Performance Evaluation 67 (2010) 759–778

Contents lists available at ScienceDirect

Performance Evaluation

journal homepage: www.elsevier.com/locate/peva

A joint moments based analysis of networks of MAP/MAP/1 queues[☆]

András Horváth [a,*], Gábor Horváth [b], Miklós Telek [b]

[a] *Dipartimento di Informatica, Università di Torino, Torino, Italy*

[b] *Department of Telecommunications, Budapest University of Technology and Economics, H-1521 Budapest, Hungary*

ARTICLE INFO

Article history:
Received 23 February 2009
Received in revised form 30 July 2009
Accepted 19 December 2009
Available online 11 January 2010

Keywords:
Queueing network
Markov arrival process
MAP/MAP/1 queue
Matrix geometric method
Joint moment

ABSTRACT

The decomposition based approximate numerical analysis of queueing networks of MAP/MAP/1 queues with FIFO scheduling is considered in this paper. One of the most crucial decisions of decomposition based queueing network analysis is the description of inter-node traffic. Utilizing recent results on Markov arrival processes (MAPs), we apply a given number of joint moments of the consecutive inter-event times to describe the inter-node traffic. This traffic description is compact (uses far less parameters than the alternative methods) and flexible (allows an easy reduction of the complexity of the model). Numerical examples demonstrate the accuracy and the computational complexity of the proposed approximate analysis method.

1. Introduction

Open queueing networks are a popular modelling tool for the performance analysis of computer and telecommunication systems. Exact solution methods are available only for networks with Poisson traffic input, specific service time distribution and service discipline [1]. These restrictive assumptions make the exact solutions unlikely to use in practice. The main reason is that in real systems the Poisson process is usually not a good model for the traffic behaviour. Instead the real traffic can be bursty and correlated, and the service times in the service stations can be correlated as well. Since these features have an impact on the performance measures, they have to be taken into consideration.

The attempts to analyse queueing networks with non-Poisson traffic and non-exponential service time distributions dates back to the second half of the last century. The first attempts considered the second moments of the inter-arrival and the service time distributions in the computations. A widely applied approximation of this kind was integrated into the QNA tool [2,3]. The intrinsic assumption in these approximations is that the consecutive inter-arrival times and the consecutive service times are independent. The evolution of packet switched communication networks during the eighties and nineties resulted in teletraffic with significant correlation which led to the development of new modelling paradigms.

Several modelling approaches were developed to better describe the properties of packet traffic [4]. One of the lines of research is based on Markovian models with the aim of extending the Poisson arrival process in order to capture more statistical properties of the traffic behaviour. A long series of efforts resulted in the introduction of Markov arrival processes (MAPs) as is surveyed in [5]. A main advantage of using Markovian models for traffic description of queues is that there are efficient numerical analysis methods, commonly referred to as matrix analytic methods, for the evaluation of a Markovian queue (see e.g., [5] for an introduction and [6] for a set implemented method).

☆ This work is partially supported by the OTKA K61709 grant, by MIUR through PRIN project Famous and by EEC project Crutial. This paper is an extended version of the conference paper presented at QEST 2008.

∗ Corresponding author.
E-mail addresses: horvath@di.unito.it (A. Horváth), ghorvath@hit.bme.hu (G. Horváth), telek@hit.bme.hu (M. Telek).

0166-5316/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.peva.2009.12.006

The availability of flexible Markovian models gave a new impulse to the research on queueing network analysis [7–10]. Several approximate analysis methods were developed to combine the results of these two fields for accurate approximation of queueing networks with Markovian node behaviour. In this paper we present a new method along this line of research which is based on a recent result about the moments based representation of MAPs [11].

The rest of the paper is organised as follows. Section 2 introduces MAPs and some of their properties that are used in the sequel. Section 3 provides a summary on the available MAP based queueing network approximation methods. The proposed approximation procedure is introduced in Section 4, which includes the exact computation of the moments and joint moments of consecutive inter-departure times of MAP/MAP/1 queues. The numerical behaviour of the method is illustrated by several examples in Section 5. Section 6 concludes the paper.

2. Summary on MAP results

The arrivals of a MAP (Markovian Arrival Process) are modulated by a background Markov chain. A transition in the background Markov chain generates an arrival with a given probability; in addition, during a sojourn in a state of the Markov chain, arrivals are generated according to a Poisson process whose intensity depends on the state. For a detailed introduction on MAP we refer, e.g., to [5]. Hereinafter we consider continuous time MAPs. The generator of the continuous time Markov chain (CTMC) that modulates the arrivals is denoted by $\mathbf{D}$, and the states of the Markov chain are referred to as the *phases* of the arrival process. MAPs are usually defined by two matrices. $\mathbf{D_0}$ describes the transition rates without an arrival and $\mathbf{D_1}$ describes the ones with an arrival event. Thus $\mathbf{D} = \mathbf{D_0} + \mathbf{D_1}$. Since $\mathbf{D}$ is a generator matrix, its row sums are equal to zero, i.e., $\mathbf{D1} = \mathbf{0}$, where $\mathbf{1}$ ($\mathbf{0}$) denotes the column vector of ones (zeros) of appropriate size. Consequently, $\mathbf{D_0 1} = -\mathbf{D_1 1}$. Let us denote the stationary distribution of the phase process by α. α is the solution of the linear system $\alpha\mathbf{D} = 0, \alpha\mathbf{1} = 1$. The average arrival intensity of the MAP is then computed by:

$$\lambda = \alpha\mathbf{D_1 1}. \tag{1}$$

In the analysis of MAPs, the phase of the background CTMC at arrival instants plays an important role. The phase process of the MAP at consecutive arrivals is referred to as the process embedded at arrival instants. The state transition probability matrix of the embedded process is $\mathbf{P} = (-\mathbf{D_0})^{-1}\mathbf{D_1}$. The stationary probability vector of the embedded process, π, is the solution of the linear system $\pi\mathbf{P} = \pi, \pi\mathbf{1} = 1$.

The stationary distribution of the underlying CTMC, α, and the stationary distribution of the underlying CTMC at arrival epochs, π, are related by [5]

$$\alpha = \lambda\pi(-\mathbf{D_0})^{-1} \quad \text{and} \quad \pi = \frac{1}{\lambda}\alpha(-\mathbf{D_0}). \tag{2}$$

In steady state, the inter-arrival time is phase type (PH) distributed with initial probability vector π and transient generator $\mathbf{D_0}$. Thus, the cumulative distribution function (cdf) of the inter-arrival time is

$$F_X(x) = P(X < x) = 1 - \pi e^{\mathbf{D_0}x}\mathbf{1}, \tag{3}$$

and its kth moment is

$$\mu_k = E(X^k) = k!\pi(-\mathbf{D_0})^{-k}\mathbf{1}. \tag{4}$$

Definition 1. A MAP is non-redundant when the order of its inter-arrival time distribution is identical with the size of $\mathbf{D_0}$.

The inter-arrival times in MAPs are not independent. The joint density function of the inter-arrival times $X_0, X_1, \ldots, X_k$ is:

$$f(x_0, x_1, \ldots, x_k) = \pi e^{\mathbf{D_0}x_0}\mathbf{D_1}e^{\mathbf{D_0}x_1}\mathbf{D_1}\cdots e^{\mathbf{D_0}x_k}\mathbf{D_1}\mathbf{1}. \tag{5}$$

From the joint density function the joint moments of the $a_0 = 0 < a_1 < a_2 < \cdots < a_k$-th inter-arrival times can be derived as

$$E(X_0^{i_0}X_{a_1}^{i_1}\cdots X_{a_k}^{i_k}) = \pi i_0!(-\mathbf{D_0})^{-i_0}\mathbf{P}^{a_1-a_0}i_1!(-\mathbf{D_0})^{-i_1}\cdots\mathbf{P}^{a_k-a_{k-1}}i_k!(-\mathbf{D_0})^{-i_k}\mathbf{1}. \tag{6}$$

Several statistical quantities can be used in practice to characterise the dependency structure of MAPs. One of the most popular ones is the lag-k autocorrelation, ρ_k, defined as

$$\rho_k = \frac{E(X_0X_k) - E(X)^2}{E(X^2) - E(X)^2}. \tag{7}$$

However, recent results on the steady state characterisation of order n non-redundant MAPs [11] showed that given $(n-1)^2$ joint moments of two consecutive inter-arrivals,

$$\eta_{ij} = E(X_0^iX_1^j), \quad i, j = 1, \ldots, n-1, \tag{8}$$

together with the first $2n-1$ moments of the inter-arrival time distribution,

$$\mu_i = E(X_0^i), \quad i = 1, \ldots, 2n-1, \tag{9}$$

completely characterise the process. Based on this set of n^2 moments and joint moments all other moments and joint moments, e.g.,

- the lag-k correlation for arbitrary k, ρ_k,
- the arbitrary joint moments, $E(X_0^{i_0} X_{a_1}^{i_1} \cdots X_{a_k}^{i_k})$,
- the derivatives of the complementary cumulative distribution function (ccdf) at $x = 0$,

$$\nu_i = \frac{\mathrm{d}^i}{\mathrm{d}x^i}(1 - F_X(x))\bigg|_{x=0} = \pi \mathbf{D_0}^i \mathbf{1}, \tag{10}$$

- the derivatives of the joint density $f(x_0, x_1)$ at $x_0 = x_1 = 0$,

$$\gamma_{ij} = \frac{\partial^i}{\partial x_0^i}\frac{\partial^j}{\partial x_1^j} f(x_0, x_1)\bigg|_{x_0=x_1=0} = \pi \mathbf{D_0}^i \mathbf{D_1} \mathbf{D_0}^j \mathbf{D_1} \mathbf{1} = -\pi \mathbf{D_0}^i \mathbf{D_1} \mathbf{D_0}^{j+1} \mathbf{1} = \pi \mathbf{D_0}^{i+1} \mathbf{P} \mathbf{D_0}^{j+1} \mathbf{1}, \tag{11}$$

- the derivatives of the joint density $f(x_0, \ldots, x_n)$ at $x_0 = \cdots = x_n = 0$,

$$\gamma_{i_0,\ldots,i_n} = \frac{\partial^i}{\partial x_0^{i_0}} \cdots \frac{\partial^{i_n}}{\partial x_n^{i_n}} f(x_0, \ldots, x_n)\bigg|_{x_0=\cdots=x_n=0} = \pi \mathbf{D_0}^{i_0} \mathbf{D_1} \mathbf{D_0}^{i_1} \mathbf{D_1} \cdots \mathbf{D_0}^{i_n} \mathbf{D_1} \mathbf{1}, \tag{12}$$

can be computed [12], where these derivatives can be considered as a kind of negative moments (comparing (4) and (10)) or joint moments (comparing (6) and (11)) with proper multiplication of the factorial terms.

Definition 2. We refer to the first $2n-1$ moments, (9), and the first $(n-1)^2$ joint moments, (8), of an order n non-redundant MAP as *basic moments set*.

Definition 3. A set of moments are called independent if the determinants of the moments matrices of size k defined in [12] are non-zero when $k \le n$.

The $\mathbf{D_0}$, $\mathbf{D_1}$ representation is not a unique description of a MAP. There are infinitely many matrix pairs which result in the same stationary behaviour. On the contrary, the representation given by the basic moments set is a unique description of a non-redundant MAP.

In Section 4 we propose an approximate analysis technique. This technique requires the computation of the joint moments of two consecutive inter-departure times from a MAP/MAP/1 queue. According to our knowledge, this step cannot be performed based on the basic moments set representation. For this reason we need to be able to generate the $\mathbf{D_0}$, $\mathbf{D_1}$ representation for a given basic moments set. To this purpose a method composed of two steps is proposed in [11]. In the first step a non-Markovian matrix representation is generated and in the second step an equivalent Markovian representation is found as a result of an optimisation procedure.

On the contrary to the above mentioned step, the proposed technique contains a step that cannot be performed, according to our current knowledge, based on the $\mathbf{D_0}$, $\mathbf{D_1}$ representation. This step is the reduction of the size of the output MAP which will be performed by simple truncation of the basic moments set defined in (9) and (8).

3. Summary on MAP based queueing network approximations

3.1. MAP/MAP/1 queues

In a MAP/MAP/1 queue the arrivals of customers is given by a MAP with matrices $\mathbf{B_0}$ and $\mathbf{B_1}$ meaning that the series of inter-arrival times are correlated. Also the service of the customers is described by a MAP with matrices denoted by $\mathbf{S_0}$ and $\mathbf{S_1}$. Thus, consecutive service times are correlated as well.

