Computers and Mathematics with Applications 64 (2012) 3852–3868

Contents lists available at SciVerse ScienceDirect

Computers and Mathematics with Applications

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

Algorithmic product-form approximations of interacting stochastic models

Andrea Marin [a], Maria Grazia Vigliotti [b,*]

[a] Università Ca' Foscari Venezia, DAIS, via Torino 155, Venezia, Italy

[b] Department of Computing, Imperial College London, South Kensington Campus, SW7 2BZ, UK

ARTICLE INFO

Keywords:
Product-form solutions
Stochastic models
Approximations

ABSTRACT

A large variety of product-form solutions for continuous-time Markovian models can be derived by checking a set of structural properties of the underlying stochastic processes and a condition on their reversed rates. In previous work (Marin and Vigliotti (2010) [9]) we have shown how to derive a large class of product-form solutions using a different formulation of the Reversed Compound Agent Theorem (GRCAT). We continue this line of work by showing that it is possible to exploit this result to approximate the steady-state distribution of non-product-form model interactions by means of product-form ones.

1. Introduction

Stochastic models with underlying Continuous Time Markov Chains (CTMCs) are widely applied for the analysis of natural or artificial systems. Specifically, the steady-state analysis plays a pivotal role in the performance engineering of software and hardware architectures in computer science. Since most systems consist of a set of interacting components, the state-space of the whole model tends to be very large, and the well-known standard techniques for the computation of the steady-state probabilities are no longer feasible. Models with product-form solutions overcome this problem since the computation of the steady-state probabilities can be independently performed for each component [1–4]. As models of real systems might not always admit a product-form solution, we address the problem of approximating general models by means of product-form ones. Several other authors have addressed the same problem (see, e.g., [5–7]), yet the technique proposed in this paper is completely novel and relies on the recent formulation of the Reversed Compound Agent Theorem (GRCAT) [8,9]. Roughly speaking, GRCAT provides sufficient conditions (on the rates and on the structure of the state-space of the stochastic models) to guarantee the product-form solution.

The result is formulated in terms of cooperating Labelled Markov Automata (LMA). Cooperating automata can be thought of as labelled graphs, which can be composed together with respect to a set of labels to obtain bigger models. The aim of this formalism is to specify complex CTMCs starting from simpler ones. GRCAT states sufficient conditions on the cooperation among CTMCs to achieve product-form solutions.

If a model is not in product-form, then it does not satisfy GRCAT conditions. In this paper we define two novel algorithms to transform any pairwise cooperating model into an approximated one that satisfies GRCAT conditions, and enjoys a product-form solution. One of the strengths of our method relies on *modularity*, i.e., the modifications required to obtain the product-form solution for an LMA model may be applied to each component in isolation. For instance, let us consider a two-node tandem network consisting of queues Q_1 and Q_2 and let the second queue have infinite capacity. The application of the proposed algorithms modifies the models to $\tilde{Q}_1$ and $\tilde{Q}_2$ so that Q_1 and $\tilde{Q}_1$ share the same steady-state distribution, while $\tilde{Q}_2$ steady-state distribution is an approximation of that of Q_2.

* Corresponding author. Tel.: +44 2075948242.
E-mail addresses: marin@dais.unive.it (A. Marin), mgv98@doc.ic.ac.uk (M.G. Vigliotti).

0898-1221/$ – see front matter Crown Copyright © 2012 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2012.03.034

To test the results of the approximation techniques proposed in this paper, we consider four case studies consisting of networks that are not in product-form. For each of them, we derive an approximated model in product-form by applying the algorithms devised in this work. Then, we compare some significant performance indices computed in the original and in the approximated model. The first example considers a tandem of queues: M/M/1 queue and a./M/1 queue with finite buffer size. The second and the third case studies model tandem of queues with customer arrivals generated by a homogeneous Poisson process. In the second example, the first queue has a Coxian service time distribution with two stages: arrivals occurring when the second stage is busy are discarded. The scheduling discipline is First Come First Served (FCFS). In the third example, the first queue has still a Coxian service time distribution but the service discipline is Last Come First Served with pre-emption (LCFS). Notice that, no resume policy is assumed, hence the queue is not quasi-reversible (in particular it is different from that presented in [2]). The fourth example consists of a network of three queues: two exponential queues competing for the service of a third one. In the competition, some customers get destroyed.

