The moments to match can not be arbitrary numbers. The Stieltjes moment problem ([16]) provides conditions to determine if the sequence $\{m_n, n = 0, 1, \dots\}$ corresponds to a distribution $[0, \infty)$. Let us define Hankel matrices $\boldsymbol{H_n^{(0)}}$

and $\boldsymbol{H_n^{(1)}}$ as

$$\boldsymbol{H_n^{(0)}} = \begin{bmatrix} m_0 & m_1 & m_2 & \dots & m_n \\ m_1 & m_2 & m_3 & \dots & m_{n+1} \\ m_2 & m_3 & m_4 & \dots & m_{n+2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ m_n & m_{n+1} & m_{n+2} & \dots & m_{2n} \end{bmatrix}, \qquad (5)$$

$$\boldsymbol{H_n^{(1)}} = \begin{bmatrix} m_1 & m_2 & m_3 & \dots & m_{n+1} \\ m_2 & m_3 & m_4 & \dots & m_{n+2} \\ m_3 & m_4 & m_5 & \dots & m_{n+3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ m_{n+1} & m_{n+2} & m_{n+3} & \dots & m_{2n+1} \end{bmatrix}. \qquad (6)$$

The condition for $\{m_n, n = 0, 1, \dots\}$ being a moment sequence on $[0, \infty)$ is that $\det(\boldsymbol{H_n^{(0)}}) > 0$ and $\det(\boldsymbol{H_n^{(1)}}) > 0$ for all n. (For simplicity, let us ignore the case when the determinants are zero).

From $\det(\boldsymbol{H_1^{(0)}}) > 0$ and $\det(\boldsymbol{H_1^{(1)}}) > 0$ we can immediately obtain conditions for the second and third moments yielding

$$m_2 > m_1^2, \qquad (7)$$
$$m_3 > m_2^2/m_1. \qquad (8)$$

Instead of using the raw moments, it will be beneficial to introduce second and third order moment expressions as

$$c_2 \triangleq \frac{m_2}{m_1^2} - 1, \qquad \ell_3 \triangleq \frac{m_3 m_1}{m_2^2} - 1. \qquad (9)$$

Note that c_2 is the squared coefficient of variation (SCV). For the moment expressions the bounds given by (7) and (8) translate to $c_2 > 0$ and $\ell_3 > 0$.

The lag correlations can not be arbitrary either. Due to the Cauchy-Schwartz inequality we have that $\rho_1 \in [-1, +1]$. However, since the inter-arrival times are non-negative, $E(\mathcal{X}_0 \mathcal{X}_1)$ is non-negative as well, consequently $\rho_1 \geq -1/c_2$ must hold, too (see (4)). We have that

$$\max\{-1, -1/c_2\} \geq \rho_1 \leq 1. \qquad (10)$$

4. THE PROPOSED MATCHING PROCEDURE

To solve the matching problem, we introduce a special MAP structure consisting of

- two PH distributions,

- and a 2-state discrete time Markov chain that determines which of the two components generates the next inter-arrival time.

Notice that similar structure has been proposed in [1] (called SMAP) and in [15] (called PH-CHMM) as well. The MAPs defined in [10] and [12] are similar in spirit as well.

The two PH distributions (referred to as "component" PH distributions) are given by initial vectors $\alpha^{(i)}$ and transient generators $\boldsymbol{A}^{(i)}$, $i = \{1, 2\}$; and their kth moments are denoted by $m_k^{(i)}, i = \{1, 2\}$. The squared coefficient of variations and the third order moment expressions of the component distributions are denoted by $c_2^{(i)}$ and $\ell_3^{(i)}$, $i = \{1, 2\}$.

The state transition probability matrix of the 2-state DTMC and its stationary distribution are denoted by $\boldsymbol{\Pi}$

and π, respectively. Denoting the transition probabilities by p_1 and p_2 we have

$$\mathbf{\Pi} = \begin{bmatrix} 1-p_1 & p_1 \\ p_2 & 1-p_2 \end{bmatrix}, \quad \pi = \begin{bmatrix} \frac{p_2}{p_1+p_2} & \frac{p_1}{p_1+p_2} \end{bmatrix}. \quad (11)$$

With these definitions and notations the $\boldsymbol{D_0}$ and $\boldsymbol{D_1}$ matrices of the MAP are

$$\boldsymbol{D_0} = \begin{bmatrix} \boldsymbol{A^{(1)}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A^{(2)}} \end{bmatrix},$$

$$\boldsymbol{D_1} = \begin{bmatrix} -\boldsymbol{A^{(1)}}\mathbb{1}\alpha^{(1)} \cdot (1-p_1) & -\boldsymbol{A^{(1)}}\mathbb{1}\alpha^{(2)} \cdot p_1 \\ -\boldsymbol{A^{(2)}}\mathbb{1}\alpha^{(1)} \cdot p_2 & -\boldsymbol{A^{(2)}}\mathbb{1}\alpha^{(2)} \cdot (1-p_2) \end{bmatrix}. \quad (12)$$

This MAP structure has the advantage that both the marginal moments and the lag autocorrelations can be obtained simpler then (2) and (4). Since the marginal distribution is the mixture of the two components, the marginal moments are

$$m_k = \frac{p_2}{p_1+p_2}m_k^{(1)} + \frac{p_1}{p_1+p_2}m_k^{(2)}. \quad (13)$$

For the lag-1 joint mean we have that

$$E(\mathcal{X}_0\mathcal{X}_k) = \begin{bmatrix} m_1^{(1)} & m_1^{(2)} \end{bmatrix} \cdot \text{diag}\langle\pi\rangle \cdot \mathbf{\Pi}^k \cdot \begin{bmatrix} m_1^{(1)} \\ m_1^{(2)} \end{bmatrix}. \quad (14)$$

