A Two-Step Fitting Approach of Batch Markovian Arrival Processes for Teletraffic Data

Gang Chen[1], Li Xia[1(✉)], Zhaoyu Jiang[2], Xi Peng[3], Li Chen[3], and Bo Bai[3]

[1] Business School, Sun Yat-Sen University, Guangzhou 510275, China
{chengang5,xiali5}@sysu.edu.cn, xial@tsinghua.edu.cn

[2] Huawei Tech. Co., Ltd., No.156 Beiqing Rd., Hai-Dian District, Beijing, China
jiangzhaoyu2@huawei.com

[3] Huawei Tech. Investment Co., Ltd., Pak Shek Kok, Shatin, N.T., Hong Kong, China
{pancy.pengxi,chen.li7}@huawei.com

Abstract. Batch Markovian arrival process (BMAP) is a powerful stochastic process model for fitting teletraffic data since its MAP structure can capture the mode of packet interarrival times and its batch structure can capture packet size distributions. Compared with Poisson-related models and simple MAP models richly studied in the literature, the BMAP model and its fitting approach are much less investigated. Motivated by a practical project collaborated with Huawei company, we propose a new two-step parameter fitting approach of the BMAP model for teletraffic data generated from IP networks. The first step is the phase-type fitting for packet interarrival times, which is implemented by the framework of EM (expectation maximization) algorithms. The second step is the approximation of the lag correlation values of packet interarrival times and packet sizes, which is implemented by the framework of MM (moment matching) algorithms. The performance of our two-step EM-MM fitting approach is demonstrated by numerical experiments on both simulated and real teletraffic data sets, and compared with the MAP and MMPP (Markovian modulated Poisson process) models to illustrate the advantages of the BMAP model. Numerical examples also show that our proposed two-step fitting approach can obtain a good balance between the computation efficiency and accuracy.

Keywords: Batch Markovian arrival process · Traffic modeling · Fitting approach · Expectation maximization · Moment matching

Supported by the Guangdong Basic and Applied Basic Research Foundation (2021A1515011984, 2020A1515110824), the National Natural Science Foundation of China (62073346, 11671404, 61573206), and a collaborated project between the Sun Yat-Sen University and the Huawei Company.

Q. Zhao and L. Xia (Eds.): VALUETOOLS 2021, LNICST 404, pp. 22–35, 2021.
https://doi.org/10.1007/978-3-030-92511-6_2

1 Introduction

Along with the widely deployed infrastructure of the Internet, Internet protocol (IP) traffic has been becoming the main source of teletraffic data. It is important to study the statistical properties of IP traffic, such as the correlation, self-similarity, and burstyness. For IP traffic, the dilemma of the traffic modeling is to capture these statistical properties of the underlying measured trace data. A variety of traffic models have been developed in the past two decades. Most of the existing traffic models focus on two research directions. One direction is the consideration of non-analytically tractable models, e.g., fractional Gaussian noise model [15]. The other direction is the study of analytically tractable models, e.g., MMPP (Markovian modulated Poisson process) [9] and MAP (Markovian arrival process) [4].

As analytical models, the MAP and MMPP are widely used for traffic modeling, and they are easy to integrate in simulation models. A myriad of useful results has been achieved on these Markovian models for fitting the IP traffic. Some effective and stable fitting approaches of MAP and MMPP for traffic modeling are developed in the literature. These fitting approaches are mainly divided into moments matching (MM) [4,10,11] and expectation maximization (EM) [3,6,18]. Most of these available approaches fit the parameters of MAP or MMPP for the characterization of packet arrivals. Actually, packet arrivals and packet sizes are both observed in the measured traffic data. As we know, the characteristics of packet sizes also paly an important role in the performance analysis of the real traffic data. While, the MAP and MMPP are only used to model the packet arrivals of the traffic flow. As an extensive process of MAP, the stationary BMAP (batch Markovian arrival process) is capable of approximating any stationary batch point process. Moreover, one addition advantage of the BMAP model provides a more comprehensive tool for capturing the packet size distribution of IP traffic, while it still remains analytically tractable. The general BMAP is a highly parameterized model. So far, there have been only a limited number of studies on BMAP fitting. Breuer [1] studied a parameter estimation approach for a class of BMAP models. Then, he proposed an EM algorithm for the BMAP and its comparison to a simpler estimation procedure [2]. Klemm et al. [13] developed an efficient and numerically stable method for estimating the parameters of BMAP with an EM algorithm. Salvador et al. [21] introduced a fitting procedure for discrete-time BMAP which allows a simultaneous matching of joint characterization of packet arrival times and packet sizes. To our knowledge, only these above papers studied the parameter estimation of BMAP models. They all showed that BMAP is analytically tractable and it closely captures the statistics of the measured traffic data.

In this paper, we show how to utilize the EM and MM algorithms for the parameter estimation of BMAPs, by using a two-step fitting procedure. In the first step, the packet interarrival time is fitted by a phase-type distribution which corresponds to the embedded arrival process of BMAP. In the second step, the approximation of the lag correlation values of packet arrivals and packet sizes is computed through a non-linear optimization problem. As we know, the EM algorithm is accurate for Markovian models fitting of teletraffic data, while its computational cost is relatively high. A challenging issue for the teletraffic fitting, especially in practice, is to identify methods to reduce the computational cost. Our approach integrates the advantages of the EM and MM algorithms during the aforementioned two-step fitting procedure. Based on our proposed algorithm, we conduct some numerical experiments which illustrate the accuracy of our fitting method on real traffic data. Furthermore, in order to show the advantage of BMAP traffic model over other widely used analytically tractable models, we compare the BMAP with the MAP and MMPP models by means of visual inspection of sample paths from a real trace data. We demonstrate that BMAP model can capture the important statistical properties of packet arrivals and packet sizes. The numerical results also show that our two-step fitting approach can obtain a good balance between the computation efficiency and accuracy, which is an important feature for implementation in practice.