1.1. Related work

Although the technique adopted in this paper is novel, the problem of deriving product-form approximations for non-product-form Markovian models has already been addressed in literature. A similar work has been carried out by van Dijk in [5] where models without product-form solutions are approximated by modifying their structure to achieve *local balance*. In the theory of queueing networks, it is well-known that if for each station it holds that for each state the rate out due to a customer departure equals the rate in due to a customer arrival, then the network has product-form [4]. While local balance is applicable to queueing networks only, our method could be applied to a wider class of Markovian models thanks to the generality of GRCAT.

Algorithms for approximating steady-state distributions have been investigated by Buchholz in [10,7]. Cooperating CTMCs [7] are modelled using the Kronecker representation. Buchholz's approximation algorithms are based on a numerical minimisation technique. The author proposes a method to find the marginal steady-state distributions of the components whose product minimises the vector of residuals of the global balance equation system. The main disadvantage of this method is the 'lack of control' over the minimisation function. In other words, once the approximated product-form solution is found, the modeller does not know how the system has been modified or whether these modifications are feasible. By contrast, our technique does not only find the approximated steady-state distribution, but also gives a concrete description of the perturbations needed on each component of the system, to have product-form.

The rest of the paper is organised as follows in Section 2 we revise the theoretical background; in Section 3 we describe our approximation methods, and in Section 4 we show applications of our methods to some concrete examples. Finally, Section 5 presents some final remarks.

2. Theoretical background

In this section we provide a brief description of LMAs and GRCAT. The presentation aims to make the paper readable, but it cannot be exhaustive due to lack of space. Any reader interested in a formal and detailed treatment of the LMAs and GRCAT should look at [8,9]. We have also included an Appendix for readability.

2.1. Labelled Markov Automata (LMAs)

First, we introduce LMAs and recall the main definition, then we briefly review the conditions for the product-form solution.

Definition 1. A Markov automaton is a tuple $\mathbf{M} = \langle \mathcal{S}, Act, \rightarrow \rangle$ such that:

1. $\mathcal{S}$ is the denumerable set of states (*state-space*) with $s_1, s_2, \ldots, s_n, \ldots$ ranging over it,
2. *Act* is the set of *action labels* (or simply *labels*) with $a, b, \ldots$ ranging over it,
3. $\rightarrow$ is the transition relation between states defined as follows:

$$\rightarrow : \mathcal{S} \times Act \times (\mathbb{R}^+ \cup \mathbf{Var}) \times \mathcal{S},$$

where $\mathbb{R}^+$ is the set of positive real numbers and **Var** is the set of variable names such that if $a \in Act$ then $x_a \in \mathbf{Var}$.

For readability, we write $(s_1, a, \lambda, s'_1) \in \rightarrow$ as $s_1 \xrightarrow{a,\lambda} s'_1$. We define two sets $\mathcal{A}(\mathbf{M})$ and $\mathcal{P}(\mathbf{M})$, *active actions labels* and *passive actions labels*, such that for every $a \in Act$ if $s \xrightarrow{a,\lambda} s'$, with $\lambda \in \mathbb{R}^+$, then $a \in \mathcal{A}(\mathbf{M})$, and if $s \xrightarrow{a,x_a} s'$, with $x_a \in \mathbf{Var}$, then $a \in \mathcal{P}(\mathbf{M})$.

We shall also use the word label instead of action label.

Labels are meaningful to specify the cooperations among Markov automata as they denote how the transitions synchronise. Informally, we can say that transitions in LMAs are divided into active and passive. *Active transitions* are those with an associated *delay*, i.e., the rate is a positive real number. *Passive transitions* are those whose delays are undefined, i.e., the rate is a variable. Passive and active transitions determine which labels are active and which ones are passive.

Two automata $\mathbf{M}_1$ and $\mathbf{M}_2$ cooperate on a set of labels L. A transition labelled $a \in L$ must be active with respect to an automaton and passive with respect to the other (e.g., $a \in \mathcal{A}(\mathbf{M}_1) \cap \mathcal{P}(\mathbf{M}_2)$). Intuitively, transitions labelled $a \in L$ can be performed by the synchronising automata only simultaneously. The automata taking part in the cooperation change their state: the passive automaton is forced to move at the rate of the active one. In LMAs all the transitions with a specified rate are carried out according to an exponentially distributed random delay.