It can be proven by induction that

$$\mathbf{\Pi}^k = \frac{1}{p_1+p_2}\begin{bmatrix} p_1(1-p_1-p_2)^k+p_2 & p_1-p_1(1-p_1-p_2)^k \\ (p_2-(1-p_1-p_2)^k p_2 & p_1+(1-p_1-p_2)^k p_2 \end{bmatrix}, \quad (15)$$

from which some algebraic manipulation gives

$$\rho_k = \frac{E(\mathcal{X}_0\mathcal{X}_k)/m_1^2 - 1}{c_2}$$

$$= \left(\frac{m_1^{(1)}}{m_1} - \frac{m_1^{(2)}}{m_1}\right)^2 \frac{p_1 p_2}{p_1+p_2}(1-p_1-p_2)^k \quad (16)$$

$$+ \left(\frac{m_1^{(1)}}{m_1}\frac{p_2}{p_1+p_2} + \frac{m_1^{(2)}}{m_1}\frac{p_1}{p_1+p_2}\right)^2.$$

Now we can formalize the matching problem to be solved in this paper. Our goal is to find

- the parameters of the component distributions $\alpha^{(i)}, \boldsymbol{A^{(i)}}, i = \{1,2\}$,

- and switching probabilities p_1 and p_2,

such that the resulting MAP matches m_1, c_2, ρ_1 and optionally ℓ_3 as well, without posing any restrictions on these parameters (apart from the natural bounds shown in Section 3).

We perform the MAP matching differently depending on c_2 and ρ_1. According to Figure 1 we can identify three regions. The case of positive correlation (Region 1 in the Figure) is discussed in Section 4.1, while Regions 2 and 3 are discussed in Section 4.2.

4.1 The case of positive auto-correlation

The matching procedure is discussed in an incremental way in this Section. First we determine the component means and the switching probabilities such that m_1 and ρ_1 are matched for $m_1 > 0$ and $\rho_1 \in [0,1]$. In the next step the SCV parameters of the components are determined such
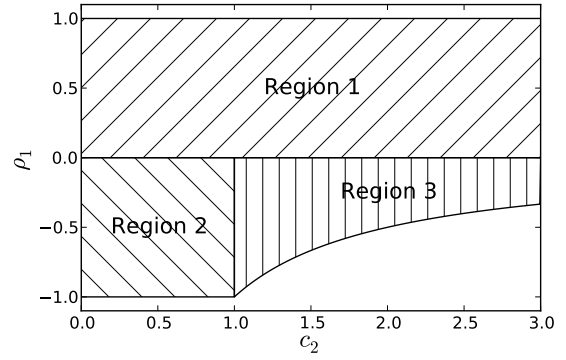
Figure 1: Regions on the c_2–ρ_1 plane needing different treatment

that c_2 is matched for $c_2 > 0$. We also provide an alternative method that matches not only c_2, but ℓ_3 as well. In case of positive ρ_1 our method is able to match any $\ell_3 > 0$.

4.1.1 Matching the mean and the auto-correlation

During the discussion of the procedure the key elements are stated as Theorems to put more emphasis on them.

The following Theorem provides the mean values of the components as well as the switching probabilities such that m_1 and ρ_1 are matched.

THEOREM 1. *The mean and the lag-k auto-correlations of the inter-arrival times of the MAP defined by component means*

$$m_1^{(1)} = m_1(1 - \sqrt{r}), \quad (17)$$

$$m_1^{(2)} = m_1(1 + c_2\sqrt{r}), \quad (18)$$

and switching probabilities

$$p_1 = \frac{1}{1+c_2}(1 - \rho_1/r), \quad (19)$$

$$p_2 = \frac{c_2}{1+c_2}(1 - \rho_1/r) \quad (20)$$

are given by m_1 and

$$\rho_k = r(\rho_1/r)^k, \quad (21)$$

for $\rho_1 < r < 1$.

PROOF. The Theorem can be proven by substituting $m_1^{(1)}, m_1^{(2)}, p_1$ and p_2 defined by (17), (18), (19) and (20) into (13) and (16).

The bounds for parameter r ensure that (17) and (18) are positive, and (19) and (20) are valid probabilities. □

The construction introduced by Theorem 1 has a free parameter, parameter r. The theorem ensures that the lag-1 autocorrelation is matched for $r \in (\rho_1, 1)$, thus r provides some additional flexibility to tune the decay of the autocorrelation function of the resulting MAP. There are several possibilities to set its value:

- Parameter r can be set such that ρ_2 is matched. Note however, that the bounds for r restrict the feasible lag-2 autocorrelation to $\rho_1^2 < \rho_2 < \rho_1$.

- Parameter r can be determined by optimization such that the (geometric) shape of the autocorrelation function is as close to the target as possible according to a distance function.

- If there is no information on the autocorrelation function apart from ρ_1, parameter r can be set arbitrarily. However, shifting r towards its bounds is not advisable, since at $r = 1$ we get $m_1^{(1)} = 0$, and at $r = \rho_1$ we have that $p_1 = p_2 = 0$. To obtain a representation with well-balanced entries in the $\boldsymbol{D_0}, \boldsymbol{D_1}$ matrices we recommend to set r such that ρ_2 falls in the middle of its feasible region, thus $\rho_2 = (\rho_1 + \rho_1^2)/2$, yielding

$$r = \frac{2\rho_1}{1 + \rho_1}. \tag{22}$$

4.1.2 Matching the second moment

In this Section we consider the case when we have only two marginal moments to match (m_1 and c_2), ℓ_3 is not available for some reason.

First we determine $c_2^{(1)}$ and $c_2^{(2)}$, the SCVs of the two components, such that the SCV of the mixture of the components is c_2. Then, we provide a simple procedure to construct PH distributions $(\alpha^{(i)}, \boldsymbol{A^{(i)}})$ based on $m_1^{(i)}$ and $c_2^{(i)}$, $i = \{1, 2\}$.