The generator of the CTMC that models the queue has the following block-tridiagonal structure:

$$\mathbf{Q} = \begin{bmatrix} \bar{\mathbf{A}}_\mathbf{0} & \mathbf{A_1} & & & \\ \mathbf{A_{-1}} & \mathbf{A_0} & \mathbf{A_1} & & \\ & \mathbf{A_{-1}} & \mathbf{A_0} & \mathbf{A_1} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}, \tag{13}$$

where the matrix blocks are given by the following Kronecker operations:

$$\begin{aligned} \mathbf{A_1} &= \mathbf{B_1} \otimes \mathbf{I}, \\ \mathbf{A_0} &= \mathbf{B_0} \oplus \mathbf{S_0}, \\ \mathbf{A_{-1}} &= \mathbf{I} \otimes \mathbf{S_1}, \\ \overline{\mathbf{A}}_\mathbf{0} &= \mathbf{B_0} \otimes \mathbf{I}, \end{aligned} \tag{14}$$

where $\mathbf{I}$ denotes the identity matrix of appropriate dimension. Before the Kronecker operator the size of the identity matrix is identical with the size of $\mathbf{B_0}$ and after the Kronecker operator the size of the identity matrix is identical with the size of $\mathbf{S_0}$.

A CTMC with generator matrix of block structure given in (13) is called *Quasi-Birth–Death* process (QBD). Solution methods exploiting the special structure of QBDs have an extensive literature (see, e.g., [13] for a recent survey). In order to compute the performance measures of interest and to analyse the departure process of a MAP/MAP/1 queue, it is necessary to compute the steady state probability vector υ. The steady state vector is partitioned as

$$\upsilon = \begin{bmatrix} \upsilon_0 & \upsilon_1 & \upsilon_2 & \cdots \end{bmatrix}, \tag{15}$$

where also υ_i is vector according to the block structure of the generator. Thus, the jth element of vector υ_i is the probability that there are i jobs in the queue and the background process (that is the product space of the background processes of the arrival and service process) is in state j.

A fundamental result of the matrix analytic methods is that the steady state distribution of QBDs is a matrix geometric distribution, thus

$$\upsilon_k = \upsilon_0 \mathbf{R}^k, \quad k > 0. \tag{16}$$

From the balance equations it follows that $\mathbf{R}$ is the minimal non-negative solution of the following matrix equation:

$$\mathbf{A_1} + \mathbf{R}\mathbf{A_0} + \mathbf{R}^2\mathbf{A_{-1}} = \mathbf{0}. \tag{17}$$

There are several efficient numerical algorithms to compute $\mathbf{R}$ [13,5]. The υ_0 part of the probability vector is the solution of the following set of linear equations:

$$\begin{aligned} 0 &= \upsilon_0 \overline{\mathbf{A}}_\mathbf{0} + \upsilon_1 \mathbf{A_{-1}} = \upsilon_0(\overline{\mathbf{A}}_\mathbf{0} + \mathbf{R}\mathbf{A_{-1}}), \\ 1 &= \sum_{k=0}^{\infty} \upsilon_0 \mathbf{R}^k \mathbf{1} = \upsilon_0 (\mathbf{I} - \mathbf{R})^{-1}\mathbf{1}. \end{aligned} \tag{18}$$

The simplicity of the matrix geometric distribution of the steady state probability vector allows a simple computation for many performance measures. E.g., the mean queue length can be computed as

$$E(N) = \sum_{k=0}^{\infty} k \upsilon_0 \mathbf{R}^k \mathbf{1} = \upsilon_0 \mathbf{R}(\mathbf{I} - \mathbf{R})^{-2}\mathbf{1}. \tag{19}$$

The steady state distribution embedded just after the departures is computed by

$$\upsilon_i^{(D)} = \frac{\upsilon_{i+1}\mathbf{A_{-1}}}{\sum_{k=1}^{\infty} \upsilon_k \mathbf{A_{-1}} \mathbf{1}} = \frac{1}{\lambda} \upsilon_{i+1} \mathbf{A_{-1}}, \quad i \geq 0, \tag{20}$$

where λ denotes the stationary intensity of the arrival MAP ($\mathbf{B_0}$, $\mathbf{B_1}$) according to (1). The denominator of (20) represents the mean departure intensity which equals to the mean arrival intensity, λ_B, when the queue is stable. The departure MAP is active while there is at least one customer in the queue and "gets frozen" when the queue is idle.

3.2. MAP models for the departure process

The exact departure process of a MAP/MAP/1 can be given by a MAP with infinitely many phases. The background Markov chain of the MAP is the Markov chain of the queueing model (see (13)). In this background process the backward level transitions correspond to the departure of a job. Thus the two matrices characterising the departure process exactly are as follows:

$$\mathbf{D_0^{(\infty)}} = \begin{bmatrix} \overline{\mathbf{A}}_\mathbf{0} & \mathbf{A_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{A_0} & \mathbf{A_1} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{A_0} & \mathbf{A_1} & \mathbf{0} & \cdots \\ \vdots & \vdots & & \ddots & \ddots & \end{bmatrix}, \qquad \mathbf{D_1^{(\infty)}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \\ \mathbf{A_{-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \\ \mathbf{0} & \mathbf{A_{-1}} & \mathbf{0} & \mathbf{0} & \cdots & \\ \vdots & \vdots & \ddots & & & \end{bmatrix}. \tag{21}$$

In [14] an approximation method is proposed for the departure process of MAP/PH/1 queues that is based on the appropriate truncation of the exact infinite MAP. Recently, two results have been published that are based on the same idea but can be applied to MAP/MAP/1 queues as well. Both of them truncate the infinite MAP at level n, but in different ways. The structure of the approximating departure process is the same

$$\mathbf{D_0^{(n)}} = \begin{bmatrix} \overline{\mathbf{A}}_\mathbf{0} & \mathbf{A_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{A_0} & \mathbf{A_1} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{A_0} & \mathbf{A_1} & \mathbf{0} & \cdots \\ \vdots & \vdots & & \ddots & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A_0} & \mathbf{A_1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \widehat{\mathbf{A}}_\mathbf{0} \end{bmatrix}, \quad \mathbf{D_1^{(n)}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{A_{-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{A_{-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \vdots & \vdots & \ddots & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A_{-1}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \widehat{\mathbf{A}}_\mathbf{-1} & \widetilde{\mathbf{A}}_\mathbf{-1} \end{bmatrix}, \tag{22}$$

only the definition of the special matrix blocks, $\widehat{\mathbf{A}}_{\mathbf{-1}}$ and $\widetilde{\mathbf{A}}_{\mathbf{-1}}$ differs.

The next two subsection provide a short overview on these methods.

3.3. Level probability based truncation method

The basic idea of the truncation method of [8] is that all the levels $i \geq n$ of the exact model are merged into the last level (referred to as the clipping level) of the truncated model. All the forward and local transitions of the infinite MAP correspond to local transitions in the truncated MAP, that gives $\widehat{\mathbf{A}}_\mathbf{0} = \mathbf{A_1} + \mathbf{A_0}$.

However, in case of departures there are two cases when the truncated model is at the clipping level. According to [8], the probability that the exact model is at level $i = n$ when the truncated process is at clipping level n is approximated using vector v_n and probability that the exact model is at level $i > n$ when the truncated process is at clipping level n is approximated using vector $v_n^+ = \sum_{k=n+1}^{\infty} v_k$. Indeed, [8] approximates the probability that a departure of the truncated process at clipping level n and phase j moves the truncated process to level $n-1$ as $[v_n]_j/([v_n^+]_j + [v_n]_j)$, where $[v_n]_j$ denotes the jth element of v_n. Thus the related blocks of the truncated MAP are the following:

$$\begin{aligned} \widehat{\mathbf{A}}_\mathbf{-1} &= \mathbf{Diag}\langle v_n \rangle \mathbf{Diag}^{-1}\langle v_n + v_n^+ \rangle \mathbf{A_{-1}}, \\ \widetilde{\mathbf{A}}_\mathbf{-1} &= \mathbf{Diag}\langle v_n^+ \rangle \mathbf{Diag}^{-1}\langle v_n + v_n^+ \rangle \mathbf{A_{-1}}, \end{aligned} \tag{23}$$

where $\mathbf{Diag}\langle vec \rangle$ denotes the diagonal matrix composed by the elements of vector vec. Since

$$\mathbf{Diag}\langle v_n \rangle \mathbf{Diag}^{-1}\langle v_n + v_n^+ \rangle + \mathbf{Diag}\langle v_n^+ \rangle \mathbf{Diag}^{-1}\langle v_n + v_n^+ \rangle = \mathbf{I},$$

this definition ensures that $\widehat{\mathbf{A}}_\mathbf{-1} + \widetilde{\mathbf{A}}_\mathbf{-1} = \mathbf{A_{-1}}$.

3.4. ETAQA truncation method

The efficiency of the method of [8] has been enhanced in [9]. The authors of [9] refer to their method as ETAQA truncation method. The blocks of the ETAQA truncated model are defined as:

$$\begin{aligned} \widehat{\mathbf{A}}_\mathbf{0} &= \mathbf{A_1} + \mathbf{A_0}, \\ \widehat{\mathbf{A}}_\mathbf{-1} &= \mathbf{A_{-1}} - \mathbf{A_1 G}, \\ \widetilde{\mathbf{A}}_\mathbf{-1} &= \mathbf{A_1 G}, \end{aligned} \tag{24}$$

where $\mathbf{G}$ can be computed from $\mathbf{R}$ using $\mathbf{G} = (-\mathbf{A_0} - \mathbf{RA_{-1}})^{-1}\mathbf{A_{-1}}$. This construction (based on the idea of the ETAQA methodology) ensures that the steady state probabilities of the truncated process $\hat{v}_k$ and of the exact model v_k are the same up to the clipping level, and for the clipping level $\hat{v}_n = \sum_{k=n}^{\infty} v_k$ holds. As a consequence, the inter-departure times and the lag-k correlations up to the truncation level n are preserved exactly.

4. Moments based queueing network approximation

In case of traffic based decomposition of queueing networks (QNs), the main elements of the computation are

- traffic aggregation,
- traffic splitting,
- output process approximation,
- model reduction.

The concrete implementation of these steps depends on the most important decision of the approximation which is the selection of the traffic description of the inter-node traffic. Similar to [9,8] in this paper we also apply MAP to describe the inter-node traffic, but in an essentially different way. In this paper we represent the inter-node traffic with its basic moments set.

The main advantage of this traffic description is that it allows a natural and flexible scaling of the size of the traffic description, i.e., the order of the MAP.

A major disadvantage of pure MAP based inter-node traffic description is that the size of the inter-node MAP model increases node by node during the evaluation and there was no efficient and accurate model reduction method available for keeping the size of the model moderate.

In the following subsections we detail the elementary steps of the analysis together with the proposed, moments based model reduction method. Both the aggregation and the traffic splitting step can be performed based purely either on the basic moments set representation or on the $\mathbf{D_0}$, $\mathbf{D_1}$ representation. We present both approaches for completeness, in spite of the fact that in practical computations we commonly apply the $\mathbf{D_0}$, $\mathbf{D_1}$ representation. According to our current knowledge the step of approximating the output process of a MAP/MAP/1 queue can be done based on the $\mathbf{D_0}$, $\mathbf{D_1}$ representation only. The model reduction step is performed based on the basic moments set representation only.

4.1. Traffic aggregation

The fact that $(\mathbf{F_0}, \mathbf{F_1})$ is the superposition of $(\mathbf{D_0}, \mathbf{D_1})$ and $(\mathbf{E_0}, \mathbf{E_1})$ means that $\lambda_F = \lambda_D + \lambda_E$, $\mathbf{F_0} = \mathbf{D_0} \oplus \mathbf{E_0}$, $\mathbf{F_1} = \mathbf{D_1} \oplus \mathbf{E_1}$, and $\alpha_F = \alpha_D \otimes \alpha_E$ [5].

The moments based description of traffic aggregation is presented in the following theorem.

Theorem 1. *Let $(\mathbf{D_0}, \mathbf{D_1})$ and $(\mathbf{E_0}, \mathbf{E_1})$ be MAPs and $(\mathbf{F_0}, \mathbf{F_1})$ their superposition. Let γ_{ij}^D, γ_{ij}^E and γ_{ij}^F denote the derivatives of the joint probability density function of two consecutive inter-arrival times at 0 as defined in (11), ν_i^D, ν_i^E and ν_i^F the derivatives of the ccdf of the marginal distribution of the inter-arrival times as defined in (10), and λ_D, λ_E and λ_F their average arrival intensity as defined in (1), for the three MAPs. Then the joint density of two consecutive inter-arrival times of the superposed process satisfies*

$$\gamma_{ij}^F = \frac{\lambda_D \lambda_E}{\lambda_D + \lambda_E} \sum_{k=0}^{i} \sum_{\ell=0}^{j+1} \binom{i}{k} \binom{j+1}{\ell} \left(\gamma_{i-k,j-\ell}^D \nu_{k+\ell-1}^E + \nu_{i-k+j+1-\ell}^D \gamma_{k-1,\ell-1}^E + \gamma_{i-k-1,j-\ell}^D \nu_{k+\ell}^E + \nu_{i-k+j-\ell}^D \gamma_{k,\ell-1}^E \right). \quad (25)$$

The proof is provided in Appendix A.