In this work, we assume that for every automaton $\mathbf{M}$ the set of active and passive labels are disjoint. In other words, a label $a \in Act$ either appears always with an associated rate in $\mathbf{M}$ ($a \in \mathcal{A}(\mathbf{M})$), or with a variable x_a ($a \in \mathcal{P}(\mathbf{M})$). Moreover, we exclude the possibility of two passive transitions with the same label outgoing from the same state. When an LMA $\mathcal{P}(\mathbf{M})$ is such that $\mathcal{P}(\mathbf{M}) = \emptyset$, then we say that $\mathbf{M}$ is *closed*, otherwise we say that it is *open*. The process underlying a closed LMA is a CTMC whose set of states is the same as that of the automaton and the transition rate from state s to s' is given by the sum of the rates of all the labelled transitions of the automaton from s to s', i.e.:

$$q(s \to s') = \sum_{\substack{(a,\lambda):s \xrightarrow{a,\lambda} s' \\ s \neq s'}} \lambda.$$

We also write $q(s \xrightarrow{a} s') = \sum_{\lambda:s \xrightarrow{a,\lambda} s'} \lambda$ for the rate respect to a label a. Before presenting the theorem on which our approximation methods are based, we need to define an operation on an automaton which we call *closure*. Informally, the closure specifies the rates of the passive transitions of an open automaton $\mathbf{M}$. Suppose that $\mathcal{P}(\mathbf{M}) = \{a_1, \dots, a_N\}$, then we write:

$$\mathbf{M}^c = \mathbf{M}\{x_{a_1} \leftarrow K_{a_1}, \dots, x_{a_N} \leftarrow K_{a_N}\}, \quad K_i \in \mathbb{R}^+$$

to refer to automaton $\mathbf{M}$ in which each transition labelled $a_i \in \mathcal{P}(\mathbf{M})$ becomes active with rate K_{a_i}. Note that, in $\mathbf{M}^c$ all the transitions sharing the same label, and that were passive in $\mathbf{M}$ have the same rate, and that $\mathbf{M}^c$ has an well-defined underlying CTMC and hence the steady-state analysis can be carried out.

Remark 1.
- Having derived the CTMC from an LMA, we will refer directly to the properties of the LMA, meaning the properties of the underlying CTMC.
- If $\mathbf{Q}$ is the generator matrix of the underlying CTMC of automaton $\mathbf{M}$, then we write $\pi(\mathbf{M})$ for *invariant measure* meaning that $\pi(\mathbf{M})\mathbf{Q} = 0$; when clear from the context, we simply write π instead of $\pi(\mathbf{M})$ and denote the component associated with state s by $\pi(s)$. If $\sum_{s \in \mathcal{S}} \pi(s) = 1$ and $\pi(s) > 0$ for all s, then the CTMC is ergodic and π is the unique steady-state distribution of $\mathbf{M}$ [11].

2.2. Product-form solutions

In this section, we informally introduce GRCAT as proved in [9]. In what follows, we implicitly assume that the models for which we consider the steady-state distributions are ergodic.

According to [9], two interacting automata $\mathbf{M}_1 \oplus_L \mathbf{M}_2$ are in product-form if:

1. *Condition* 1: a is a passive label, $a \in \mathcal{P}(\mathbf{M}_l)$ with $l = 1, 2$, then for every state s of the automaton $\mathbf{M}_l$ there exists exactly one transition labelled a outgoing from s.
2. *Condition* 2: There exists the set of rates $\{K_{a_1}, \dots, K_{a_n}\}$ with $\{a_1, \dots, a_n\} = L$ which satisfies the following equations:

$$\forall s_k \in \mathcal{S}_l, \forall a_i \in \mathcal{A}(\mathbf{M}_l), \quad \frac{\sum_{s_j \in \mathcal{S}_l} q(s_j \xrightarrow{a_i} s_k)\pi_l(s_j)}{\pi_l(s_k)} = K_{a_i} \tag{1}$$

with $l = 1, 2$ and π_l the steady-state solution of $\mathbf{M}_l^c = \{x_a \leftarrow K_a, \ \forall a \in \mathcal{P}(\mathbf{M}_l)\}$.

The following remark discusses the implication of GRCAT sufficient conditions for the product-form.

Remark 2 (*GRCAT Conditions*)**.** Observe that the GRCAT imposes two conditions: one on the passive and one on the active transitions. The former is just structural, and hence easy to check. The latter is more complicated since it requires that for every active transition, and for every state of an automaton, Eq. (1) must be satisfied. Testing this condition may require the solution of non-linear systems of equations because, in general, neither of the two automata could be closed. Nevertheless, we can observe that if $a \in \mathcal{A}(\mathbf{M})$ then each state of $\mathbf{M}$ must have at least one incoming transition labelled a. Finally, we point out that for $a \in \mathcal{A}(\mathbf{M})$ we can interpret K_a as the reversed incoming flow in all the states in $\mathcal{S}$ of $\mathbf{M}^c$ as noted in [3].