Let us start by expressing the second marginal moment with p_1 and p_2 given by Theorem 1 as

$$m_2 = \frac{c_2}{1 + c_2} m_2^{(1)} + \frac{1}{1 + c_2} m_2^{(2)},$$

that, divided by m_1^2, provides

$$\frac{m_2}{m_1^2} = \frac{c_2}{1 + c_2} \frac{m_2^{(1)}}{m_1^{(1)2}} \frac{m_1^{(1)2}}{m_1^2} + \frac{1}{1 + c_2} \frac{m_2^{(2)}}{m_1^{(2)2}} \frac{m_1^{(2)2}}{m_1^2}.$$

After substituting (17) and (18) we arrive to the following equation for the SCVs of the components.

$$
\begin{aligned}
c_2 + 1 = {} & \frac{c_2}{1 + c_2}(c_2^{(1)} + 1)(1 - \sqrt{r})^2 \\
& + \frac{1}{1 + c_2}(c_2^{(2)} + 1)(1 + c_2\sqrt{r})^2.
\end{aligned} \tag{23}
$$

Any $c_2^{(1)} > 0$ and $c_2^{(1)} > 0$ satisfying (23) can be chosen for the components. However, it is well known that the minimal SCV an order-n PH distribution can realize is $1/n$, thus choosing too low component SCVs can lead to overly large representations. The exact cost of given component SCVs in terms of number of states of the resulting MAP is

$$\mathcal{C}(c_2^{(1)}, c_2^{(2)}) = \max\left\{\left\lceil\frac{1}{c_2^{(1)}}\right\rceil, 2\right\} + \max\left\{\left\lceil\frac{1}{c_2^{(2)}}\right\rceil, 2\right\},$$

which, due to the ceiling function and the maximum, is difficult to handle analytically. However, the simplified cost function given by

$$\mathcal{C}'(c_2^{(1)}, c_2^{(2)}) = 1/c_2^{(1)} + 1/c_2^{(2)}, \tag{24}$$

which is clearly in a close relation with the size of the resulting MAP, allows us to express the optimal component SCVs explicitly.

THEOREM 2. *The component SCVs minimizing the cost function $1/c_2^{(1)} + 1/c_2^{(2)}$ are given by*

$$c_2^{(1)} = \frac{(1 + c_2)\sqrt{c_2}(1 + \sqrt{r})}{1 + \sqrt{c_2}(1 - \sqrt{r}) + c_2\sqrt{r}}, \tag{25}$$

$$c_2^{(2)} = \frac{c_2(1 + c_2)(1 - r)}{(1 + c_2\sqrt{r})(1 + \sqrt{c_2}(1 - \sqrt{r}) + c_2\sqrt{r})}. \tag{26}$$

PROOF. The theorem can be proven by using the Lagrange multiplier technique. Let us introduce the function

$$
\begin{aligned}
f(c_2^{(1)}, c_2^{(2)}, \lambda) = {} & \frac{1}{c_2^{(1)}} + \frac{1}{c_2^{(2)}} + \lambda\left(\frac{c_2}{1 + c_2}(c_2^{(1)} + 1)(1 - \sqrt{r})^2\right. \\
& \left. + \frac{1}{1 + c_2}(c_2^{(2)} + 1)(1 + c_2\sqrt{r})^2 - (c_2 + 1)\right),
\end{aligned}
$$

where the last term comes from the constraint (23).

The optimal solution is located at the point where the derivatives of $f(c_2^{(1)}, c_2^{(2)}, \lambda)$ taken according to all three variables are equal to zero. The equations to solve are

$$0 = -\frac{1}{c_2^{(1)2}} + \frac{c2(1 - \sqrt{r})^2\lambda}{1 + c2}, \tag{27}$$

$$0 = -\frac{1}{c_2^{(2)2}} + \frac{c2(1 + c_2\sqrt{r})^2\lambda}{1 + c2}, \tag{28}$$

$$0 = \frac{c_2(1 + c_2^{(1)})(1 - \sqrt{r})^2 + (1 + c_2^{(2)})(1 + c_2\sqrt{r})^2}{1 + c2} - 1 - c_2. \tag{29}$$

First we can express λ from (27), solve $c_2^{(2)}$ from (28) (care has to be taken to select the positive solution), finally, $c_2^{(1)}$ is provided by (29).

Note that both (25) and (26) are positive since $c_2 > 0$ and $r \in (0, 1)$. $\square$

The $\alpha^{(i)}, \boldsymbol{A^{(i)}}$ parameters of the component PH distributions are determined from $m_1^{(i)}$ and $c_2^{(i)}$ based on the next theorem.

THEOREM 3. *The mean and the SCV of the size N PH distribution given by $(\alpha, \boldsymbol{A})$ are $m_1 = 1/\lambda$ and c_2, where*

$$\alpha = \begin{bmatrix} p & 0 & \dots & 0 & 1 - p \end{bmatrix}, \tag{30}$$

$$\boldsymbol{A} = \begin{bmatrix} -N\lambda p & N\lambda p & & & \\ & -N\lambda p & N\lambda p & & \\ & & \ddots & \ddots & \\ & & & -N\lambda p & N\lambda p \\ & & & & -N\lambda \end{bmatrix} \tag{31}$$

with

$$p = \frac{1}{c_2 + 1 + (c_2 - 1)/(N - 1)} \tag{32}$$

and

$$N = \max\left\{\left\lceil\frac{1}{c_2}\right\rceil, 2\right\}. \tag{33}$$

PROOF. As there are only two states with non-zero initial probabilities, the mean value can be expressed by

$$
\begin{aligned}
m_1 &= p \cdot \left((N - 1)\frac{1}{N\lambda p} + \frac{1}{N\lambda}\right) + (1 - p) \cdot \frac{1}{N\lambda} \\
&= (N - 1)\frac{1}{N\lambda} + \frac{1}{N\lambda} = \frac{1}{\lambda}.
\end{aligned} \tag{34}
$$

For the second moment we have

$$m_2 = p \cdot \left[\left((N-1)\frac{1}{N^2\lambda^2 p^2} + \frac{1}{N^2\lambda^2} \right) \right.$$

$$+ \left. \left((N-1)\frac{1}{N\lambda p} + \frac{1}{N\lambda} \right)^2 \right] + (1-p) \cdot \frac{2}{N^2\lambda^2} \quad (35)$$

$$= \frac{N-1+2p}{Np\lambda^2} = \frac{1+c_2}{\lambda^2},$$

which, divided by m_1^2 and decreased by 1 gives c_2. In the last step of (35) we applied the definition of p given by (32). $\square$

4.1.3 Matching the third moment

In this Section we are matching m_1, c_2 and ℓ_3 as well.

The concept of the solution is the same as before, we are again looking for the appropriate component moments, that, with the given switching probabilities provide c_2 and ℓ_3.