4.2. Traffic splitting

Markovian traffic splitting at the exit of a node of the queueing network means that departing customers are directed to a given consecutive node with probability p. If $(\mathbf{D_0}, \mathbf{D_1})$ represents the departure process then $(\mathbf{D_0} + (1-p)\mathbf{D_1}, p\mathbf{D_1})$ characterises the traffic towards the given consecutive node [5].

The moments based description of traffic splitting is demonstrated here for a low order term, $i = 2$, $j = 1$. The description of the general case requires cumbersome notation which we avoid here.

Let $(\mathbf{D_0}, \mathbf{D_1})$ be a MAP and $(\mathbf{E_0}, \mathbf{E_1}) = (\mathbf{D_0} + (1-p)\mathbf{D_1}, p\mathbf{D_1})$ its split with probability p. Consequently, $\alpha_D = \alpha_E$, since $\mathbf{D} = \mathbf{D_0} + \mathbf{D_1} = \mathbf{E_0} + \mathbf{E_1}$, and $\lambda_E = p\lambda_D$.

$$\gamma_{2,1}^E = \pi_E \mathbf{E_0^2} \mathbf{E_1} \mathbf{E_0^1} \mathbf{E_1} \mathbf{1} = \pi_D (\mathbf{D_0} + (1-p)\mathbf{D_1})^2 p \mathbf{D_1} (\mathbf{D_0} + (1-p)\mathbf{D_1})^1 p \mathbf{D_1} \mathbf{1},$$

where

$$(\mathbf{D_0} + (1-p)\mathbf{D_1})^2 = \mathbf{D_0^2} + \mathbf{D_0}((1-p)\mathbf{D_1}) + ((1-p)\mathbf{D_1})\mathbf{D_0} + ((1-p)\mathbf{D_1})^2$$

and

$$\begin{aligned} \pi_E &= \frac{1}{\lambda_E} \alpha_E (-\mathbf{E_0}) = \frac{1}{p\lambda_D} \alpha_D (-\mathbf{D_0} - (1-p)\mathbf{D_1}) \\ &= \frac{1}{p\lambda_D} \lambda_D \pi_D (-\mathbf{D_0})^{-1} (-\mathbf{D_0} - (1-p)\mathbf{D_1}) \\ &= \pi_D \left(\frac{1}{p} \mathbf{I} - \frac{1-p}{p} (-\mathbf{D_0})^{-1} \mathbf{D_1} \right) = \pi_D, \end{aligned} \quad (26)$$

since $\pi_D(-\mathbf{D_0})^{-1}\mathbf{D_1} = \pi_D$. Using these

$$\begin{aligned}
\gamma^E_{2,1} &= \pi_D(\mathbf{D_0} + (1-p)\mathbf{D_1})^2 p\mathbf{D_1}(\mathbf{D_0} + (1-p)\mathbf{D_1})^1 p\mathbf{D_1}\mathbf{1} \\
&= \pi_D(\mathbf{D_0})^2 p\mathbf{D_1}\mathbf{D_0}p\mathbf{D_1}\mathbf{1} + \pi_D\mathbf{D_0}(1-p)\mathbf{D_1}p\mathbf{D_1}\mathbf{D_0}p\mathbf{D_1}\mathbf{1} + \pi_D(1-p)\mathbf{D_1}\mathbf{D_0}p\mathbf{D_1}\mathbf{D_0}p\mathbf{D_1}\mathbf{1} \\
&\quad + \pi_D(1-p)^2(\mathbf{D_1})^2 p\mathbf{D_1}\mathbf{D_0}p\mathbf{D_1}\mathbf{1} + \pi_D(\mathbf{D_0})^2 p\mathbf{D_1}(1-p)\mathbf{D_1}p\mathbf{D_1}\mathbf{1} \\
&\quad + \pi_D\mathbf{D_0}(1-p)\mathbf{D_1}p\mathbf{D_1}(1-p)\mathbf{D_1}p\mathbf{D_1}\mathbf{1} + \pi_D(1-p)\mathbf{D_1}\mathbf{D_0}p\mathbf{D_1}(1-p)\mathbf{D_1}p\mathbf{D_1}\mathbf{1} \\
&\quad + \pi_D(1-p)^2(\mathbf{D_1})^2 p\mathbf{D_1}(1-p)\mathbf{D_1}p\mathbf{D_1}\mathbf{1} \\
&= p^2\gamma^D_{2,1} + p^2(1-p)\gamma^D_{1,0,1} + p^2(1-p)\gamma^D_{0,1,1} + p^2(1-p)^2\gamma^D_{0,0,0,1} \\
&\quad + p^2(1-p)\gamma^D_{2,0,0} + p^2(1-p)^2\gamma^D_{1,0,0,0} + p^2(1-p)^2\gamma^D_{0,1,0,0} + p^2(1-p)^3\gamma^D_{0,0,0,0,0}.
\end{aligned}$$

This example demonstrates that the derivatives of the joint densities of the split process can be computed from the ones of the original process without knowing its $\mathbf{D_0}$, $\mathbf{D_1}$ representation.

4.3. Output process approximation

The moments based description of the departure process differs significantly from the approximation approaches described in Sections 3.3 and 3.4. Those techniques construct an approximate departure process directly based on the behaviour of the MAP/MAP/1 queue. Our approach instead is first to compute dominant parameters of the departure process, namely the joint moments of the consecutive inter-departure times, and then to construct a MAP that realizes these parameters.

To describe the moments of the departure process we need the following notations. The row vector $b^{(D)}_k$ ($s^{(D)}_k$) denotes the phase distribution of the arrival (departure) MAP after a departure which left k customers in the system. The ith element of $b^{(D)}_k$ and $s^{(D)}_k$ are

$$[b^{(D)}_k]_i = v^{(D)}_k(e^T_i \otimes \mathbf{1}) \quad \text{and} \quad [s^{(D)}_k]_i = v^{(D)}_k(\mathbf{1} \otimes e^T_i). \tag{27}$$

e_i is the row vector whose ith element is one and the others are zero. Matrix $\mathbf{U^{(D)}_0}$ of size $n \times m$ is composed by the elements of vector $v^{(D)}_0$, such that $[\mathbf{U^{(D)}_0}]_{i,j} = [v^{(D)}_0]_{(i-1)n+j}$, i.e., $[\mathbf{U^{(D)}_0}]_{i,j}$ is the probability that a departure leaves the MAP/MAP/1 queue empty, the phase of the arrival MAP is i and the phase of the departure MAP is j. Further more $b_0 = -\mathbf{B_0}\mathbf{1}$ and $s_0 = -\mathbf{S_0}\mathbf{1}$.

Theorem 2. *The stationary inter-departure time of a MAP/MAP/1 queue has a matrix exponential representation of order $n+m$ with initial vector u, generator $\mathbf{M}$ and closing vector v. That is, the CDF of the inter-departure time distribution is $1 - ue^{\mathbf{M}t}v$, where*

$$u = \begin{bmatrix} b^T_0 & s^{(D)}_{1+} \end{bmatrix}, \tag{28}$$

$$\mathbf{M} = \begin{bmatrix} \mathbf{B}^T_\mathbf{0} & \mathbf{U^{(D)}_0} \\ \mathbf{0} & \mathbf{S_0} \end{bmatrix}, \tag{29}$$

$$v = \begin{bmatrix} (-\mathbf{B}^T_\mathbf{0})^{-1}\mathbf{U^{(D)}_0}\mathbf{1} \\ \mathbf{1} \end{bmatrix}, \tag{30}$$

where the subscript $1+$ refers to the cases when there is at least on customer in the system

$$v^{(D)}_{1+} = \sum_{k=1}^{\infty} v^{(D)}_k = \frac{1}{\lambda} v_0 \mathbf{R}^2(\mathbf{I} - \mathbf{R})^{-1}\mathbf{A}_{-1}, \tag{31}$$

and $s^{(D)}_{1+}$ is obtained according to (27).

Proof. If a departure leaves the MAP/MAP/1 queue busy such that the phase of the service MAP is j then the time to the next departure is phase type distributed time with initial vector e_j and generator $\mathbf{S_0}$.

If a departure leaves the MAP/MAP/1 queue empty such that the phase of the arrival MAP is i and service MAP is j then the time to the next departure is the sum of two phase type distributed times, the first one with initial vector e_i and generator $\mathbf{B_0}$ and the second one with initial vector e_j and generator $\mathbf{S_0}$.

The Laplace transform of the stationary inter-departure time, T_D, is

$$E(\mathrm{e}^{-sT_D}) = \sum_{j=1}^{m}[s^{(D)}_{1+}]_j e_j(s\mathbf{I} - \mathbf{S_0})^{-1}s_0 + \sum_{i=1}^{m}\sum_{j=1}^{m}[\mathbf{U^{(D)}_0}]_{i,j} e_i(s\mathbf{I} - \mathbf{B_0})^{-1}b_0 e_j(s\mathbf{I} - \mathbf{S_0})^{-1}s_0$$

where

$$e_i(s\mathbf{I} - \mathbf{B_0})^{-1}b_0 = b^T_0((s\mathbf{I} - \mathbf{B_0})^{-1})^T e^T_i = b^T_0(s\mathbf{I} - \mathbf{B}^T_\mathbf{0})^{-1}e^T_i.$$

Using this we have

$$E(e^{-sT_D}) = s_{1+}^{(D)}(s\mathbf{I} - \mathbf{S_0})^{-1}s_0 + b_0^T(s\mathbf{I} - \mathbf{B_0^T})^{-1}\mathbf{U_0^{(D)}}(s\mathbf{I} - \mathbf{S_0})^{-1}s_0.$$

Partitioning the Laplace transform of the matrix exponential distribution with representation u, $\mathbf{M}$, v we have

$$\begin{aligned} u(s\mathbf{I} - \mathbf{M})^{-1}(-\mathbf{M})v &= \begin{bmatrix} b_0^T & s_{1+}^{(D)} \end{bmatrix} \begin{bmatrix} s\mathbf{I} - \mathbf{B_0^T} & -\mathbf{U_0^{(D)}} \\ \mathbf{0} & s\mathbf{I} - \mathbf{S_0} \end{bmatrix}^{-1} \begin{bmatrix} -\mathbf{B_0^T} & -\mathbf{U_0^{(D)}} \\ \mathbf{0} & -\mathbf{S_0} \end{bmatrix} \begin{bmatrix} (-\mathbf{B_0^T})^{-1}\mathbf{U_0^{(D)}}\mathbf{1} \\ \mathbf{1} \end{bmatrix} \\ &= \begin{bmatrix} b_0^T & s_{1+}^{(D)} \end{bmatrix} \begin{bmatrix} (s\mathbf{I} - \mathbf{B_0^T})^{-1} & (s\mathbf{I} - \mathbf{B_0^T})^{-1}\mathbf{U_0^{(D)}}(s\mathbf{I} - \mathbf{S_0})^{-1} \\ \mathbf{0} & (s\mathbf{I} - \mathbf{S_0})^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ s_0 \end{bmatrix} \\ &= b_0^T(s\mathbf{I} - \mathbf{B_0^T})^{-1}\mathbf{U_0^{(D)}}(s\mathbf{I} - \mathbf{S_0})^{-1}s_0 + s_{1+}^{(D)}(s\mathbf{I} - \mathbf{S_0})^{-1}s_0. \quad \square \end{aligned}$$

Corollary 1. *When in a MAP/MAP/1 queue the order of the arrival MAP is n and that of the service MAP is m, then the order of the phase type distributed inter-departure time distribution is at most $n + m$ and consequently the number of independent inter-departure time moments is at most $2(m + n) - 1$.*

Proof. The corollary is a straight forward consequence of Theorem 2. □

Theorem 3. *The stationary joint moments of two consecutive inter-departure of a MAP/MAP/1 queue can be computed as*

$$E(X_0^i X_1^j) = zi!(-\mathbf{M_0})^{-i-1}\mathbf{M_1}j!(-\mathbf{M_0})^{-j}\mathbf{1}, \tag{32}$$

where

$$z = \begin{bmatrix} v_0^{(D)} & v_1^{(D)} & v_{2+}^{(D)} \end{bmatrix}, \tag{33}$$

$$v_{2+}^{(D)} = \sum_{k=2}^{\infty} v_k^{(D)} = \frac{1}{\lambda} v_0 \mathbf{R}^3(\mathbf{I} - \mathbf{R})^{-1}\mathbf{A_{-1}}, \tag{34}$$

$$\mathbf{M_0} = \begin{bmatrix} \bar{\mathbf{A}}_\mathbf{0} & \mathbf{A_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A_0} & \mathbf{A_1} \\ \mathbf{0} & \mathbf{0} & \mathbf{A_0} + \mathbf{A_1} \end{bmatrix}, \tag{35}$$

$$\mathbf{M_1} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{A_{-1}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A_{-1}} & \mathbf{0} \end{bmatrix}. \tag{36}$$

Proof. Since we focus on the joint moments of two consecutive inter-departure times we have to consider the following three cases:

- a departure leaves the queue empty, with probability $v_0^{(D)}$;
- a departure leaves one customer in the queue, with probability $v_1^{(D)}$;
- a departure leaves at least two customers in the queue, with probability $v_{2+}^{(D)}$.