Our main contributions are summarized as follows. First, we introduce a traffic model and a fitting procedure that provide a detailed characterization of packet arrivals and packet sizes. Second, we show that the proposed traffic model BMAP and its parameters estimation algorithm are capable of matching closely the simulated and real traffic traces. Finally, we make a comparison study with the MAP, MMPP models fitted by the existing EM algorithm. The numerical results show that our proposed two-step EM-MM algorithm is both accurate and numerically efficient. The paper is organized as follows. The mathematical model and some properties of BMAP are introduced in Sect. 2. In Sect. 3, we analyze the implementation details of our proposed two-step fitting procedure. A number of numerical examples are conducted in Sect. 4, which demonstrate the efficiency of our proposed BMAP fitting approach. Finally, we draw concluding remarks with some future discussions in Sect. 5.

2 Basic Properties of BMAP

BMAP was first introduced by Neuts [17]. Its current and more tractable description can refer to the work by Lucantoni [16]. In this section, we firstly present a brief introduction of the BMAP model. The moments and autocorrelations characterization for the BMAP model are then provided.

2.1 Description of BMAP

We denote an m-state BMAP as symbol $BMAP_m(K)$, where K is the maximum batch size. With the model of $BMAP_m(K)$, a doubly stochastic process can be identified as $X(t) = \{J(t), N(t)\}$, where $J(t)$ represents an irreducible Markov process with finite state space $\mathcal{S} = \{1, 2, ..., m\}$, and $N(t)$ denotes the total number of arrivals up to time t. Moreover, $J(t)$ is normally called the *phase process* and $N(t)$ is the *counting process*.

For each state $i \in \mathcal{S}$, the process $J(t)$ spends an exponentially distributed amount of time in state i with rate λ_i. The transition that follows this sojourn can be one of two following types. For the first type, an arrival of batch size k ($k \geq 1$) occurs, and the process transitions to state $j \in \mathcal{S}$ with probability $p_{ij}(k)$. For the second type, the batch size is 0 indicating no arrival and the process transitions to state $j \neq i$ with probability $p_{ij}(0)$. For each state $i \in \mathcal{S}$, the probabilities $p_{ij}(k)$ satisfy

$$\sum_{k=1}^{K}\sum_{j=1}^{m} p_{ij}(k) + \sum_{j\in\mathcal{S}\setminus\{i\}} p_{ij}(0) = 1.$$

In the context of a $BMAP_m(K)$, the matrices, $\{\mathbf{D}_0, \mathbf{D}_k, k = 1, ..., K\}$, are said to form a representation of the $BMAP_m(K)$, i.e., the $BMAP_m(K)$ is completely specified by these matrices. Here, $\mathbf{D}_0$ is an $m \times m$ matrix, with negative diagonal elements and non-negative off-diagonal elements, which represents the state transitions that correspond to no arrival occurring. $\mathbf{D}_k$ ($1 \leq k \leq K$) is an $m \times m$ matrix, with non-negative elements, which represents the state transitions that correspond to a batch arrival of size k. For $k = 0, 1, ..., K$, these matrices $\mathbf{D}_k = [d_{i,j}(k)]_{i,j\in\mathcal{S}}$ are defined as follows,

$$d_{i,j}(0) = \begin{cases} -\lambda_i, & j = i, \\ \lambda_i p_{i,j}(0) & j \neq i, \end{cases}$$

$$d_{i,j}(k) = \lambda_i p_{i,j}(k), i, j \in \mathcal{S}.$$

The matrix $\mathbf{D}_0$ is assumed to be stable and nonsingular. The definition of the rate matrices implies that

$$\mathbf{Q} = \sum_{k=0}^{K} \mathbf{D}_k,$$

which is the infinitesimal generator of the underlying Markov process $J(t)$. The steady state probability vector $\boldsymbol{\pi}$ of the process $J(t)$ is the solution of the linear system $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$, $\boldsymbol{\pi}\mathbf{e} = 1$ where $\mathbf{e}$ is a column vector of ones.

2.2 Performances of the Interarrival Times and Batch Sizes

Next, we provide some properties of the interarrival times and batch sizes of $BMAP_m(K)$. Its discrete time process embedded at arrival instants play an important role in the performance analysis of the $BMAP_m(K)$ model.

An important property of the $BMAP_m(K)$ concerns Markovian renewal theory. Let S_n denote the state of the process $J(t)$ at the time of the nth arrival event. Then $\{S_n\}_{n=0}^{\infty}$ is a Markov chain and its transition matrix is $\mathbf{P} = (-\mathbf{D}_0)^{-1}\mathbf{D}$ where $\mathbf{D} = \sum_{k=1}^{K} \mathbf{D}_k$. The steady state probability vector $\boldsymbol{\phi}$ of the embedded process is the solution of the linear equations $\boldsymbol{\phi}\mathbf{P} = \boldsymbol{\phi}$, $\boldsymbol{\phi}\mathbf{e} = 1$. From the results in [14], the steady state distributions of the original process $J(t)$ and the embedded process S_n are related as follows

$$\boldsymbol{\phi} = (\boldsymbol{\pi}\mathbf{D}\mathbf{e})^{-1}\boldsymbol{\pi}\mathbf{D}.$$