For the component SCVs the constraint given by (23) must still be satisfied, but there is an additional constraint for $\ell_3^{(i)}$ as well. Applying a derivation similar to the one leading to (23) we have

$$(\ell_3+1)(c_2+1)^2 =$$
$$\frac{c_2}{1+c_2}(\ell_3^{(1)}+1)(c_2^{(1)}+1)^2(1-\sqrt{r})^3 \quad (36)$$
$$+ \frac{1}{1+c_2}(\ell_3^{(2)}+1)(c_2^{(2)}+1)^2(1+c_2\sqrt{r})^3.$$

Any $c_2^{(i)} > 0$ and $\ell_3^{(i)} > 0$ parameters can be chosen which satisfy (23) and (36), but we have to keep in mind that these parameters affect the size of the resulting MAP significantly. Unfortunately, the relation between ℓ_3 and the size of the minimal APH distribution realizing it is quite complex (see [2]), which makes the selection of the optimal moments for the components difficult. To simplify the analysis, we pick one particular bound: the one that says that the minimal ℓ_3 parameter that can be realized by an order-n APH is $1/n$ (case b. of Theorem 3.1 in [2]), and perform the optimization of $\ell_3^{(i)}$ parameters according to this rule. This leads to cost function (simplified as in the previous section)

$$\mathcal{C}'(c_2^{(1)}, \ell_3^{(1)}, c_2^{(2)}, \ell_3^{(2)}) = \min\{\frac{1}{c_2^{(1)}}, \frac{1}{\ell_3^{(1)}}\} + \min\{\frac{1}{c_2^{(2)}}, \frac{1}{\ell_3^{(2)}}\}.$$

Unfortunately we were not able to solve this optimization problem such a straight forward way as we did in Section 4.1.2, we found that it does not even always have a unique solution. What we could do is that we determined the optimal $\ell_3^{(i)}$ parameters given that the $c_2^{(i)}$ parameters are fixed.

THEOREM 4. *The $\ell_3^{(i)}$ parameters of the components minimizing the cost function $1/\ell_3^{(1)} + 1/\ell_3^{(2)}$ are given by*

$$\ell_3^{(1)} = a/b, \quad (37)$$

$$\ell_3^{(2)} = \ell_3^{(1)} \frac{c_2^{(1)}+1}{c_2^{(2)}+1} \frac{1-\sqrt{r}}{1+c_2\sqrt{r}} \sqrt{\frac{c_2(1-\sqrt{r})}{1+c_2\sqrt{r}}}. \quad (38)$$

with

$$a = (\ell_3+1)(c_2+1)^3 - c_2(c_2^{(1)}+1)^2(1-\sqrt{r})^3$$
$$- (c_2^{(2)}+1)^2(1+c_2\sqrt{r})^3,$$

$$b = (c_2^{(1)}+1)(1-\sqrt{r})\left(c_2(c_2^{(1)}+1)(1-\sqrt{r})^2 \right.$$
$$+ \left. (c_2^{(2)}+1)(1+c_2\sqrt{r})^2 \sqrt{\frac{c_2(1-r)}{1+c_2\sqrt{r}}} \right),$$

PROOF. The theorem can be proven by the same way as Theorem 2. $\square$

We were not able to find component SCVs $c_2^{(i)}$ minimizing the size of the MAP representation. However, we were able to find component SCVs such that the $c_2^{(i)}$ and $\ell_3^{(i)}$ are always feasible. Thus, by setting the component moments according to the next theorem we always get a valid MAP matching m_1, c_2, ℓ_3 and ρ_1.

THEOREM 5. *When the SCVs of the two components are*

$$c_2^{(1)} = \frac{c_2 + \sqrt{r}}{1 - \sqrt{r}}, \quad (39)$$

$$c_2^{(2)} = c_2 \frac{1 - \sqrt{r}}{1 + c_2\sqrt{r}}, \quad (40)$$

the expressions for the cost-optimal $\ell_3^{(1)}$ and $\ell_3^{(2)}$ simplify to

$$\ell_3^{(1)} = \frac{l_3(1+c_2)}{c_2(1-\sqrt{r}) + \sqrt{c_2(1+c_2\sqrt{r})(1-\sqrt{r})}}, \quad (41)$$

$$\ell_3^{(2)} = \frac{l_3(1+c_2)}{1 + c_2\sqrt{r} + \sqrt{c_2(1+c_2\sqrt{r})(1-\sqrt{r})}}. \quad (42)$$

Furthermore, we have that $c_2^{(1)} > 0, c_2^{(2)} > 0, \ell_3^{(1)} > 0$ and $\ell_3^{(2)} > 0$ always hold.

PROOF. Equations (41) and (42) are obtained by substituting (39) and (40) into (37) and (38).

The positivity of $c_2^{(i)}$ and $\ell_3^{(i)}$ follows from $c_2 > 0, \ell_3 > 0$ and $r \in (0, 1)$. $\square$

At this point we have three (feasible) marginal moments. Based on these moments the appropriate $(\alpha^{(i)}, \boldsymbol{A^{(i)}})$ representations of the component PH distributions are obtained by applying the results of [2].

4.2 The case of negative auto-correlation

4.2.1 Matching the mean and the auto-correlation

The MAP structure introduced in Section 4.1.1 in not always able to match negative lag-1 autocorrelation. Thus, for $\rho < 0$ (Region 2 and Region 3 in Figure 1) we introduce a slightly different (and actually simpler) MAP structure.

THEOREM 6. *The mean and the lag-k auto-correlations of the inter-arrival times of the MAP defined by component means*

$$m_1^{(1)} = m_1(1 - \sqrt{c_2 r}), \quad (43)$$

$$m_1^{(2)} = m_1(1 + \sqrt{c_2 r}), \quad (44)$$

and switching probabilities

$$p_1 = \frac{1}{2}(1 - \rho_1/r), \tag{45}$$

$$p_2 = \frac{1}{2}(1 - \rho_1/r) \tag{46}$$

are given by m_1 and

$$\rho_k = r(\rho_1/r)^k, \tag{47}$$

for

$$-\rho_1 < r < \min\{1, 1/c_2\}. \tag{48}$$

PROOF. The theorem can proven the same way as Theorem 1.