For all the three cases, the computation of the joint moments of inter-departure times is based on constructing the MAP that generates the departures and then computing the joint moments based on (6).

The process evolution up to the second departure is different in the three cases. Let us first consider the third case which is the simplest. If there are at least two customers in the queue at a departure, then the queue does not become empty before the next two departures. For this reason the joint moments of the next two inter-departure times do not depend on the arrivals. Consequently, in this case, it is enough to consider the state transitions which are assigned with a departure, $\mathbf{A_{-1}}$, and the ones which are not, $\mathbf{A_0} + \mathbf{A_1}$. As a result, in this case the joint moments can be computed as

$$\begin{aligned} E(X_0^i X_1^j I_{\{N(0)\geq 2\}}) &= v_{2+}^{(D)} i!(-\mathbf{A_0} - \mathbf{A_1})^{-i}(-\mathbf{A_0} - \mathbf{A_1})^{-1}\mathbf{A_{-1}}j!(-\mathbf{A_0} - \mathbf{A_1})^{-j}\mathbf{1} \\ &= v_{2+}^{(D)} i!(-\mathbf{A_0} - \mathbf{A_1})^{-i-1}\mathbf{A_{-1}}j!(-\mathbf{A_0} - \mathbf{A_1})^{-j}\mathbf{1}, \end{aligned} \tag{37}$$

where $N(t)$ denotes the number of customers at time t, we assume that a departure occurred at $t = 0$ and $I_{\{A\}}$ equals to one when A is true and to zero otherwise. In the second case, i.e., when a departure leaves one customer in the queue, we need to take into consideration one arrival as well in order to compute the joint moments of the next two inter-departure times. This arrival can happen either before or after the first departure and is taken into account by the block $\mathbf{A_1}$ in position (2, 3) of $\mathbf{M_0}$ in (35).

Since in the third case the queue is left empty, for the calculation of the joint moments of the next two inter-departure times we have to consider two arrivals. The first happens before the first departure and is taken into account by the block

$\mathbf{A_1}$ in position (1, 2) of $\mathbf{M_0}$ in (35). The second arrival can happen either before the first departure or after the first departure and is considered the same way as the arrival in the second case.

The three cases can be organised in a single compact form as presented in (32)–(36). □

Note that also the moments of the inter-departure times can be computed based on Theorem 3 by setting j to 0 in (32). Having computed the moments and joint moments of the departure process of a queue, we apply the method described in [11] to construct a MAP with such parameters and use this MAP as approximation of the output process.

It is important to note that

- the MAP defined by $\mathbf{M_0}$ and $\mathbf{M_1}$ is not a good output process model of the MAP/MAP/1 queue,
- the embedded stationary distribution of the MAP defined by $\mathbf{M_0}$ and $\mathbf{M_1}$ is different from z,
- the finite dimensional matrix expression in (32) is exact, because vector z represents the effect of the infinite queue.

4.4. Maximal order

We can compute any moments and lag-1 joint moments of the MAP/MAP/1 queue's departure process based on (32). Having these moments we can compose the basic moment set for various order and apply the procedure to generate Markovian representation from the basic moments set [11]. This procedure requires basic moments set composed by independent moments which limits its applicability. According to Corollary 1 the maximal order of the basic moments set is sum of the orders of the arrival and departure MAPs. Theoretically, the maximal order can be lower, e.g., if (28) and (29) is a redundant representation of a lower order phase type distribution, but it rarely occurs in practice.

4.5. Model reduction

The applied model reduction is based on the natural assumption that the lower moments carry more information on the traffic behaviour than the higher ones. Consequently, the moments based model reduction is a very natural procedure. It is simply dropping the higher moments and joint moments from the basic moments set. Namely, starting from the order n basic moments set, $\mu_i, i = 1, \ldots, 2n-1$, and $\eta_{ij}, i, j = 1, \ldots, n-1$, the reduced traffic description is the order $k < n$ basic moments set, $\mu_i, i = 1, \ldots, 2k-1$, and $\eta_{ij}, i, j = 1, \ldots, k-1$.

4.6. Approximating the output process by a MAP(2)

The moment set of the output traffic can be such that the moment matching algorithm presented in [11] fails. In these cases we modify the moment set to make it feasible for a MAP. The only class of MAPs for which symbolic moment and correlation bounds are known is the set of order 2 MAPs (MAP(2)). These moment and correlation bounds are provided in [15]. A MAP(2) is determined by four parameters, μ_1, μ_2, μ_3 and ρ_1, and we modify these four parameters in order to make them feasible for a MAP(2) according to the following procedure.

1. Adjust μ_2 into the interval $\left[\frac{3}{2}\mu_1^2, \infty\right]$,
2. adjust μ_3 into the feasible interval:

$$\begin{cases} \mu_3 \in \left[9\mu_1\mu_2 - 12\mu_1^3 + 3\sqrt{2}(2\mu_1^2 - \mu_2)^{3/2}, 6\mu_1\mu_2 - 6\mu_1^3\right], & \text{if } \dfrac{3}{2}\mu_1^2 \le \mu_2 \le 2\mu_1^2 \\ \mu_3 \in \left[\dfrac{3}{2}\dfrac{\mu_2^2}{\mu_1}, \infty\right], & \text{if } \mu_2 > 2\mu_1^2 \end{cases}$$

3. adjust the correlation ρ_1 according to the bounds given in [15].

4.7. Steps of a network queue analysis

The proposed analysis of a network node is composed by the following steps:

1. *Traffic aggregation of input MAPs*
   $\mathbf{F_0} = \mathbf{D_0} \oplus \mathbf{E_0}$, $\mathbf{F_1} = \mathbf{D_1} \oplus \mathbf{E_1}$.
2. *MAP/MAP/1 queue analysis*
   Computation of the required performance measures.
3. *Moments analysis of the MAP/MAP/1 output process*
   Based on Theorem 3.
4. *Output process truncation and matrix representation*
   Computation of a MAP representation of the output process based on its moments and joint moments [11].
5. *Splitting of the approximate MAP output process*
   $\mathbf{E_0} = \mathbf{D_0} + (1-p)\mathbf{D_1}$, $\mathbf{E_1} = p\mathbf{D_1}$.

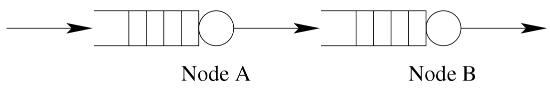

Fig. 1. Tandem network used in Example 1.

5. Numerical examples

5.1. Tandem networks

In this section the presented joint moments based MAP/MAP/1 queueing network analysis method is evaluated on the three tandem queueing network examples provided in [9]. The basic setup is depicted by Fig. 1. The service times at Node B are given by an Erlang-2 distribution with mean 1.25. Three different arrival and service MAPs are defined for Node A as follows.

- Case (a)
  The service times at Node A are exponentially distributed with mean 1, and the MAP generating the arrivals is given by the following matrices:

$$\mathbf{D_0^a} = \begin{bmatrix} -6.9375 & 0.9375 \\ 0.0625 & -0.1958 \end{bmatrix}, \qquad \mathbf{D_1^a} = \begin{bmatrix} 6 & 0 \\ 0 & 0.1333 \end{bmatrix}.$$

  The arrival intensity, squared coefficient of variation and lag-1 correlation coefficient of this MAP are: $\lambda = 0.5$, $c_v^2 = 4.1$, $\rho_1 = 0.23$.
- Case (b)
  The properties of the arrival process of Node A are $\lambda = 0.5$, $c_v^2 = 18.86$, $\rho_1 = 0.34$, it is characterised by the following matrices:

$$\mathbf{D_0^b} = \begin{bmatrix} -0.542409519 & 0.0037279 & 0 \\ 0.004349217 & -0.02298872 & 0.000621317 \\ 0 & 0.001242633 & -2.269670072 \end{bmatrix},$$

$$\mathbf{D_1^b} = \begin{bmatrix} 0.020503453 & 0 & 0.518178166 \\ 0 & 0.017396869 & 0.000621317 \\ 2.259107688 & 0.004970534 & 0.004349217 \end{bmatrix}.$$

  The service time of Node A is hyperexponentially distributed with a mean of 1 and $c_v^2 = 2.62$, thus the matrices of the service MAP are:

$$\mathbf{S_0^{\mathit{b}}} = \begin{bmatrix} -10 & 0 \\ 0 & -0.52632 \end{bmatrix}, \qquad \mathbf{S_1^b} = \begin{bmatrix} 5 & 5 \\ 0.26316 & 0.26316 \end{bmatrix}.$$

- Case (c)
  The arrival MAP is the same as in case (b), but the service times are correlated ($\rho = -0.31$), given by the following MAP:

$$\mathbf{S_0^c} = \begin{bmatrix} -10 & 0 \\ 0 & -0.52632 \end{bmatrix}, \qquad \mathbf{S_1^c} = \begin{bmatrix} 0 & 10 \\ 0.52632 & 0 \end{bmatrix}.$$

The method presented in this paper is compared to the ones presented in [8,9] and summarised in Sections 3.3 and 3.4.

First the mean queue length of Node B is investigated using different output approximation methods and different truncation levels for Node A. The results are summarised in Table 1. The accuracy of the two truncation based methods increases with increasing truncation level. However, with these methods the order of the MAP representing the departure traffic of Node A is much larger than the one of the moments based representation. The traffic approximation with large MAPs has two negative consequences:

- It slows down (or makes infeasible) the analysis of Node B. When the truncation is at level $n = 20$ the computation of the mean queue length of Node B took about 1 min, as opposed to the prompt results obtained with MAPs with $n \leq 5$ phases.
- It does not scale well. If we have a larger queueing network with more than just 2 nodes, the size of the departure MAP grows exponentially with the number of hops. With the truncation methods it becomes impossible to analyse a network composed by three tandem nodes if the clipping level is $n > 10$.

As reflected by the results in Table 1, our MAP approximation of the departure process results in a compact MAP having only a few (2 or 3) states, and even with 2 states we get reasonably accurate results. In Case (a) the moments based approximation with two states is more accurate than the truncation methods with 12 states.

We need to mention that with our current approach the output process of Node A can be approximated only with MAP(2) and MAP(3) because the moments and joint moments of the departure process of Node A are such that there is no MAP(4)

Table 1
Mean queue length on Node B in Example 1.

	#States	Case (a)	#States	Case (b)	Case (c)
Simulation	n/a	0.9517	n/a	3.48825	3.08063
Moments $n = 2$	2	0.93967	2	2.5053	2.55597
Based $n = 3$	3	0.954241	3	3.48803	3.01978
ETAQA $n = 2$	6	0.833259	18	2.58742	2.61587
$n = 5$	12	0.900164	36	2.91293	2.73691
$n = 10$	22	0.936189	66	3.20054	2.95097
$n = 20$	42	0.949793	126	3.41015	3.04765
Level $n = 2$	6	0.902632	18	3.52804	3.05992
Prob. $n = 5$	12	0.939841	36	3.53408	3.08245
Based $n = 10$	22	0.947761	66	3.5002	3.0771
$n = 20$	42	0.951109	126	3.4889	3.07611

whose basic moments set is identical to the one of the departure process. In this paper we restrict our attention to the cases when the basic moments set of the output process is feasible for MAPs of a given order (i.e., the moments matching procedure of [11] is applicable). If it is not the case, then the same moments based approximation could be applied together with a MAP fitting method (which finds a valid MAP whose basic moments set is as close to the one of the departure process as possible). This possibility is out of the scope of this paper (mainly because our current MAP fitting procedures are not stable enough yet). In the consecutive examples we use only MAP(2) and MAP(3) approximations due to the same reason. It is important to note that this is not a limitation of the moments based approximation approach. The moments based approximation approach is applicable with any order MAPs if a stable MAP fitting procedure provides the valid MAP representation of the basic moments set.

Another important advantage of the moments based approximation method is that the model size does not grow with the number of nodes of the network. We can apply arbitrary compact description for the output process of all nodes. Thus, moments based approximation procedure does not have scaling problems due to state space explosion.