Interarrival Time. Let T_n denote the interarrival time between the $(n-1)$th and nth events in the $BMAP_m(K)$ model. The variables T_n's are phase-type distributed with representation $\{\boldsymbol{\phi}, \mathbf{D}_0\}$ (see [22] for more details).
The distribution of the interarrival times T_n in the stationary case is

$$P(T < t) = 1 - \boldsymbol{\phi} e^{\mathbf{D}_0 t}\mathbf{e}, \quad t > 0.$$

The moments of T_n in the stationary case are given by

$$\mu_r = E[T^r] = r!\boldsymbol{\phi}(-\mathbf{D}_0)^{-r}\mathbf{e}, \quad r \geq 1.$$

The lag-r correlation function of the sequence of interarrival times is

$$\rho_T(r) = \rho(T_1, T_{r+1}) = \frac{\mu_1 \boldsymbol{\pi}[(-\mathbf{D}_0)^{-1}\mathbf{D}]^{r-1}(-\mathbf{D}_0)^{-1}\mathbf{e} - \mu_1^2}{\mu_2 - \mu_1^2}, \quad r \geq 1. \tag{1}$$

Batch Size. Let B_n denote the batch size of the nth arrival in the $BMAP_m(K)$ model. From the results in [20], we know that
The mass probability function of the stationary batch size, B, is

$$P(B = k) = \boldsymbol{\phi}(-\mathbf{D}_0)^{-1}\mathbf{D}_k\mathbf{e}, \quad k = 1, 2, ..., K.$$

The moments of the stationary batch size B are obtained as

$$\beta_r = E[B^r] = \boldsymbol{\phi}(-\mathbf{D}_0)^{-1}\mathbf{D}_r^*\mathbf{e}, \quad r \geq 1, \tag{2}$$

where $\mathbf{D}_r^* = \sum_{k=1}^{K} k^r \mathbf{D}_k$. Also, the lag-$r$ correlation function in the stationary version of the batch process is given by

$$\rho_B(r) = \rho(B_1, B_{r+1}) = \frac{\boldsymbol{\phi}(-\mathbf{D}_0)^{-1}\mathbf{D}_1^*[(-\mathbf{D}_0)^{-1}\mathbf{D}]^{r-1}(-\mathbf{D}_0)^{-1}\mathbf{D}_1^*\mathbf{e} - \beta_1^2}{\beta_2 - \beta_1^2}. \tag{3}$$

By the joint Laplace Stieltjes transform (LST) of the interarrival times and batch sizes of a stationary $BMAP_m(K)$ given in the Lemma 1 of [20], we obtain $E[TB]$ as follows

$$\eta = E[TB] = \boldsymbol{\phi}(-\mathbf{D}_0)^{-2}\mathbf{D}_1^*\mathbf{e}. \tag{4}$$

Moreover, the covariance between T and B is derived as

$$cov(T, B) = \boldsymbol{\phi}(-\mathbf{D}_0)^{-2}\mathbf{D}_1^*\mathbf{e} - \boldsymbol{\phi}(-\mathbf{D}_0)^{-1}\mathbf{e}\boldsymbol{\phi}(-\mathbf{D}_0)^{-1}\mathbf{D}_1^*\mathbf{e}. \tag{5}$$

All the performances of the interarrival times and batch sizes for the BMAP model obtained above are very important. It will be used in our proposed algorithm for estimating BMAP parameters in the next section. Especially, the analytically tractable BMAP makes it possible for us to take the MM algorithm to consider the parameters estimation. That is, we can compute a BMAP that matches or approximates the performances of the observed processes including the interarrival times and packet sizes for a traffic data.

3 The Fitting Procedure

In terms of the observed information of the teletraffic trace, the sequences of packet interarrival times $\mathbf{t} = (t_1, t_2, ..., t_n)$ and packet sizes $\mathbf{b} = (b_1, b_2, ..., b_n)$ constitute the available observed samples. In this section, we summarize the theoretical issues of the fitting procedure for the $BMAP_m(K)$ model with the observing time series. Based on the properties of the $BMAP_m(K)$ (marginal distribution and lag correlation), we present a general framework of two-step $BMAP_m(K)$ fitting algorithm. Generally speaking, the main idea of the applied approach is that the matrix $\mathbf{D}_0$ and the matrices $\mathbf{D}_k$, $k = 1, 2, ..., K$ are constructed separately.

- In the first step, the interarrival time distribution is fitted by a phase-type distribution from the sequences of interarrival times, This procedure could determine the $\mathbf{D}_0$ matrix and the $\boldsymbol{\phi}$ vector.
- Then, the $\mathbf{D}_k$, $k = 1, 2, ..., K$ matrices are constructed by solving a non-linear optimization problem, such that the interarrival time distribution of the resulting $BMAP_m(K)$ remains the same, and its lag correlation functions of the interarrival times and batch sizes approximate the traffic data.

3.1 Constructing the D_0 Matrix and the ϕ Vector

At the first step of the procedure, we try to fit a phase type distribution of arrival times. The interarrival time distribution of the original process can be given with its samples or by a given number of moments. Among a number of past research results on PH fitting, we find some fitting methods to solve the phase-type fitting problem [8,23].

For the estimation of the $\mathbf{D}_0$ matrix and the $\boldsymbol{\phi}$ vector, we can use the following software tools for analyzing the PH fitting, which are based on EM algorithm or MM algorithm [5]. By reviewing the relative literature, we find three fitting tools (KPC-Toolbox [7], BUTools [12], Mapfit [19]), which could solve the phase-type fitting problem. In our paper, we will use the fitting method from BUTools [12], which provides the fast EM algorithms for PH fitting with teletraffic data.