Condition $r > -\rho_1$ is necessary for $p_1 < 1, p_2 < 1$, and $r < 1/c_2$ is required to get $m_1^{(1)} > 0$. However, showing why the remaining condition $r < 1$ is necessary is a bit more involved.

Let us express the SCV of the MAP with the component SCVs. First, the second moment is

$$m_2 = \frac{1}{2}m_2^{(1)} + \frac{1}{2}m_2^{(2)},$$

which, divided by m_1^2, yields

$$\begin{aligned}\frac{m_2}{m_1^2} = c_2 + 1 &= \frac{1}{2}\frac{m_2^{(1)}}{m_1^{(1)2}}\frac{m_1^{(1)2}}{m_1^2} + \frac{1}{2}\frac{m_2^{(2)}}{m_1^{(2)2}}\frac{m_1^{(2)2}}{m_1^2}\\ &= \frac{1}{2}(c_2^{(1)}+1)(1-\sqrt{c_2r})^2 + \frac{1}{2}(c_2^{(2)}+1)(1+\sqrt{c_2r})^2.\end{aligned} \tag{49}$$

The right hand side of (49) is minimized at $c_2^{(1)} = c_2^{(2)} = 0$, thus

$$c_2 + 1 > \frac{1}{2}(1-\sqrt{c_2r})^2 + \frac{1}{2}(1+\sqrt{c_2r})^2 = c_2r + 1, \tag{50}$$

from which $r < 1$ follows. $\square$

As in case of positive autocorrelation, parameter r can be used to adjust the decay of the autocorrelation function. The feasible range of lag-2 autocorrelation, however, depends on the SCV in this case.

- If $c_2 \leq 1$ (Region 2 in Figure 1), we have that $\rho_1^2 < \rho_2 < -\rho_1$ (from $-\rho < r < 1$).

- If $c_2 > 1$ (Region 3 in Figure 1), we get $\rho_1^2 c_2 < \rho_2 < -\rho_1$ (as $-\rho < r < 1/c_2$ in this case).

The recommended "default" value for parameter r is such that ρ_2 is set to the middle of its feasible region (by similar considerations as in Section 4.1.1), thus our suggestion is

$$r = \begin{cases} -\dfrac{2\rho_1}{1-\rho_1} & \text{if } c_2 \leq 1, \\ -\dfrac{2\rho_1}{1-\rho_1 c_2} & \text{if } c_2 > 1. \end{cases} \tag{51}$$

4.2.2 Matching the second moment

We are again following the approach taken in the case of positive lag-1 autocorrelation in Section 4.1.2.

The condition for the component SCVs is already derived in (49), that is

$$c_2 + 1 = \frac{1}{2}(c_2^{(1)}+1)(1-\sqrt{c_2r})^2 + \frac{1}{2}(c_2^{(2)}+1)(1+\sqrt{c_2r})^2. \tag{52}$$

The next theorem provides the component SCVs that are optimal with regards to the simplified cost function defined by (24).

THEOREM 7. *The component SCVs minimizing the cost function $1/c_2^{(1)} + 1/c_2^{(2)}$ are given by*

$$c_2^{(1)} = c_2\frac{1-r}{1-\sqrt{c_2r}}, \tag{53}$$

$$c_2^{(2)} = c_2\frac{1-r}{1+\sqrt{c_2r}}. \tag{54}$$

PROOF. The results can be proven by the Lagrange multiplier technique as in case of Theorem 2.

Note that both (53) and (54) are positive since $c_2 > 0$ and $c_2r \in (0,1)$ holds both in Region 2 and Region 3. $\square$

The $(\alpha^{(i)}, \boldsymbol{A^{(i)}})$ representations of the component PH distributions are obtained by using Theorem 3.

4.2.3 Matching the third moment

Along the same line as Section 4.1.3, we can derive the constraint for $\ell_3^{(i)}$ parameters as

$$\begin{aligned}(\ell_3+1)(c_2+1)^2 = &\frac{1}{2}(\ell_3^{(1)}+1)(c_2^{(1)}+1)^2(1-\sqrt{c_2r})^3\\ &+ \frac{1}{2}(\ell_3^{(2)}+1)(c_2^{(2)}+1)^2(1+\sqrt{c_2r})^3,\end{aligned} \tag{55}$$

and arrive to a theorem similar to Theorem 4:

THEOREM 8. *If $\rho_1 < 0$, the $\ell_3^{(i)}$ parameters of the components minimizing the cost function $1/\ell_3^{(1)} + 1/\ell_3^{(2)}$ are given by*

$$\ell_3^{(1)} = a/b, \tag{56}$$

$$\ell_3^{(2)} = \ell_3^{(1)}\frac{c_2^{(1)}+1}{c_2^{(2)}+1}\left(\frac{1-\sqrt{c_2r}}{1+\sqrt{c_2r}}\right)^{\frac{3}{2}} \tag{57}$$

with

$$\begin{aligned}a = &2(\ell_3+1)(c_2+1)^2 - (c_2^{(1)}+1)^2(1-\sqrt{c_2r})^3\\ &- (c_2^{(2)}+1)^2(1+\sqrt{c_2r})^3\\ b = &(c_2^{(1)}+1)^2(1-\sqrt{c_2r})^3\\ &+ (c_2^{(1)}+1)(c_2^{(2)}+1)(1+\sqrt{c_2r})^3\left(\frac{1-\sqrt{c_2r}}{1+\sqrt{c_2r}}\right)^{3/2}\end{aligned}$$

The only remaining question is how to set the component SCVs such that $c_2^{(i)}$ and $\ell_3^{(i)}$ are all positive, yielding valid component PH distributions. Unfortunately we were not able to find such a solution. The solution we found (given by the next theorem) is able to satisfy these requirements for Region 3, but not always in Region 2. Extensive numerical experiments make us believe that this is a structural restriction. According to us, no MAP exists that has a structure of (12) and is able to implement both near-zero c_2, near-zero ℓ_3 and close to -1 lag-1 autocorrelation.