Fig. 2 depicts the queue length distribution of Node B and the autocorrelation of the internal traffic between Node A and Node B. As expected, by increasing the number of states more statistical quantities of the traffic are matched and therefore the accuracy of the approximation improves. In these examples 3 phases are enough to capture the shape of the autocorrelation function. Fig. 2 presents the autocorrelation for low order lags, but the MAP representation of the output process makes it very simple to obtain also the asymptotic decay rate of the autocorrelation function, since it is the real part of the subdominant eigenvalue of $\mathbf{P} = (-\mathbf{D_0})^{-1}\mathbf{D_1}$, where $\mathbf{D_0}$ and $\mathbf{D_1}$ are the matrices of the moments based MAP approximation of the departure process of Node A.

5.2. A three-node network with superposition

As a second example we consider a simple network composed by three nodes as depicted in Fig. 3.
The MAPs of the arrival and service at the nodes are as follows.

- At Node A the arrival process is given by

$$\mathbf{D_0^A} = \begin{bmatrix} -25 & 3 & 10 \\ 1 & -6 & 0 \\ 0 & 4 & -10 \end{bmatrix}, \qquad \mathbf{D_1^A} = \begin{bmatrix} 10 & 0 & 2 \\ 2 & 3 & 0 \\ 5 & 0 & 1 \end{bmatrix},$$

  with $\lambda = 6.63,\ c_v^2 = 1.31,\ \rho_1 = 0.027$, the service MAP is defined by

$$\mathbf{S_0^A} = \begin{bmatrix} -30 & 12 \\ 0 & -9 \end{bmatrix}, \qquad \mathbf{S_1^A} = \begin{bmatrix} 15 & 3 \\ 2 & 7 \end{bmatrix},$$

  with $\lambda = 10,\ c_v^2 = 1.16,\ \rho_1 = 0.025$.
- The arrival and service MAPs at Node B are

$$\mathbf{D_0^B} = \begin{bmatrix} -60 & 10 \\ 1 & -5 \end{bmatrix}, \qquad \mathbf{D_1^B} = \begin{bmatrix} 50 & 0 \\ 0 & 4 \end{bmatrix},$$

$$\mathbf{S_0^B} = \begin{bmatrix} -80 & 40 \\ 6 & -20 \end{bmatrix}, \qquad \mathbf{S_1^B} = \begin{bmatrix} 20 & 20 \\ 7 & 7 \end{bmatrix},$$

  with the arrival process having properties of $\lambda = 8.18,\ c_v^2 = 2.2,\ \rho_1 = 0.19$ and the basic properties of the service MAP are $\lambda = 18.63,\ c_v^2 = 1.23,\ \rho_1 = 0$.
- The MAP describing the service process of Node C is given by

$$\mathbf{S_0^C} = \begin{bmatrix} -100 & 10 \\ 1 & -16 \end{bmatrix}, \qquad \mathbf{S_1^C} = \begin{bmatrix} 80 & 10 \\ 1 & 14 \end{bmatrix}.$$

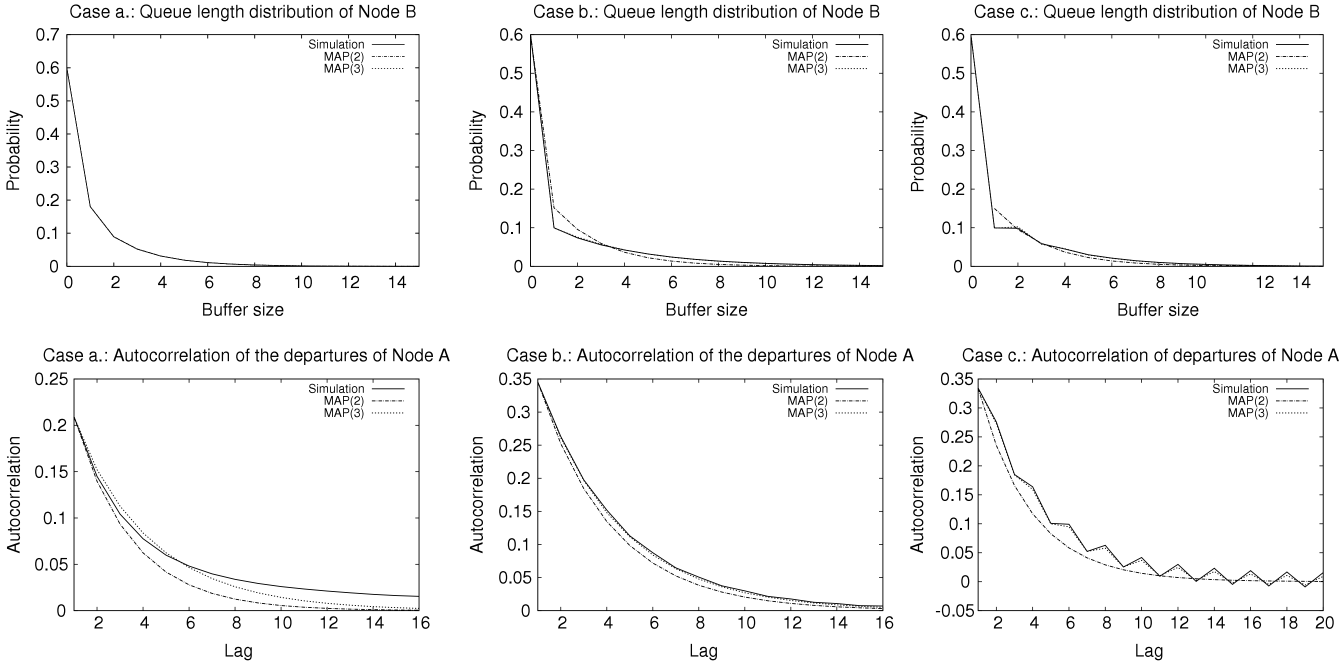

Fig. 2. Queue length distribution and autocorrelation in Example 1.

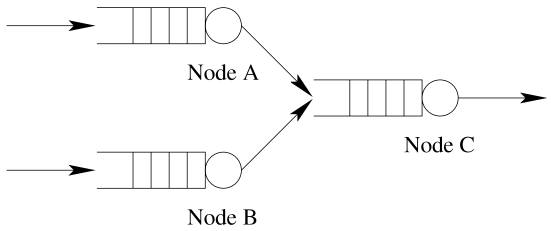

Fig. 3. The queueing network used in Example 2.

Table 2
Mean queue length of Node C in Example 2.

Node A, B output	MAP(2)	MAP(3)
Simulation	4.63527	
Compressed $n = 3$	4.313268 (−7%)	4.06334 (−12%)
Aggregate $n = 5$	n/a	4.23841 (−9%)
$n = 7$	n/a	4.31843 (−7%)
Non-compressed		
Aggregate ($n = 4, 9$)	4.32768 (−6.5%)	4.44595 (−4%)
Renewal output approx.	3.5791 (−22.8%)	3.5847 (−22.6%)

The basic properties of this MAP are $\lambda = 21.8,\ c_v^2 = 1.58,\ \rho_1 = 0.13$.

The performance measure of interest is the same as before, the mean queue length at Node C. Unfortunately our trial to compare our results with the ones of the truncation based methods failed because we were not able to perform the analysis even at the lowest possible truncation level $n = 2$ (which results 432 phases at Node C) due to infeasible computation time. Our Mathematica implementation did not provide sufficiently precise result with normal precision (16 digits) and did not terminate in an hour with extended precision (100 digits). Mathematica allows to evaluate the precision of the results and we needed 100 digits precision at the input to obtain reasonable (>10 digits) precision of the result.

For the superposition of the output traffic of Node A and Node B we used both the direct method based on Kronecker algebra and the moments based superposition method of Theorem 1 obtaining the same results.

The results of the moments based approximation are summarised in Table 2. The header indicates how many phases have been used to approximate the output of Node A and B. "Compression" refers to the number of phases the superposed MAP is compressed to ("n/a" indicates that the field has no meaning, e.g., compression of the superposed traffic to 5 states is not possible when a MAP(2) approximation is used since the superposed traffic has only 4 states in this case).

According to the expectations the results are more accurate when MAPs(3) are used for the output process approximation of Node A and B. The compression decreases the accuracy of the approximation. Surprisingly, the 2-state output approxi-

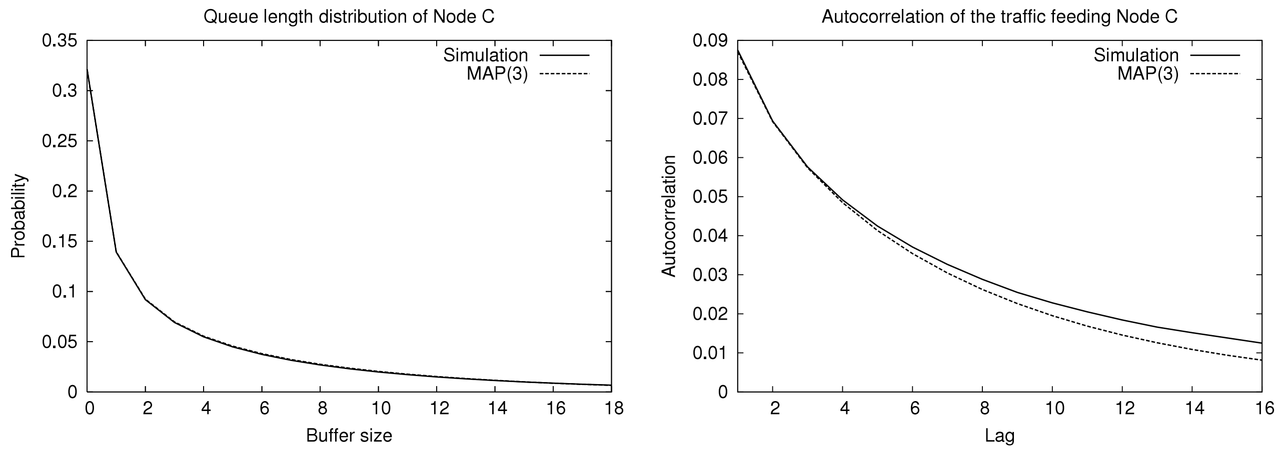

Fig. 4. Queue length distribution and autocorrelation of arrivals of Node C in Example 2.

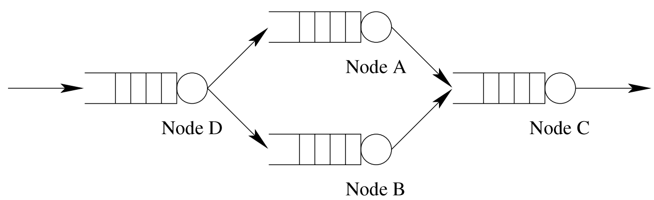

Fig. 5. The queueing network used in Example 3.

Table 3
Mean queue length of nodes in Example 3.

	Node D	Node A	Node B	Node C
Simulation	4.24696	1.0709	1.94556	5.4563
MAP(2)	4.24962	1.06936	1.9342	5.23628
Rel. error	−0.06%	−0.1%	−0.5%	−4%
MAP(3)	4.24962	1.07144	1.94196	5.25906
Rel. error	−0.06%	0.05%	−0.2%	−3.6%

mation provides better results than the 3-state one when the superposed traffic is compressed. We do not have explicit explanation for this phenomena, we believe however that it is due to the random interplay of the two approximations, the one of the output process and the one of the compression of the superposed process. The table also contains the mean queue length when the output of Node A and Node B is approximated by a renewal process. In spite of the low lag-1 correlation of the MAPs of this example the results indicate that the renewal output assumption is less accurate than the MAP model capturing some correlation measures of the traffic.

The queue length distribution and the autocorrelation of the traffic feeding Node C are depicted in Fig. 4.

5.3. *A four-node network with splitting and superposition*

As a last example we consider a queueing network with both splitting and superposition. Fig. 5 depicts the structure of this network. The MAP describing the arrivals entering to the network (i.e., the traffic of Node D) is given by

$$\mathbf{D_0^D} = \begin{bmatrix} -62.5 & 7.5 & 25 \\ 2.5 & -15 & 0 \\ 0 & 10 & -25 \end{bmatrix}, \qquad \mathbf{D_1^D} = \begin{bmatrix} 25 & 0 & 5 \\ 5 & 7.5 & 0 \\ 12.5 & 0 & 2.5 \end{bmatrix},$$

with average intensity, coefficient of variation and lag-1 correlation coefficient of $\lambda = 16.6,\ c_v^2 = 1.31,\ \rho_1 = 0.027$. The matrices of the service MAP are

$$\mathbf{S_0^D} = \begin{bmatrix} -62.5 & 25 \\ 0 & -17.5 \end{bmatrix}, \qquad \mathbf{S_1^D} = \begin{bmatrix} 25 & 12.5 \\ 10 & 7.5 \end{bmatrix}.$$

The basic properties of the service MAP are $\lambda = 21.71,\ c_v^2 = 1.31,\ \rho_1 = 0.007$.
The service processes of Nodes A, B and C are the same as in Example 2. Each departing customer of Node D is directed to Node A with probability 0.3 and to Node B with probability 0.7.