3.2 Constructing the D_k Matrices

In this subsection, we provide the detailed estimating process of the matrices $\mathbf{D}_k$, $k = 1, 2, ..., K$. The characterization of $BMAP_m(K)$ can be extended from the case $K = 2$ to the case with an arbitrary maximum batch size K. The generalization of such process is due to the fact that given a $BMAP_m(K)$ represented by $\mathbf{B}_K = \{\mathbf{D}_0, \mathbf{D}_1, ..., \mathbf{D}_K\}$, then K different $BMAP_m(2)$s can be obtained as

$$\mathbf{B}_2^i = \{\mathbf{D}_0, \mathbf{D}_i, \sum_{k \neq i} \mathbf{D}_k\}, \quad i = 1, 2, ..., K.$$

Then, we show the constructing process of $\mathbf{D}_1$ in the $BMAP_m(2)$ model, i.e., $\mathbf{B}_2^1 = \{\mathbf{D}_0, \mathbf{D}_1, \mathbf{D}_{-1}\}$ where $\mathbf{D}_{-1} = \mathbf{D}_2 + \mathbf{D}_3 + \cdots + \mathbf{D}_K$. It should be pointed out that, in order to compute the empirical moment β_r and lag-r correlation $\overline{\rho}_B(r)$, all batch sizes in $\mathbf{b}$ larger than 2 are considered as equal to 2.

Constraints of the $\mathbf{D}_1$ and $\mathbf{D}_{-1}$ Matrices. Once the matrix $\mathbf{D}_0$ and the vector $\boldsymbol{\phi}$ are obtained by the first step in Sect. 3.1, we consider the constraints of the matrices $\mathbf{D}_1$ and $\mathbf{D}_{-1}$ based on the performances of the $BMAP_m(K)$ model. The matrices $\mathbf{D}_1$ and $\mathbf{D}_{-1}$ have to satisfy the following two constraints to maintain the interarrival time distribution determined in the first step:

- C1: $(\mathbf{D}_1 + \mathbf{D}_{-1})\mathbf{e} = -\mathbf{D}_0\mathbf{e}$,
- C2: $\boldsymbol{\phi}(-\mathbf{D}_0)^{-1}(\mathbf{D}_1 + \mathbf{D}_{-1}) = \boldsymbol{\phi}$.

Exact Performance Measures Fitting. From Sect. 2.2, we find that the moments of the batch size β_r and the joint expected of interarrival time and batch size $E[TB]$ can also be expressed as linear constraints. Specially, by Eq. (3), we know that the lag correlation function of the batch process depends on the first and second moments of the batch sizes. Here we provide the following linear constraints to exact β_1, β_2 (related to the batch size distribution) and $\eta = E[TB]$ (joint moments concerning interarrival times and batch sizes) fitting.

- C3: $\beta_1 = \boldsymbol{\phi}(-\mathbf{D}_0)^{-1}(\mathbf{D}_1 + 2\mathbf{D}_{-1})\mathbf{e}$,
- C4: $\beta_2 = \boldsymbol{\phi}(-\mathbf{D}_0)^{-1}(\mathbf{D}_1 + 4\mathbf{D}_{-1})\mathbf{e}$,
- C5: $\eta = \boldsymbol{\phi}(-\mathbf{D}_0)^{-2}(\mathbf{D}_1 + 2\mathbf{D}_{-1})\mathbf{e}$.

We formulate these constrains $C1 \sim C5$ as a linear system of equations. To do so, we introduce a column vector $\mathbf{x}$ (of size $2m^2$), which is composed by the columns of the matrices $\mathbf{D}_1$ and $\mathbf{D}_{-1}$ as below.

$$\mathbf{x} = \begin{pmatrix} \{\mathbf{D}_1\}_1 \\ \vdots \\ \{\mathbf{D}_1\}_m \\ \{\mathbf{D}_{-1}\}_1 \\ \vdots \\ \{\mathbf{D}_{-1}\}_m \end{pmatrix}.$$

All possible $\mathbf{x}$ vectors (thus, $\mathbf{D}_1$ and $\mathbf{D}_{-1}$ matrices) satisfying constraints $C1 \sim C5$ are the solutions of the following linear equations with coefficient matrix $\mathcal{A}$

$$\begin{bmatrix} I & I & \cdots & I & I & I & \cdots & I \\ \omega & & & & \omega & & & \\ & \omega & & & & \omega & & \\ & & \ddots & & & & \ddots & \\ & & & \omega & & & & \omega \\ \omega & \omega & \cdots & \omega & 2\omega & 2\omega & \cdots & 2\omega \\ \omega & \omega & \cdots & \omega & 4\omega & 4\omega & \cdots & 4\omega \\ \varphi & \varphi & \cdots & \varphi & 2\varphi & 2\varphi & \cdots & 2\varphi \end{bmatrix}_{(2m+3)\times 2m^2} \cdot \begin{bmatrix} \\ \mathbf{x} \\ \\ \end{bmatrix}_{2m^2\times 1} = \begin{bmatrix} d \\ \\ \phi \\ \\ \beta_1 \\ \beta_2 \\ \eta \end{bmatrix}_{(2m+3)\times 1}, \qquad (6)$$