THEOREM 9. *When the SCVs of the two components are*

$$c_2^{(1)} = \frac{c_2 + \sqrt{c_2r}}{1 - \sqrt{c_2r}}, \tag{58}$$

$$c_2^{(2)} = \frac{c_2 - \sqrt{c_2r}}{1 + \sqrt{c_2r}}, \tag{59}$$

the expressions for the cost-optimal $\ell_3^{(1)}$ and $\ell_3^{(2)}$ simplify to

$$\ell_3^{(1)} = \frac{2\ell_3}{1 - \sqrt{c_2 r} + \sqrt{1 - c_2 r}}, \tag{60}$$

$$\ell_3^{(2)} = \frac{2\ell_3}{1 + \sqrt{c_2 r} + \sqrt{1 - c_2 r}}. \tag{61}$$

Furthermore, we have that $c_2^{(1)} > 0, c_2^{(2)} > 0, \ell_3^{(1)} > 0$ and $\ell_3^{(2)} > 0$ holds if $r < c_2$.

PROOF. Equations (60) and (61) are obtained by substituting (58) and (59) into (56) and (57).

Since $c_2 > 0, \ell_3 > 0, r < 1/c_2$ and $r > 0$, all four expressions are positive except $c_2^{(2)}$, which is positive only if the condition $r < c_2$ holds. $\square$

Note that the condition of the Theorem $r < c_2$ always holds in Region 3, but depending on the lag-1 autocorrelation and the SCV it may not always hold in Region 2. Thus, for $c_2 \geq 1$ we are able to match any c_2, ℓ_3, ρ_1 combinations, while for $c_2 < 1$ we might not able to do that.

4.3 Summary of the matching procedure

The pseudo-code in Figure 2 gives an overview of the presented procedure. The main decision points of the method are as follows.

- We are using different structures for positive and negative ρ_1 values, thus the procedure checks ρ_1 right at the beginning.

- If we do not want to set parameter r (because of the lack of knowledge of the ACF), a default value has to be assigned.

- Different methods are used to obtain the component distributions when two (m_1, c_2) or three (m_1, c_2, ℓ_3) moments are matched.

5. NUMERICAL EXAMPLES

Our implementation of the matching procedure is available online[1] for Mathematica, Matlab and Python environments.

5.1 Size of the resulting MAP representation

In the first experiment we investigate how the size of the final MAP representation depends on the input parameters. The results are depicted in Figure 3. The different shades of grays represent the size of the MAP, with the lightest being the smallest one (with 4 states), and the darker shade correspond to the larger MAP size (the completely black means that the result is larger than 50 states). The dashed line indicates the region that can be covered by the MAP-2 as published in [3].

It is clearly visible in the Figure that our procedure was able to cover the entire c_2–ρ_1 plane (with respect to the natural bounds). The closer c_2 is to zero and the larger $|\rho_1|$ is, the more states are required to match them (up to infinity at the bounds).

In the second scenario the ℓ_3 parameter is also involved into the matching. We fixed the lag-1 autocorrelation to $\rho_1 = 0.2$, and recorded the size of the resulting MAP by

```
1:  procedure MAPMATCH(m_1, c_2, ℓ_3, ρ_1, r)
2:      if ρ_1 ≥ 0 then
3:          if r is given and r ≤ 0 or r ≥ 1 then
4:              error "Parameter r is out of range!"
5:          else if r is not given then
6:              r ← (2ρ_1)/(1+ρ_1)
7:          end if
8:          Set m_1^(1), m_1^(2) according to (17) and (18)
9:          Set p_1, p_2 according to (19) and (20)
10:         if ℓ_3 is not given then
11:             Calculate c_2^(1) and c_2^(2) by (25) and (26)
12:             Obtain α^(1), A^(1), α^(2), A^(2) by Theorem 3
13:         else
14:             Calculate c_2^(1) and c_2^(2) by (39) and (40)
15:             Calculate ℓ_3^(1) and ℓ_3^(2) by (41) and (42)
16:             Obtain α^(1), A^(1), α^(2), A^(2) by [2]
17:         end if
18:     else
19:         if r is given and r ≤ 0 or r ≥ min{1, 1/c_2} then
20:             error "Parameter r is out of range!"
21:         else if r is not given then
22:             if c_2 ≤ 1 then
23:                 r ← -(2ρ_1)/(1-ρ_1)
24:             else
25:                 r ← -(2ρ_1)/(1-ρ_1 c_2)
26:             end if
27:         end if
28:         Set m_1^(1), m_1^(2) according to (43) and (44)
29:         Set p_1, p_2 according to (45) and (46)
30:         if ℓ_3 is not given then
31:             Calculate c_2^(1) and c_2^(2) by (53) and (54)
32:             Obtain α^(1), A^(1), α^(2), A^(2) by Theorem 3
33:         else if r ≥ c_2 then
34:             error "Can not match ℓ_3 by the given c_2, ρ_1!"
35:         else
36:             Calculate c_2^(1) and c_2^(2) by (58) and (59)
37:             Calculate ℓ_3^(1) and ℓ_3^(2) by (60) and (61)
38:             Obtain α^(1), A^(1), α^(2), A^(2) by [2]
39:         end if
40:     end if
41:     Construct matrices D_0 and D_1 according to (12)
42:     return D_0, D_1
43: end procedure
```

Figure 2: The matching procedure

different c_2 and ℓ_3 parameters. The conclusion is the same as in the previous case (see Figure 4). The MAP-2 is able to match only a limited range of these parameters (striped area in the Figure), while our procedure is flexible and successfully matches any c_2–ℓ_3 combinations by increasing the size of the MAP. Again, the closer c_2 and ℓ_3 are to 0, the more states are needed to match them.