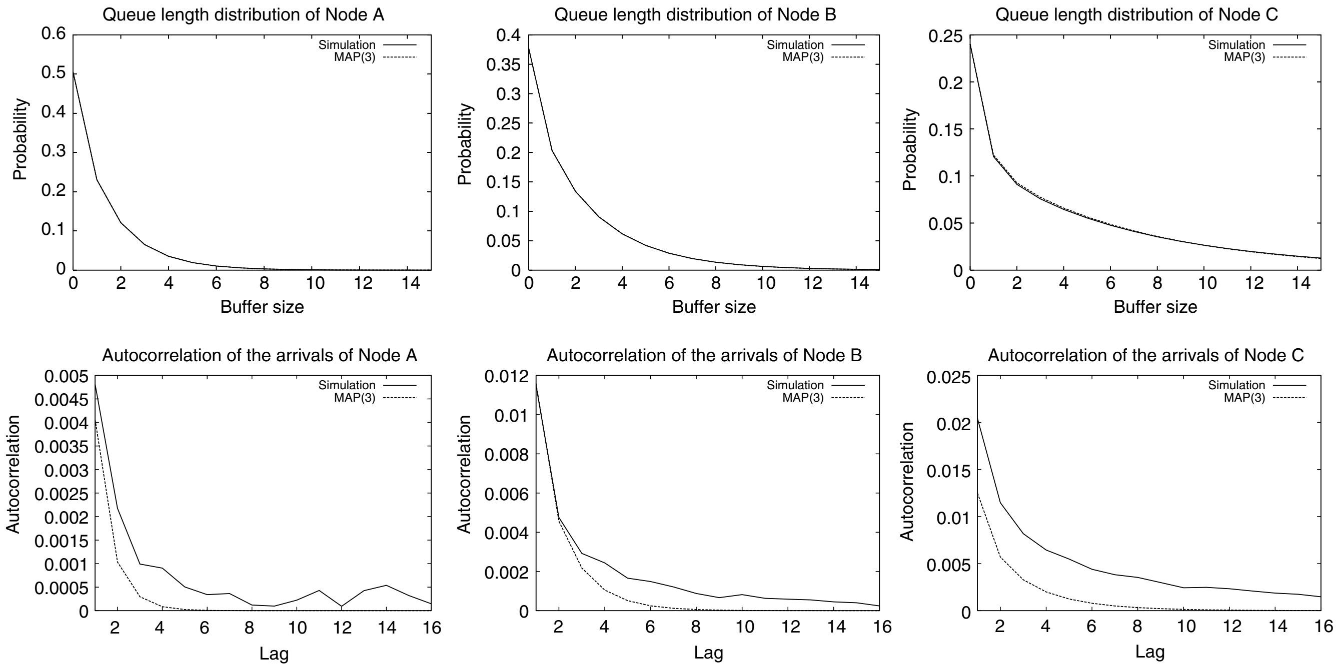

Fig. 6. Autocorrelation of arrivals and queue length distribution of nodes in Example 3.

The mean queue length results of the nodes are summarised in Table 3. In this example the accuracy is reasonable high (the error is below 4% compared to the simulation) both when MAPs(2) and MAPs(3) are used to approximate the departure traffic of the queues.

Fig. 6 depicts the queue length distribution of the nodes and the autocorrelation of the arriving traffic. The queue length distribution is approximated very accurately even if the high lag correlations are not captured exactly. Computation time of these results was between 1–2 s which indicates that this approximation method does not have scaling problems and can be applied for more complex queueing networks. This is not the case with the truncation based methods.

5.4. Networks with feedback

In this section we analyse the same networks considered so far but with the introduction of feedback. Accordingly, there are mutual dependencies between the nodes and the network cannot be evaluated by a single round of analysis of the nodes. A fixed-point iteration scheme is applied to resolve the mutual dependencies. We proceed as follows. In the first round the network is evaluated considering only those traffic streams that arrive from external sources. Based on this first round we obtain a first approximation of the feedback traffic streams. These approximations are used then in the second round of evaluation which provides a better approximation of the behaviour of the network and also of the feedback traffic. The iteration is stopped when for all the nodes of the network we have that the queue length distribution does not change more than a predefined measure from one iteration to another. As distance measure we apply the sum of the absolute differences, i.e., having two probability mass functions, p_i and q_i, their distance is calculated as $\sum_{i=0}^{\infty} |p_i - q_i|$. In the examples presented hereinafter we stopped the calculations when the distance of two consecutive queue length distributions was below 10^{-6} for all the nodes.

All the calculations were performed applying 2-state MAPs. We failed to use MAPs with higher number of states because the moment sets of the output traffic are such that the applied moment matching algorithm [11] fails. More precisely, the moment matching algorithm generates a pair of matrices, whose moments (4) and joint moments (6) are identical with the required ones, but these matrices do not define a MAP (e.g., contains negative off-diagonal element). In some cases the method given in [11] failed even with two states. In these cases we modified the moment set according to the procedure outlined in Section 4.6.

We consider first Case (a) of Example 1 with the assumption that customers leaving Node B return to Node A with probability p_{BA}. Figs. 7 and 8 report, with two different values of p_{BA}, the queue length distributions after 1, 2 and 8 rounds of calculation and the results obtained by simulation. For this topology the feedback we have introduced increases the traffic of both nodes by 25% in case of $p_{BA} = 0.2$ and by 66.6% in case of $p_{BA} = 0.4$. The arrival rate in the ith round of iteration is given simply by $\lambda \sum_{j=0}^{i-1} p_{BA}^j$ where λ is the intensity of the external source. For example, in case of $p_{BA} = 0.4$ and $\lambda = 0.5$ the amount of traffic we consider grows as 0.5, 0.7, 0.78, 0.812, 0.8248, 0.82992, 0.831968, 0.832787, 0.833115, ... This geometric series provides an estimate for the number of necessary iterations. Mean number of clients is reported in Table 4. In this case the moment based approximation yields reasonably accurate results.

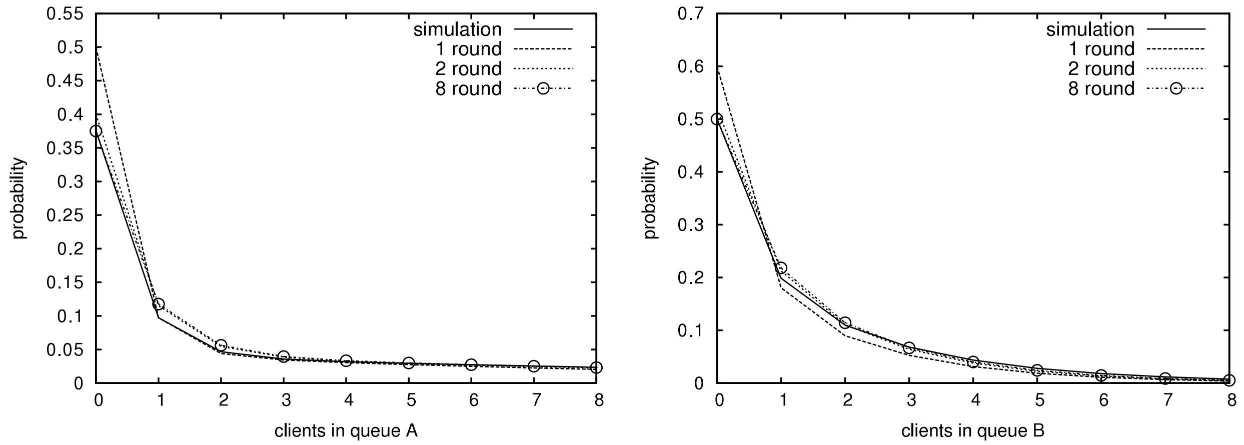

Fig. 7. Example 1(a) with feedback $p_{BA} = 0.2$.

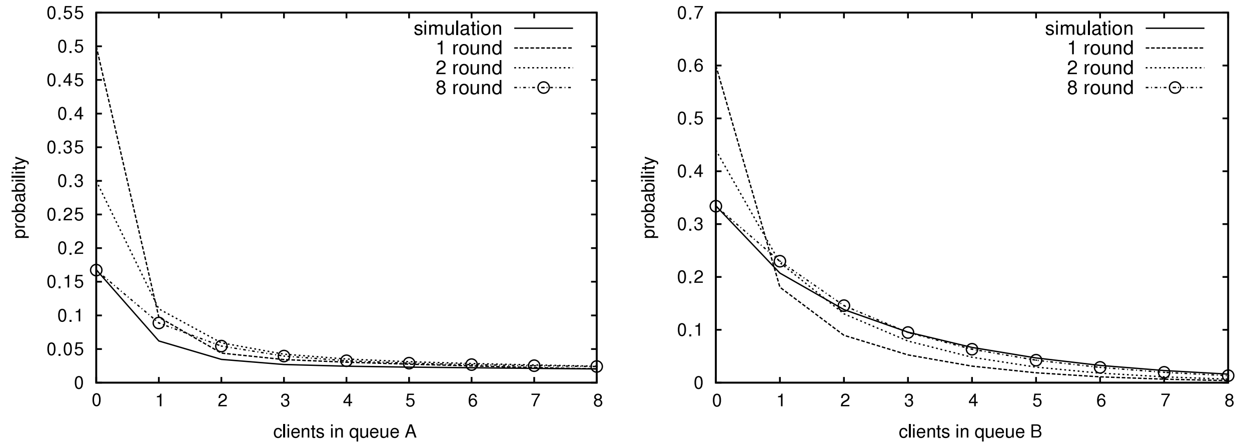

Fig. 8. Example 1(a) with feedback $p_{BA} = 0.4$.

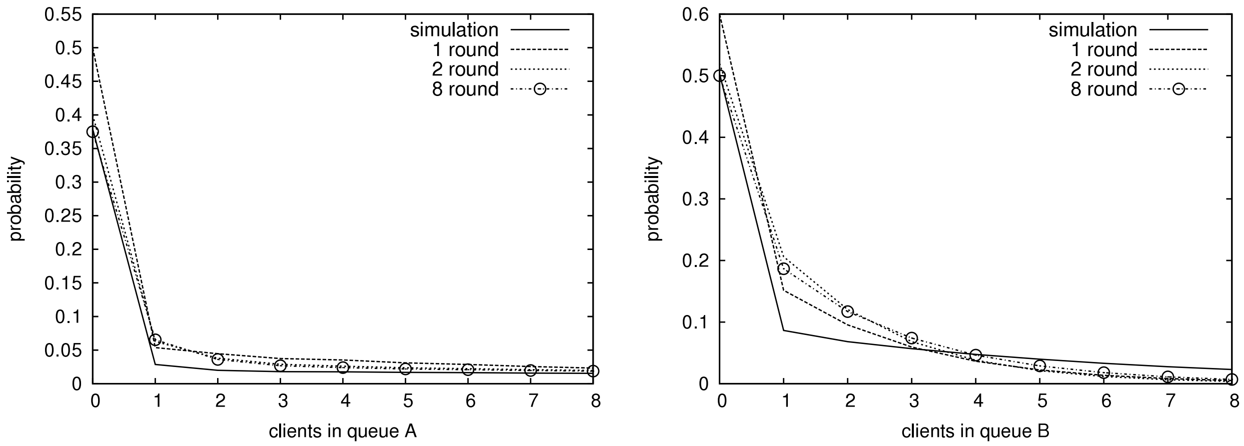

Fig. 9. Example 1(b) with feedback $p_{BA} = 0.2$.

Table 4
Example 1: Mean number of clients at the nodes for various values of feedback probabilities.