where $\omega = \boldsymbol{\phi}(-\mathbf{D}_0)^{-1}$, $\varphi = \boldsymbol{\phi}(-\mathbf{D}_0)^{-2}$ and $d = -\mathbf{D}_0\mathbf{e}$. The first m lines of $\mathcal{A}$ correspond to constraint $C1$, and the second m lines of $\mathcal{A}$ are related to constraint $C2$. The last three lines of $\mathcal{A}$ correspond to constraint $C3 \sim C5$. Proper matrices $\mathbf{D}_1$ and $\mathbf{D}_{-1}$ (i.e., $\mathbf{x}$ vector) satisfy the following set of linear equations and inequalities:

$$\mathcal{A}\mathbf{x} = b, \quad \mathbf{x} \geq 0. \tag{7}$$

As we can see, the $\mathcal{A}\mathbf{x} = b$ equation is under-determined for $m \geq 2$, since we have $2m+3$ equations and $2m^2$ unknowns. So that, the $\mathbf{D}_0$ matrix, the $\boldsymbol{\phi}$ vector, and the performance measures $\{\beta_1, \beta_2, \eta\}$ cannot determine the $\mathbf{D}_1$ and $\mathbf{D}_{-1}$ matrices when $m \geq 2$. Moreover, the study in [20] explores the identifiability of the stationary two-state $BMAP_m(K)$. The results in [20] show that the $BMAP_m(K)$ is not identifiable if $m = 2$ and $K \geq 2$.

Fitting with Lag Correlations. In order to estimate the matrices $\mathbf{D}_1$ and $\mathbf{D}_{-1}$, we use the lag correlation functions of the interarrival times and batch sizes. To fit lag correlation values, we define an optimization problem with the linear constraints (7) such that a properly chosen goal function ensures the approximation of lag correlation values. The fitting of a given number of lag correlations is a linearly constrained nonlinear optimization problem.

We apply the objective function $c(\mathbf{x})$, which is the squared difference between the lag correlations of the original process for $(\overline{\rho}_T(r), \overline{\rho}_B(r))$ and the fitted $(\rho_T(r), \rho_B(r))$ of the $BMAP_m(K)$ model weighted with w_r and v_r, respectively. Then we can set up the following nonlinear optimization problem

$$c(\mathbf{x}) = \sum_{r=1}^{R_1} w_r \left[\frac{\rho_T(r) - \overline{\rho}_T(r)}{\overline{\rho}_T(r)}\right]^2 + \sum_{r=1}^{R_2} v_r \left[\frac{\rho_B(r) - \overline{\rho}_B(r)}{\overline{\rho}_B(r)}\right]^2. \tag{8}$$

In the objective function (8), $\rho_T(r)$ and $\rho_B(r)$ are the functions w.r.t. $\mathbf{x}$ (thus, $\mathbf{D}_1$ and $\mathbf{D}_{-1}$ matrices), presented in Eqs. (1) and (3), respectively. The largest lag correlation coefficient of interarrival times and batch sizes considered in this objective function are the lag-R_1 and lag-R_2 correlation coefficients, respectively. The weights w_r and v_r can be used to increase the importance of the accuracy of lower lag-r correlation with respect to higher ones or vice-versa.

Once the $\mathbf{D}_1$ matrix is obtained as the solutions of the nonlinear optimization problem with linear constraints (7), the approach will be repeated for estimating $\mathbf{D}_2$ (using the representation of $\mathbf{B}_2^2 = \{\mathbf{D}_0, \mathbf{D}_2, \mathbf{D}_{-2}\}$), $\mathbf{D}_3, ...,$ and finally $\mathbf{D}_K$. The computation process is summarized in Algorithm 1. It is important to comment that the optimization problems of (9) in Algorithm 1 for $k = 1, ..., K$ are straightforward problems in $2m^2$ variables each, solved by using standard optimization routines (fmincon in MATLAB), where a multistart with 100 randomly chosen starting points was executed.

4 Numerical Experiments

To illustrate the fitting properties of our proposed approach, we conduct two numerical experiments in this section. The first one approximates the sample data generated by a given $BMAP_m(K)$. We show how close the result of

Algorithm 1. A two-step EM-MM algorithm for fitting $BMAP_m(K)$.

1. **Input:**
Trace: the sequences of interarrival times $\mathbf{t} = (t_1, t_2, ..., t_n)$ and batch sizes $\mathbf{b} = (b_1, b_2, ..., b_n)$.
2. **A phase-type fitting for interarrival times**
• Obtain $\mathbf{D}_0$ and $\boldsymbol{\phi}$ by the EM algorithm (from BUTools) on the trace of interarrival times $\mathbf{t} = (t_1, t_2, ..., t_n)$.
3. **The lag correlation fitting for interarrival time and batch sizes**
For $k = 1, ..., K-1$ repeat:
• Compute the empirical moments $\{\beta_1, \beta_2, \eta\}$ and lag-r correlation $\{\rho_B(r), \rho_T(r)\}$ of the trace of interarrival times $\mathbf{t} = (t_1, t_2, ..., t_n)$ and batch sizes $\mathbf{b} = (b_1, b_2, ..., b_n)$ by Eqs. (1)–(4).
• From the obtained $\mathbf{D}_0, ..., \mathbf{D}_{k-1}$, consider the $BMAP_m(2)$: $\mathbf{B}_2^k = \{\mathbf{D}_0, \mathbf{D}_i, \sum_{i \neq k} \mathbf{D}_i\}$, and obtain $\mathbf{D}_k$ as the solutions of

$$\min \sum_{r=1}^{R_1} w_r \left[\frac{\rho_T(r) - \overline{\rho}_T(r)}{\overline{\rho}_T(r)}\right]^2 + \sum_{r=1}^{R_2} v_r \left[\frac{\rho_B(r) - \overline{\rho}_B(r)}{\overline{\rho}_B(r)}\right]^2 \tag{9}$$
$$s.t. \quad \mathcal{A}\mathbf{x} = b, \quad \mathbf{x} \geq 0.$$