The situation is similar in case of negative correlation (Figure 5), In this case, the region feasible with MAP-2 is even smaller. As mentioned before, our procedure can not match all c_2, ℓ_3 parameters. As also visible in the Figure, only $c_2 > 1/3$ is covered (remember that this value depends on ρ_1), but if it holds, the ℓ_3 parameter can be arbitrary.
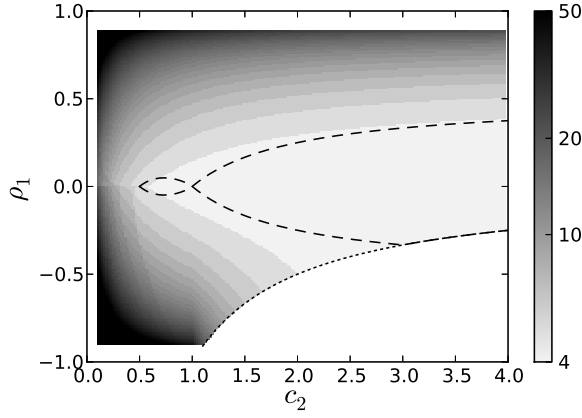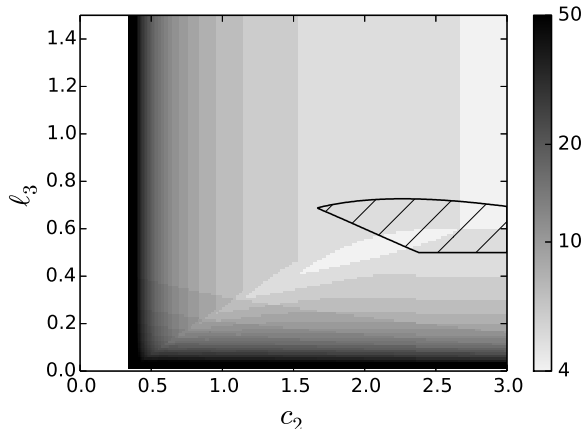
Figure 3: The size of the MAP as a function of the SCV and the lag-1 autocorrelation
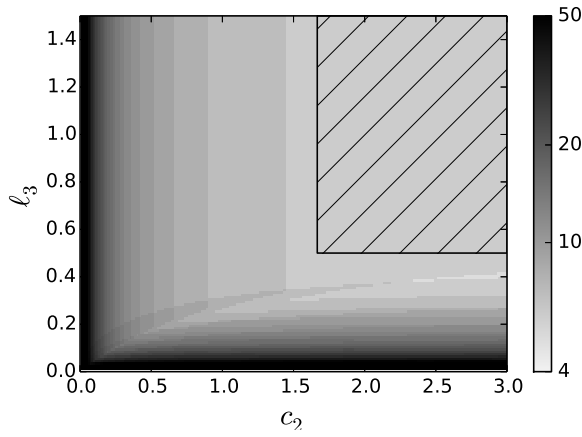
Figure 4: The size of the MAP as a function c_2 and ℓ_3 at $\rho = 0.2$

Figure 5: The size of the MAP as a function c_2 and ℓ_3 at $\rho = -0.2$

5.2 Application for fitting traffic traces

Although there are methods which are much better suited for trace fitting, we can apply the proposed matching procedure to real traffic traces as well. In this example we have chosen a traffic trace which is frequently used as a bench-

mark, the BC-pAug89 trace[2].

We investigate the following three scenarios depending on which statistics are involved in the matching and how.

- Only three parameters, m_1, c_2 and ρ_1 are matched according to the proposed procedure.

- Five parameters, m_1, c_2, ℓ_3, ρ_1 and ρ_2 are matched according to the proposed procedure.

- In this scenario we are trying to capture the characteristics of the trace with our MAP structure as much as possible. In the first step we are looking for the optimal ρ_1 and r parameters such that the Kullback–Leibler distance (also called as relative entropy) of the geometric ACF given by (21) and the ACF of the trace are minimized. Then, we are looking for the optimal $c_2^{(i)}, \ell_3^{(i)}$ parameters for the marginal distribution (respecting the constraints of (23) and (36)) such that the Kullback–Leibler distance between the resulting and the empirical density functions are minimized. This means that two separate nonlinear optimization is required in this case, both of them involves only two variables. On our average computer both terminates in 1-2 seconds.

The results are depicted in Figures 6 and 7. As expected, the worst results are obtained in Case 1, where only 3 parameters are matched. Matching 5 parameters improves the approximation of the density function and the ACF as well, the improvement in the ACF is substantial. The best approximation is achieved by the third scenario, which is not surprising, as it utilizes the most information from the trace.
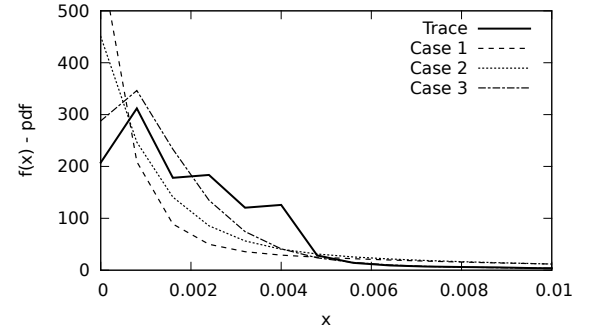
Figure 6: Marginal density functions of the 3 cases

We emphasize again that the main purpose of the presented procedure is to match a few statistical parameters, which is not always enough to capture the dominant characteristics of a measurement trace. There are several existing procedures that have been developed specifically for trace fitting.

5.3 Sensitivity analysis of the M/MAP/1 queue

This section demonstrates an other possible application area of the presented procedure. It is well known, that the mean waiting time of the M/GI/1 queue depends only on the

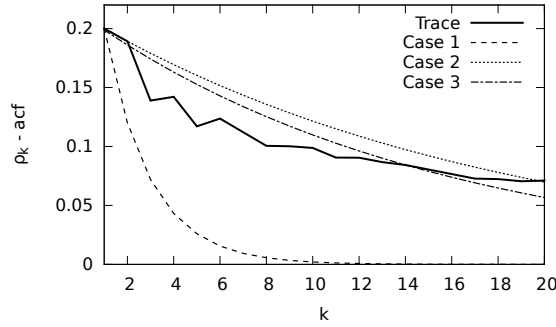**Figure 7: Autocorrelation functions of the 3 cases**

Figure 9: Mean waiting time in case of lower load

mean and the SCV of the inter-arrival times (as stated by the Pollaczek-Khintchine formula, [4]). We consider, however, the case when the service times are correlated, and investigate how the mean waiting time of a M/MAP/1 queue depends on the c_2 ("burstiness") and on the ρ_1 ("correlation") parameters of the service times.