Example 1		Simulation		MAP(2) approximation	
Case	p_{BA}	Node A	Node B	Node A	Node B
(a)	0.2	7.89277	1.34513	6.83832	1.20222
(a)	0.4	24.1497	2.2041	16.7656	2.00624
(b)	0.2	19.5783	2.96365	12.6555	1.33503
(b)	0.4	139.503	4.97111	72.1915	3.48855
(c)	0.2	18.4693	2.43225	11.1764	1.45485
(c)	0.4	137.936	3.80106	69.4735	2.66769

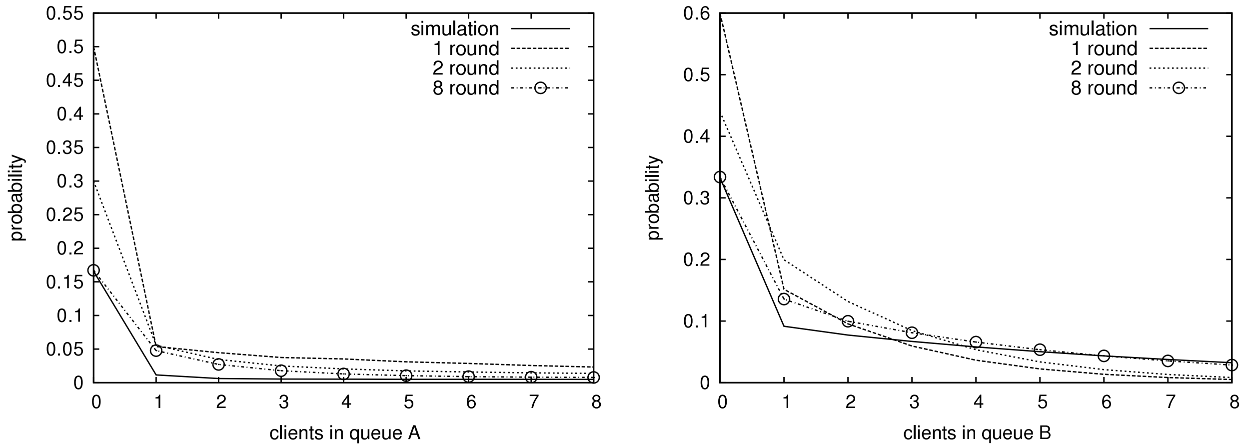

Fig. 10. Example 1(b) with feedback $p_{BA} = 0.4$.

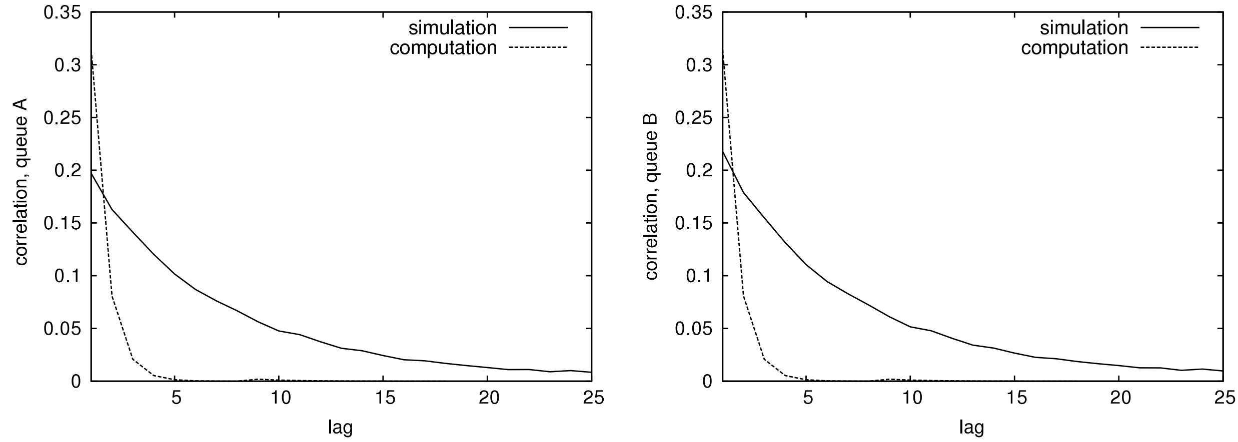

Fig. 11. Correlation of clients leaving Node B for Example 1(b) with feedback $p_{BA} = 0.4$.

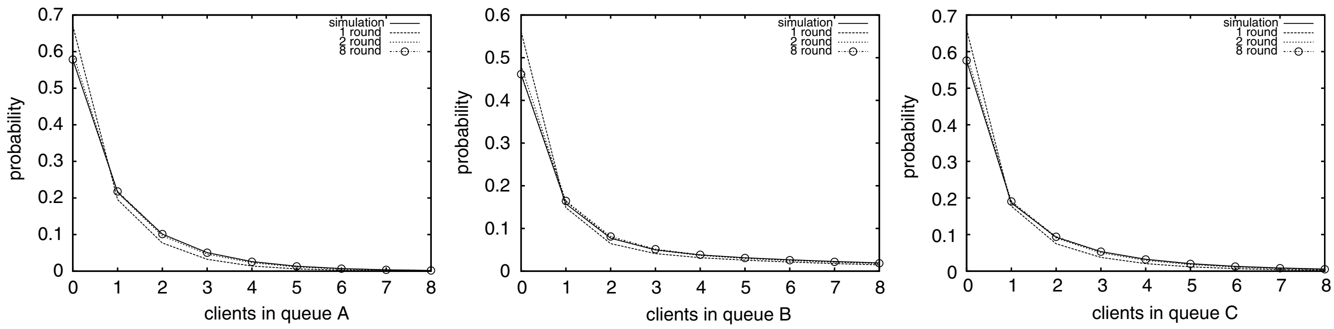

Fig. 12. Example 2 with feedback $p_{CA} = 0.1$ and $p_{CB} = 0.1$.

In Figs. 9 and 10 we have depicted the results for Case (b) of Example 1. Both the figures and Table 4 demonstrate that for this case the approximation is not accurate. The reason is that in this case the slightly more correlated external traffic together with the more complex nature of the server at node A result in such correlation structure that cannot be approximated well enough by a 2-state MAP. Indeed, we had to resort to the procedure given in Section 4.6 to perform the analysis. The correlation of the traffic stream leaving node B is depicted in Fig. 11. For Case (c) of Example 1 the results are very similar to those for Case (b).

For what concerns Example 2, we introduced feedback from Node C to Node A and B with probability p_{CA} and p_{CB}, respectively. In order to allow for more significant feedback probabilities the speed of the servers in Node A and C are doubled. This was achieved simply by multiplying their matrix descriptors by 2. Figs. 12 and 13 depict cases with different feedback probabilities. The mean number of clients are reported in Table 5. In this case both with light feedback, which results in rather low utilization, and stronger feedback, which results in heavy load, the approximation provides good results both in terms in distributions and mean values.

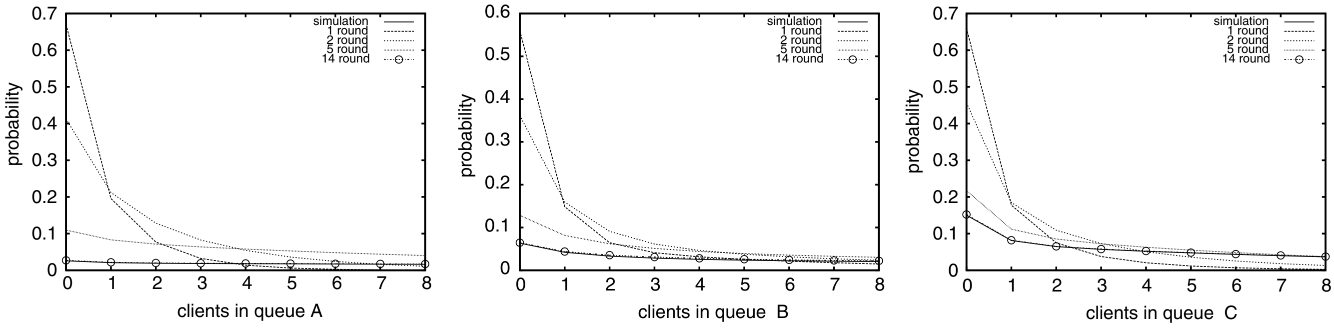

Fig. 13. Example 2 with feedback $p_{CA} = 0.35$ and $p_{CB} = 0.25$.

Table 5
Example 2, mean number of clients at the nodes for various values of feedback probabilities.

Example 2		Simulation			MAP(2) approximation		
p_{CA}	p_{CB}	Node A	Node B	Node C	Node A	Node B	Node C
0.2	0.2	1.61072	5.9741	2.05407	1.5771	5.22479	1.94212
0.1	0.1	0.84514	3.0113	1.08031	0.83576	2.8155	1.04058
0.35	0.25	53.8258	34.1808	10.5818	49.231	30.4029	10.2632

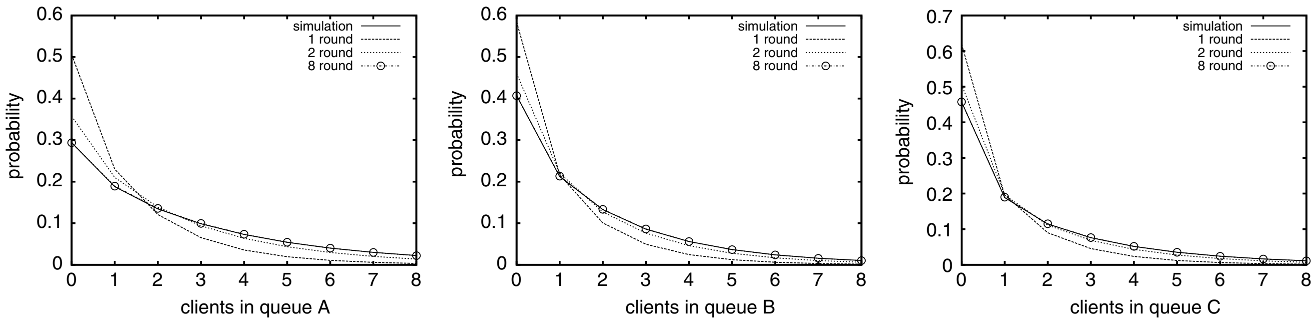

Fig. 14. Example 3 with feedback $p_{CD} = 0.3$.

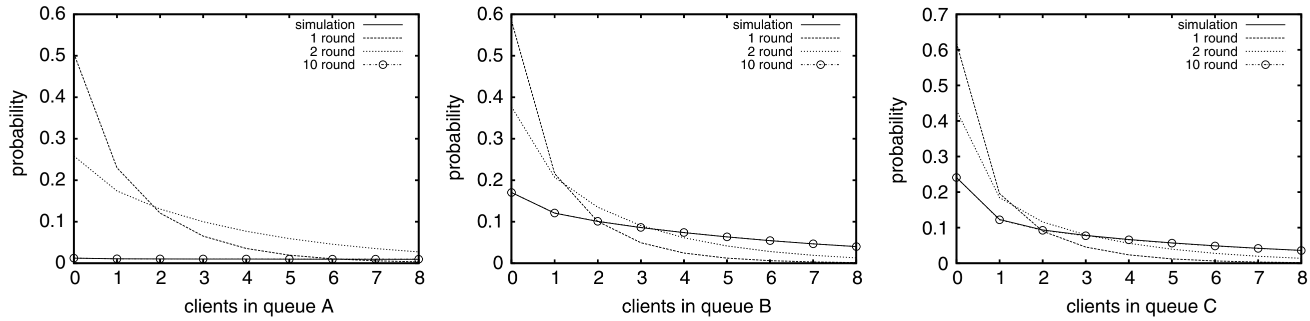

Fig. 15. Example 3 with feedback $p_{CD} = 0.5$.

In the following we consider the fork and join network (Example 3) with feedback from Node C to Node D. Nodes B, C and D are speeded up by a factor of 1.5, 2 and 2, respectively, in order to allow for more significant feedback. Figs. 14 and 15 and Table 6 report results with various feedback levels. In this case as well the approximation yields good results both for heavy and light loads. Figs. 16 and 17 depict the correlation of the traffic leaving the nodes with various levels of feedback probabilities. In this case as well the correlation is not captured accurately. However, since the correlation is low its impact on the performance indices is not as strong as in case of Example 1(b). It is interesting to note how the load of the network changes the correlation structure of the traffic leaving the nodes. Consider, for example, Fig. 16. Node A exhibits the slowest decaying correlation with medium load ($p_{CD} = 0.3$). Both with higher load ($p_{CD} = 0.5$) and lower load ($p_{CD} = 0.1$) the correlation decays faster. This is because with medium loads the output traffic is effected by both the correlation of the input traffic and the correlation of the server.