It can be solved by using standard optimization routines (fmincon in MATLAB), where a multistart with 100 randomly chosen starting points is executed.
4. **Output:**
The parameters of the fitted $BMAP_m(K)$ model: $\mathrm{M} = \{\mathbf{D}_0, \mathbf{D}_1, ..., \mathbf{D}_K\}$.

our fitting algorithm is in terms of the performances on the interarrival times and packet sizes. In the second experiment, a real teletraffic data is fitted of a $BMAP_m(K)$ model by our proposed Algorithm 1. This teletraffic trace data is measured from a popular video game, which is provided from a practical project collaborated with Huawei company.

4.1 BMAP Traffic Modeling Framework

In our numerical experiments, the first data set was simulated from a $BMAP_3(2)$ model represented by the rate matrices $\{\mathbf{D}_0, \mathbf{D}_1, \mathbf{D}_2\}$ shown in the second column of the top part of Table 1. Based on this parameter set, a trace file with $n = 100,000$ arrivals and corresponding batch sizes is generated. This trace file is used as input for our proposed Algorithm 1 in order to derive the (known) parameter set of the BMAP model. This numerical experiment was conduct to demonstrate the accuracy of our proposed two-step EM-MM algorithm for fitting BMAP. The results obtained by our proposed approach are shown in Table 1. The second column in Table 1 presents the generator process and the characterizing theoretical performances according to Sect. 2.2. The third column in Table 1 presents the empirical moments from the simulated traces. The rest of columns present the estimated rate matrices and estimated characterizing performances by using the Algorithm 1. As can be seen in Table 1, very good results were obtained since the fitted BMAP matches closely the performances of the empirical data. Moreover, we should note that the estimated parameters of the fitted BMAP are different from that of the generator process. This is due to the fact that the $BMAP_m(K)$ is not identifiable when $K \geq 2$.

Table 1. Performances of the estimation method for a simulated trace from a $BMAP_3(2)$.

	Generator process	Empirical	Fitted process
$\mathbf{D}_0$	$\begin{pmatrix} -10 & 0 & 0 \\ 0 & -20 & 0 \\ 0 & 0 & -30 \end{pmatrix}$	–	$\begin{pmatrix} -8.6109 & 0 & 0 \\ 0 & -12.7691 & 0 \\ 0 & 0 & -26.1805 \end{pmatrix}$
$\mathbf{D}_1$	$\begin{pmatrix} 2 & 3 & 1 \\ 0 & 3 & 5 \\ 10 & 8 & 2 \end{pmatrix}$	–	$\begin{pmatrix} 0.5926 & 0.7304 & 2.5585 \\ 0.0000 & 3.0802 & 4.5875 \\ 0.5646 & 3.6611 & 8.9966 \end{pmatrix}$
$\mathbf{D}_2$	$\begin{pmatrix} 1 & 1 & 2 \\ 3 & 7 & 2 \\ 5 & 2 & 3 \end{pmatrix}$	–	$\begin{pmatrix} 1.1467 & 1.7818 & 1.8008 \\ 0.0000 & 3.2711 & 1.8304 \\ 3.1485 & 3.6417 & 6.1681 \end{pmatrix}$
μ_1	5.9935e−2	5.9796e−2	6.0122e−2
β_1	1.4207e−1	9.6640e−2	1.3730e−1
$\rho_B(1)$	1.4658	1.4680	1.4660
$\rho_T(1)$	5.5148e−5	7.8551e−3	6.3589e−4
$E[TB]$	8.7514e−2	8.7414e−2	8.7812e−2
$Cov(T,B)$	3.3882e−4	3.7050e−4	3.2825e−4

4.2 Comparative Study of IP Traffic Modeling

As a real example, we use the trace from a popular game which is provided from our project collaborated with Huawei company. The underlying trace traffic including 393206 packets, comprises packet interarrival times and the corresponding packet lengths. The analysis of the packet length distributions reveals that the packet lengths of the real traffic trace follow to a large extent of discrete values. Recalling the definition of $BMAP_m(K)$ in Sect. 2, the mapping process of packet lengths to $BMAP_m(K)$ rewards results in K kinds of different batch sizes. We map the packet lengths from a real traffic trace onto the discrete packet lengths s_k, for $1 \leq k \leq K$, where s_k is the average length of all packets of the considered trace comprising packet lengths between $\frac{L(k-1)}{K}$ bytes and $\frac{Lk}{K}$ bytes, where L denotes the maximum packet length of the real traffic trace. The aggregated traffic model utilizes these observations where different reward values of the $BMAP_m(K)$ represent different discrete packet lengths. In our numerical study, the parameter estimation procedure is applied for the $BMAP_m(K)$ model with $K = 2$ distinct batch sizes. The largest packet length is $L = 1506$ bytes in the original WZRY trace.