First of all we checked whether the mean waiting time is still insensitive to ℓ_3 when $\rho_1 \neq 0$. We found that the insensitivity does not hold any more when the service times are correlated, however, the impact of ℓ_3 on the mean waiting time is negligible. Hence, we are matching only m_1, c_2 and ρ_1 in the sequel.

Figure 8 and Figure 9 depicts the mean waiting time as the function of the lag-1 autocorrelation by higher and lower load, respectively. According to the results the higher the correlation is, the higher is the mean waiting time. Interestingly, the mean waiting time with negatively correlated service times is lower than in the uncorrelated case. Furthermore, the higher the SCV of the service time is, the more sensitive the waiting time is to the lag-1 autocorrelation.
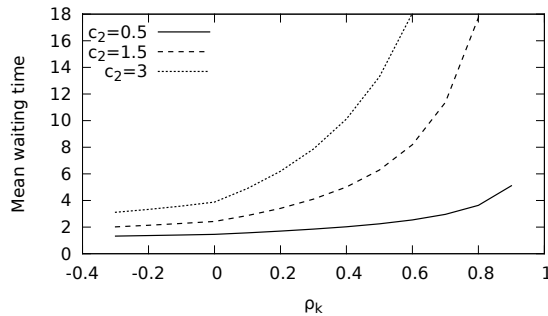

Figure 8: Mean waiting time in case of higher load

6. CONCLUSION

The proposed procedure is able to construct a Markovian arrival process based on the mean inter-arrival times (m_1), the "burstiness" (SCV,c_2) and the "correlation" (ρ_1) parameters. The resulting MAP is always valid for all $c_2 > 0$ and $\rho_1 \in (-1, +1)$. The applied structure has some degrees of freedom left that can be used to adjust the decay of the autocorrelation function and to match the third moment as well. Apart from the case when $c_2 < 1, \rho_1 < 0$, the third
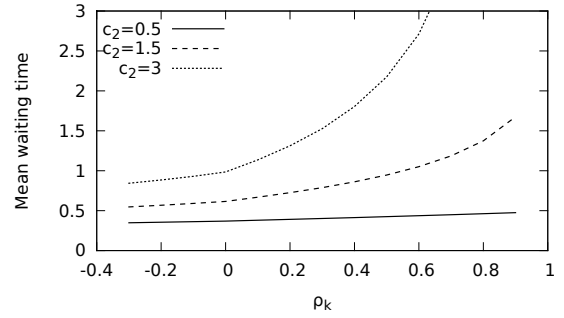
moment can be arbitrary as well.

The application fields of the procedure include different fitting problems when only a low number of parameters are available, or sensitivity analysis when we are interested in how the system reacts on different burstiness and correlation parameters.

Acknowledgment

7. REFERENCES

[1] F. Bause and G. Horváth. Fitting Markovian Arrival Processes by Incorporating Correlation into Phase Type Renewal Processes. In *Quantitative Evaluation of Systems (QEST), 2010 Seventh International Conference on the*, pages 97–106. IEEE, 2010.

[2] A. Bobbio, A. Horváth, and M. Telek. Matching three moments with minimal acyclic phase type distributions. *Stochastic models*, 21(2-3):303–326, 2005.

[3] L. Bodrog, A. Heindl, G. Horváth, and M. Telek. A Markovian canonical form of second-order matrix-exponential processes. *European Journal of Operational Research*, 190(2):459–477, 2008.

[4] G. Bolch, S. Greiner, H. de Meer, and K. S. Trivedi. *Queueing networks and Markov chains: modeling and performance evaluation with computer science applications*. John Wiley & Sons, 2006.

[5] P. Buchholz. An EM-algorithm for MAP fitting from real traffic data. In *Computer Performance Evaluation. Modelling Techniques and Tools*, pages 218–236. Springer, 2003.

[6] P. Buchholz, P. Kemper, and J. Kriege. Multi-class Markovian arrival processes and their parameter fitting. *Performance Evaluation*, 67(11):1092–1106, 2010.

[7] G. Casale, E. Z. Zhang, and E. Smirni. KPC-toolbox: Simple yet effective trace fitting using Markovian arrival processes. In *Quantitative Evaluation of Systems, 2008. QEST'08. Fifth International Conference on*, pages 83–92. IEEE, 2008.

[8] J. Diamond and A. Alfa. On approximating higher order MAPs with MAPs of order two. *Queueing Systems*, 34(1-4):269–288, 2000.

[9] H.-W. Ferng and J.-F. Chang. Connection-wise end-to-end performance analysis of queueing networks with MMPP inputs. *Performance Evaluation*, 43(1):39–62, 2001.

[10] M. K. Girish and J.-Q. Hu. Modeling of correlated arrival processes in the Internet. In *Decision and Control, 1999. Proceedings of the 38th IEEE Conference on*, volume 5, pages 4454–4459. IEEE, 1999.

[11] G. Horváth and M. Telek. Acceptance-rejection methods for generating random variates from matrix exponential distributions and rational arrival processes. In *Matrix-Analytic Methods in Stochastic Models*, pages 123–143. Springer, 2013.

[12] M. A. Johnson. Markov MECO: a simple Markovian model for approximating nonrenewal arrival processes. *Stochastic Models*, 14(1-2):419–442, 1998.

[13] J. Kriege and P. Buchholz. Simulating stochastic processes with omnet++. In *Proceedings of the 4th International ICST Conference on Simulation Tools and Techniques*, pages 367–374. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011.

[14] G. Latouche and V. Ramaswami. *Introduction to matrix analytic methods in stochastic modeling*, volume 5. Society for Industrial and Applied Mathematics, 1987.

[15] H. Okamura and T. Dohi. Faster maximum likelihood estimation algorithms for Markovian arrival processes. In *Quantitative Evaluation of Systems, 2009. QEST'09. Sixth International Conference on the*, pages 73–82. IEEE, 2009.

[16] T. J. Stieltjes and G. van Dijk. *Œuvres complètes: Collected papers*. Springer-Verlag, 1993.

[17] M. Telek and G. Horváth. A minimal representation of Markov arrival processes and a moments matching method. *Performance Evaluation*, 64(9):1153–1168, 2007.