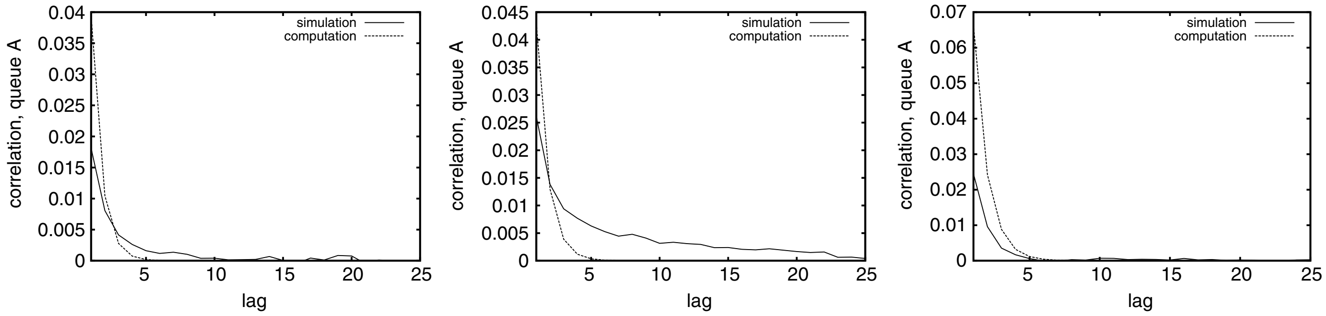

Fig. 16. Example 3 with feedback $p_{CD} = 0.1, 0.3, 0.5$, correlation of clients leaving Node A.

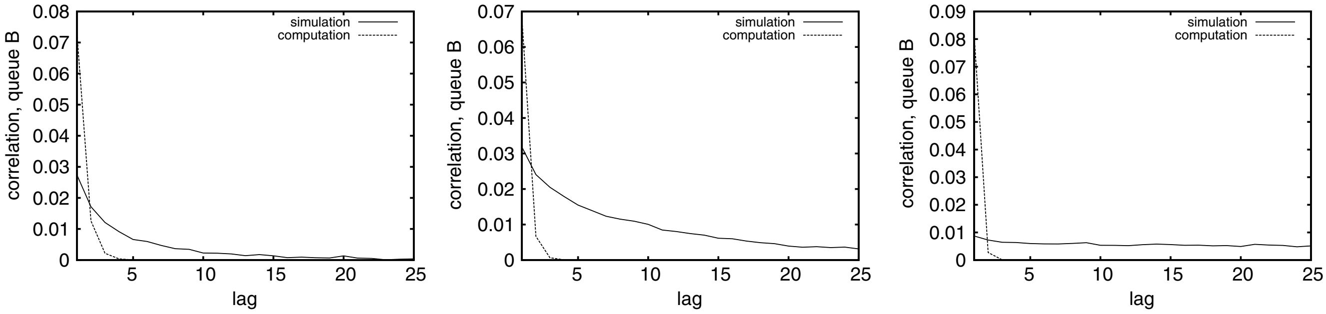

Fig. 17. Example 3 with feedback $p_{CD} = 0.1, 0.3, 0.5$, correlation of clients leaving Node B.

Table 6
Example 3, mean number of clients at the nodes for various values of feedback probabilities.

Ex. 3	Simulation				MAP(2) approximation			
p_{CD}	Node D	Node A	Node B	Node C	Node D	Node A	Node B	Node C
0.1	0.879425	1.35113	0.982433	0.947525	0.878006	1.34292	0.977575	0.927992
0.3	1.45455	2.72948	1.69404	1.6556	1.44636	2.68637	1.67931	1.63799
0.5	4.1277	91.7678	5.85162	5.14396	4.22406	102.483	5.85292	5.2919

6. Conclusions

This paper provides an approximation for the output process of MAP/MAP/1 queues. In particular, we propose approximating the output process of a MAP/MAP/1 queue based on the moments of the inter-departure time and the joint moments of two consecutive inter-departure times. Then this approximation is used for the analysis of queueing networks with traffic superposition and splitting with and without feedback.

The proposed moments based approximation method was tested in numerical examples and showed reasonable accuracy compared to simulation results. An important feature of the proposed method is that the size of the traffic models remains small during the analysis of larger queueing networks. This was not the case with the previously proposed approximations. Due to this property moments based approximation provides a fast approximation of larger queueing networks than the previously analysable ones.

Appendix A. Proof of Theorem 1

Based on (2) and the properties of the superposition of MAPs we have that

$$\begin{aligned}\pi_F &= \frac{-1}{\lambda_D + \lambda_E}(\alpha_D \otimes \alpha_E)(\mathbf{D_0} \oplus \mathbf{E_0}) \\ &= \frac{-1}{\lambda_D + \lambda_E}(\alpha_D \otimes \alpha_E)(\mathbf{D_0} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{E_0}) \\ &= \frac{-1}{\lambda_D + \lambda_E}(\alpha_D \mathbf{D_0} \otimes \alpha_E + \alpha_D \otimes \alpha_E \mathbf{E_0}),\end{aligned}$$

where we applied compatibility of the Kronecker product [16]. By further application of (2)

$$\begin{aligned}\pi_F &= \frac{1}{\lambda_D + \lambda_E}\left(\lambda_D \pi_D \otimes \lambda_E \pi_E (-\mathbf{E_0})^{-1} + \lambda_D \pi_D (-\mathbf{D_0})^{-1} \otimes \lambda_E \pi_E\right) \\ &= \frac{\lambda_D \lambda_E}{\lambda_D + \lambda_E}\left(\pi_D \otimes \pi_E (-\mathbf{E_0})^{-1} + \pi_D (-\mathbf{D_0})^{-1} \otimes \pi_E\right).\end{aligned} \tag{38}$$

The left hand side of (25) can be calculated as

$$\gamma_{ij}^F = -\pi_F \mathbf{F_0}^i \mathbf{F_1} \mathbf{F_0}^{j+1} \mathbf{1}$$
$$= -\pi_F (\mathbf{D_0} \oplus \mathbf{E_0})^i (\mathbf{D_1} \oplus \mathbf{E_1})(\mathbf{D_0} \oplus \mathbf{E_0})^{j+1} \mathbf{1}. \tag{39}$$

Since

$$(\mathbf{D_0} \otimes \mathbf{I})(\mathbf{I} \otimes \mathbf{E_0}) = (\mathbf{D_0}\mathbf{I}) \otimes (\mathbf{I}\mathbf{E_0}) = (\mathbf{I}\mathbf{D_0}) \otimes (\mathbf{E_0}\mathbf{I}) = (\mathbf{I} \otimes \mathbf{E_0})(\mathbf{D_0} \otimes \mathbf{I}),$$

$(\mathbf{D_0} \otimes \mathbf{I})$ and $(\mathbf{I} \otimes \mathbf{E_0})$ commutes and we can expand $(\mathbf{D_0} \oplus \mathbf{E_0})^i$ as

$$(\mathbf{D_0} \oplus \mathbf{E_0})^i = (\mathbf{D_0} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{E_0})^i$$
$$= \sum_{k=0}^{i} \binom{i}{k} (\mathbf{D_0} \otimes \mathbf{I})^{i-k} (\mathbf{I} \otimes \mathbf{E_0})^k$$
$$= \sum_{k=0}^{i} \binom{i}{k} \mathbf{D_0}^{i-k} \otimes \mathbf{E_0}^k.$$

Further more

$$\gamma_{ij}^F = -\pi_F \left(\sum_{k=0}^{i} \binom{i}{k} \mathbf{D_0}^{i-k} \otimes \mathbf{E_0}^k \right) (\mathbf{D_1} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{E_1}) \left(\sum_{\ell=0}^{j+1} \binom{j+1}{\ell} \mathbf{D_0}^{j+1-\ell} \otimes \mathbf{E_0}^\ell \right) \mathbf{1},$$

which, based on the compatibility of ordinary and Kronecker product and applying (38), can be written as

$$\gamma_{ij}^F = \frac{\lambda_D \lambda_E}{\lambda_D + \lambda_E} \sum_{k=0}^{i} \sum_{\ell=0}^{j+1} \binom{i}{k} \binom{j+1}{\ell} \left(\left(\pi_D \mathbf{D_0}^{i-k} \mathbf{D_1} \mathbf{D_0}^{j+1-\ell} \mathbf{1} \right) \left(\pi_E \mathbf{E_0}^{k+\ell-1} \mathbf{1} \right) + \left(\pi_D \mathbf{D_0}^{i-k+j+1-\ell} \mathbf{1} \right) \left(\pi_E \mathbf{E_0}^{k-1} \mathbf{E_1} \mathbf{E_0}^\ell \mathbf{1} \right) \right.$$
$$\left. + \left(\pi_D \mathbf{D_0}^{i-k-1} \mathbf{D_1} \mathbf{D_0}^{j+1-\ell} \mathbf{1} \right) \left(\pi_E \mathbf{E_0}^{k+\ell} \mathbf{1} \right) + \left(\pi_D \mathbf{D_0}^{i-k+j-\ell} \mathbf{1} \right) \left(\pi_E \mathbf{E_0}^k \mathbf{E_1} \mathbf{E_0}^\ell \mathbf{1} \right) \right).$$

This final expression is equal to the right hand side of (25).

References

[1] Forest Baskett, K. Mani Chandy, Richard R. Muntz, Fernando G. Palacios, Open, closed, and mixed networks of queues with different classes of customers, Journal of the ACM 22 (2) (1975) 248–260.
[2] W. Whitt, Approximating a point process by a renewal process, I: Two basic methods, Operations Research (1982) 125–147.
[3] W. Whitt, Approximations for departure processes and queues in series, Naval Research Logistics Quarterly (1984) 499–521.
[4] J. Roberts, U. Mocci, J. Virtamo (Eds.), Broadband Network Teletraffic, Springer, 1996.
[5] G. Latouche, V. Ramaswami, Introduction to Matrix Analytic Methods in Stochastic Modeling, SIAM, Philadelphia, 1999.
[6] D.A. Bini, B. Meini, S. Steffé, B. Van Houdt, Structured Markov chains solver: Software tools, in: SMCtools Workshop, ACM Press, Pisa, Italy, 2006.
[7] A. Heindl, Traffic-Based Decomposition of General Queueing Networks with Correlated Input Processes, Shaker Verlag, Aachen, 2001.
[8] R. Sadre, B.R. Haverkort, Characterizing traffic streams in networks of MAP/MAP/1 queues, in: Proceedings 11th GI/ITG Conference on Measuring, Modelling and Evaluation of Computer and Communication Systems, MMB 2001, VDE Verlag, 2001, pp. 195–208.
[9] A. Heindl, Q. Zhang, E. Smirni, Etaqa truncation models for the MAP/MAP/1 departure process, in: QEST '04: Proceedings of The Quantitative Evaluation of Systems, First International Conference on (QEST'04), IEEE Computer Society, Washington, DC, USA, 2004, pp. 100–109.
[10] P. Buchholz, Bounding stationary results of tandem networks with map input and PH service time distributions, in: Proc. Sigmetrics/Performance 2006, ACM Press, 2006, pp. 191–202.
[11] M. Telek, G. Horváth, A minimal representation of Markov arrival processes and a moments matching method, Performance Evaluation 64 (9–12) (2007) 1153–1168.
[12] L. Bodrog, A. Horváth, M. Telek, Moment characterization of matrix exponential and Markovian arrival processes, Annals of Operations Research 160 (2008) 51–68.
[13] D. Bini, G. Latouche, B. Meini, Numerical Methods for Structured Markov Chains, (Numerical Mathematics and Scientific Computation), Oxford University Press, Inc., New York, NY, USA, 2005.
[14] N.G. Bean, D.A. Green, P.G. Taylor, Approximations to the output process of MAP/PH/1 queues, in: 2nd International Conference on Matrix Analytic Methods, Notable Publications Inc., NJ, 1998, pp. 151–169.
[15] G. Horvath, L. Bodrog, A. Heindl, M. Telek, A Markovian canonical form of second-order matrix-exponential processes, European Journal of Operational Research 190 (2) (2008) 459–477.
[16] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, 1991.

András Horváth was born in 1974 in Budapest where he received the M.Sc. degree in Computer Science from the University of Technology and Economics. From 1998 to 2002 he was a Ph.D. student supervised by Miklós Telek at the same university. From 2003 he is a researcher at the University of Turin (Italy). His research interests are in the area of stochastic processes including performance analysis of non-Markovian systems and modeling issues of communication networks.

Gábor Horváth is an assistant professor at the Department of Telecommunications of the Budapest University of Technology and Economics. He received his M.Sc. and Ph.D. degrees in Computer Science from the Budapest University of Technology and Economics in 2001 and 2005, respectively. He is a member of the stochastic modeling laboratory of the department and he plays an active role in the international and industrial research projects of the laboratory. He is a project leader in the co-operation with Nokia Research Center Budapest on traffic modeling of mobile telecommunication networks. His research interests include traffic modeling and performance analysis of telecommunication systems using matrix geometric methods.

Miklós Telek received the M.Sc. degree in Electrical Engineering from the Technical University of Budapest in 1987. In the same year he joined the Hungarian Post Research Institute where he studied the modelling, analysis and planning aspects of communication networks. Since 1990 he has been with the Department of Telecommunications of the Technical University of Budapest, where he is a full professor now. He received the C.Sc. and D.Sc. degree from the Hungarian Academy of Science in 1995 and 2004, respectively. His current research interest includes stochastic performance modeling and analysis of computer and communication systems.