We consider a $BMAP_3(2)$ model to capture the packet arrival process of the considered trace by using Algorithm 1. We take the parameters $R_1 = R_2 = 5$ in our numerical experiments. We use weights $w_r = v_r = 10^{-(r-1)}$ for the fitting of lag-r correlations for interarrival times and batch sizes. The average packet lengths of our measurements are as follows: $s_1 = 77.4887$ bytes and $s_2 = 1113.3234$ bytes. To show the efficiency and accuracy of the BMAP traffic model and our proposed fitting method, we make a numerical comparison study with the other existing Markovian models, MAP and MMPP. Note that the parameter matrices $\mathbf{D}_0$ and $\mathbf{D}_1$ of the MAP, MMPP were estimated by means of the EM algorithm from BUTools [12]. For the BMAP, MAP, MMPP, the corresponding fitting time and estimated parameter sets are derived as follows:

(1) BMAP model: the fitting time is 292.32 s and the estimated parameters are as follows

$$\mathbf{D}_0 = \begin{pmatrix} -237.00 & 141.46 & 74.94 \\ 135.72 & -13792.60 & 6216.24 \\ 113.55 & 2032.53 & -14919.27 \end{pmatrix}, \mathbf{D}_1 = \begin{pmatrix} 8.34 & 9.18 & 3.10 \\ 0.89 & 6089.52 & 899.54 \\ 0.11 & 631.49 & 8.85 \end{pmatrix}.$$

$$\mathbf{D}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0.02 & 441.38 & 9.29 \\ 0 & 0 & 0 \end{pmatrix}.$$

(2) MAP model: the fitting time is 536.92 s and the estimated parameters are as follows

$$\mathbf{D}_0 = \begin{pmatrix} -17011.63 & 8.98 & 3573.99 \\ 1.46 & -335.49 & 1.06 \\ 43.97 & 0.73 & -2052.95 \end{pmatrix}, \mathbf{D}_1 = \begin{pmatrix} 9083.65 & 40.13 & 4304.88 \\ 1.89 & 329.97 & 1.10 \\ 738.95 & 0.68 & 1268.62 \end{pmatrix}.$$

(3) MMPP model: the fitting time is 433.13 s and the estimated parameters are as follows

$$\mathbf{D}_0 = \begin{pmatrix} -3042.07 & 2150.33 & 0.31 \\ 6511.04 & -16053.45 & 22.14 \\ 2.29 & 3.26 & -334.04 \end{pmatrix}, \mathbf{D}_1 = \begin{pmatrix} 891.43 & 0 & 0 \\ 0 & 9520.26 & 0 \\ 0 & 0 & 328.49 \end{pmatrix}.$$

Table 2. The moments of interarrival times of the fitted models and original process.

Moment ordinal	Raw trace	BMAP	MAP	MMPP
$E[T]$	$6.0254e^{-4}$	$6.0328e^{-4}$	$6.5187e^{-4}$	$6.1063e^{-4}$
$E[T^2]$	$2.8704e^{-6}$	$2.6430e^{-6}$	$2.4471e^{-6}$	$2.1487e^{-6}$
$E[T^3]$	$6.5608e^{-8}$	$2.9795e^{-8}$	$2.0150e^{-8}$	$1.6958e^{-8}$
$E[T^4]$	$6.6922e^{-9}$	$4.8607e^{-10}$	$2.4367e^{-10}$	$1.9352e^{-10}$

In this numerical experiment, the parameters of a BMAP model are estimated by our proposed two-step EM-MM algorithm, and the parameters of the MAP and MMPP models are estimated by the EM algorithm from [12]. The numerical results show that the fitting time of the BMAP model is almost half of the

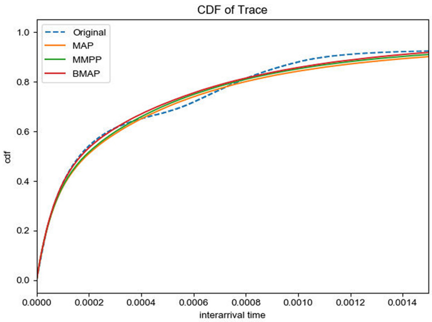

Fig. 1. The CDF of interarrival times of the fitted MMPP, MAP, BMAP and the original process.

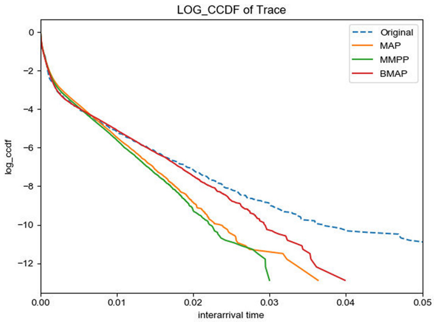

Fig. 2. The Log(ccdf) of interarrival times of the fitted MMPP, MAP, BMAP and the original process.

fitting time of the MAP, MMPP models, which illustrates that our proposed two-step EM-MM algorithm for fitting the BMAP model is very efficient. Moreover, we present statistical properties for the measured traffic data, and the fitted BMAP, MAP, MMPP models in terms of the interarrival times. As we can see, Table 2 shows the moments of the interarrival times of the fitted MMPP, MAP, BMAP models and the original process. We observe that the mean and standard deviation of the interarrival times approximation are very accurate. Figure 1 presents the CDF (cumulative distribution function) of the interarrival times of the fitted BMAP, MAP, MMPP models and the original process. From Fig. 1, we see that the three statistic models perform very well w.r.t. the CDF of interarrival times. Moreover, Fig. 2 gives the Log of CCDF (complementary CDF) of the interarrival times of the fitted MMPP, MAP, BMAP models and the original process. It is obvious to find that the CCDF of the measured traffic and the BMAP, MAP, MMPP models perform different. The BMAP model is slightly better than the MAP and MMPP models.

5 Conclusions

In this paper, we present a two-step parameter fitting approach for BMAP to model real teletraffic traces. The approach is based on a two-step procedure combining EM and MM algorithms, where the first step is the phase-type fitting of the interarrival times and the second step is the lag correlation fitting of interarrival times and packet sizes. We evaluate our proposed BMAP fitting approach in two ways. We perform a fitting experiment for BMAP based on a trace data set generated by a given BMAP, such that we can evaluate the accuracy of our fitting method with theoretical values. We also conduct the second experiment fitting the models of BMAP, MAP, MMPP based on a real teletraffic trace, respectively. The numerical results illustrate the advantages of the BMAP modeling approach over other widely used analytically tractable models, MAP and MMPP. Such advantage mainly comes from the strong approximation capability of BMAP, since the class of BMAP includes the well-known Poisson process, MMPP, and MAP as special cases, and BMAP is able to capture the interarrival time and packet size distributions simultaneously. The experiment results also show that our two-step fitting approach can achieve a good balance between the computation efficiency and accuracy. Based on the results in this paper, one direction for future research is to study the performance analysis and optimization of the queueing system with BMAP inputs.

References

1. Breuer, L.: Parameter estimation for a class of BMAPs. In: Latouche, G., Taylor, P. (eds.) Advances in Algorithmic Methods for Stochastic Models: Proceedings of the 3rd International Conference on Matrix Analytic Methods, pp. 87–97. Notable Publications, Neshanic Station, New Jersey, USA, July 2000
2. Breuer, L.: An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure. Ann. Oper. Res. **112**(1), 123–138 (2002)
3. Buchholz, P.: An EM-algorithm for MAP fitting from real traffic data. In: Kemper, P., Sanders, W.H. (eds.) TOOLS 2003. LNCS, vol. 2794, pp. 218–236. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45232-4_14
4. Buchholz, P., Kriege, J.: A heuristic approach for fitting MAPs to moments and joint moments. In: 2009 Sixth International Conference on the Quantitative Evaluation of Systems, pp. 53–62. IEEE (2009)
5. Buchholz, P., Kriege, J., Felko, I.: Software tools. In: Input Modeling with Phase-Type Distributions and Markov Models. SM, pp. 111–114. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-06674-5_7
6. Buchholz, P., Panchenko, A.: A two-step EM algorithm for MAP fitting. In: Aykanat, C., Dayar, T., Körpeoğlu, İ (eds.) ISCIS 2004. LNCS, vol. 3280, pp. 217–227. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30182-0_23
7. Casale, G., Zhang, E.Z., Smirni, E.: KPC-toolbox: simple yet effective trace fitting using Markovian arrival processes. In: 2008 Fifth International Conference on Quantitative Evaluation of Systems, pp. 83–92. IEEE (2008)
8. Fiondella, L., Puliafito, A.: Principles of Performance and Reliability Modeling and Evaluation. Springer (2016). https://doi.org/10.1007/978-3-319-30599-8
9. Fischer, W., Meier-Hellstern, K.: The Markov-modulated Poisson process (MMPP) cookbook. Performance Eval. **18**(2), 149–171 (1993)
10. Heindl, A., Horvath, G., Gross, K.: Explicit inverse characterizations of acyclic MAPs of second order. Formal methods, pp. 108–122 (2006)
11. Horvath, G., Buchholz, P., Telek, M.: A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation, pp. 124–133 (2005)
12. Horváth, G., Telek, M.: BuTools 2: a rich toolbox for Markovian performance evaluation. In: VALUETOOLS (2016)
13. Klemm, A., Lindemann, C., Lohmann, M.: Modeling IP traffic using the batch Markovian arrival process. Perform. Eval. **54**(2), 149–173 (2003)
14. Latouche, G., Ramaswami, V.: Introduction to matrix analytic methods in stochastic modeling. J. Am. Stat. Assoc. **95**(452), 1379 (1999)

15. Ledesma, S., Liu, D.: Synthesis of fractional Gaussian noise using linear approximation for generating self-similar network traffic. ACM SIGCOMM Comput. Commun. Rev. **30**(2), 4–17 (2000)
16. Lucantoni, D.M.: New results on the single server queue with a batch Markovian arrival process. Commun. Stat. Stochas. Mod. **7**(1), 1–46 (1991)
17. Neuts, M.F.: A versatile Markovian point process. J. Appl. Probab. **16**(4), 764–779 (1979)
18. Okamura, H., Dohi, T.: Faster maximum likelihood estimation algorithms for Markovian arrival processes, pp. 73–82, September 2009
19. Okamura, H., Dohi, T.: Fitting phase-type distributions and Markovian arrival processes: algorithms and tools. In: Fiondella, L., Puliafito, A. (eds.) Principles of Performance and Reliability Modeling and Evaluation. SSRE, pp. 49–75. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30599-8_3
20. Rodriguez, J., Lillo, R.E., Ramirezcobo, P.: Nonidentifiability of the two-state BMAP. Methodol. Comput. Appl. Probab. **18**(1), 81–106 (2016)
21. Salvador, P., Pacheco, A., Valadas, R.: Modeling IP traffic: joint characterization of packet arrivals and packet sizes using BMAPs. Comput. Netw. **44**(3), 335–352 (2004)
22. Telek, M., Horvath, G.: A minimal representation of Markov arrival processes and a moments matching method. Perform. Eval. **64**(9), 1153–1168 (2007)
23. Thummler, A., Buchholz, P., Telek, M.: A novel approach for fitting probability distributions to real trace data with the EM algorithm. In: 2005 International Conference on Dependable Systems and Networks (DSN 2005), pp. 712–721. IEEE (2005)