In the rest of the paper, we shall investigate models that violate one or more of the conditions specified above.

3. Product-form approximations

In this section, we illustrate an algorithmic technique to approximate a non-product-form model by means of a product-form one.

We consider the cases where one or more of the conditions stated in Section 2.2 are not satisfied in the cooperating LMAs. Conditions of GRCAT cannot be satisfied in two ways:

1. For some synchronising label $a \in \mathcal{P}(\mathbf{M}_i)$ with $i = 1, 2$ there exists a subset of states of the state space $\mathcal{S}_i$ from which there are no outgoing passive transitions labelled a. This means that Condition 1 of GRCAT is not satisfied.
2. There exists a label $a \in \mathcal{A}(\mathbf{M}_i)$ with $i = 1, 2$ such that the reversed incoming flow is not constant for transitions labelled a and for all the states, i.e., Eq. (1) of Condition 2 is not satisfied.

In what follows, we use $\tilde{\mathbf{M}}$ to denote automaton $\mathbf{M}$ modified in order to satisfy the GRCAT product-form conditions. Analogously, we adopt the notation $\tilde{\pi}_l$ and $\tilde{K}_a$ to denote respectively the steady-state probability distribution and the constant reversed incoming flow due to an active label a in the approximate automaton. Moreover, for sake of compactness we write $\mathcal{A}_i$, $\mathcal{P}_i$ instead of $\mathcal{A}(\mathbf{M}_i)$ and $\mathcal{P}(\mathbf{M}_i)$. In the following two sections we introduce the main results of this work, i.e., the algorithms for perturbation of cooperating models to meet GRCAT product-form conditions. For clarity, we illustrate our approach on pairs of models with one synchronising label. Generalisation to synchronisation with multiple labels or with a finite number of synchronising automata should not add any technical difficulty.

3.1. Missing passive outgoing transitions

We consider the case in which an automaton does not satisfy Condition 1 of GRCAT. Let us consider $\mathbf{M}_1 \oplus_{\{a\}} \mathbf{M}_2$, with $a \in \mathcal{A}_1 \cap \mathcal{P}_2$, where $\mathbf{M}_2$ does not satisfy Condition 1 of GRCAT, i.e. in each state of the automaton there must be an outgoing passive transition. Then, let $\mathcal{S}_2$ be the state-space of $\mathbf{M}_2$ and we define the set of all the states s without an outgoing passive transition labelled a:

$$\mathcal{S}_2^{\neg a} = \{s \in \mathcal{S}_2 : \forall s' \in \mathcal{S}_2,\ (s, a, x_a, s') \not\in \rightarrow_2\}.$$

We aim to obtain the joint-model $\tilde{\mathbf{M}}_1 \oplus_{\{a\}} \tilde{\mathbf{M}}_2$ such that:

$$\tilde{\pi}(\tilde{\mathbf{M}}_1 \oplus_{\{a\}} \tilde{\mathbf{M}}_2) \approx \pi(\mathbf{M}_1 \oplus_{\{a\}} \mathbf{M}_2).$$

$\tilde{\mathbf{M}}_2$ differs from $\mathbf{M}_2$ because for all $s \in \mathcal{S}_2^{\neg a}$ a self-loop passive transition with label a is added. Hence, $\tilde{\mathbf{M}}_2$ satisfies Condition 1 of GRCAT.

If $\mathbf{M}_1$ satisfies Condition 2 for label a, then $\mathbf{M}_1 \oplus_{\{a\}} \tilde{\mathbf{M}}_2$ is in product-form. Yet, the approximation $\mathbf{M}_1 \oplus_{\{a\}} \tilde{\mathbf{M}}_2$ can be quite inaccurate, as transitions labelled a are observed in $\mathbf{M}_1 \oplus_{\{a\}} \tilde{\mathbf{M}}_2$ more frequently than in $\mathbf{M}_1 \oplus_{\{a\}} \mathbf{M}_2$. In fact, the introduction of self-loop transitions in $\tilde{\mathbf{M}}_2$ is key in allowing $\mathbf{M}_1$ to perform all active transitions labelled a.

To improve the accuracy of the approximation we propose to modify also $\mathbf{M}_1$ to $\tilde{\mathbf{M}}_1$ by the system of equations below:

$$\begin{cases} \tilde{q}_1(s \xrightarrow{a} s') = q_1(s \xrightarrow{a} s') \displaystyle\sum_{s' \not\in \mathcal{S}_2^{\neg a}} \tilde{\pi}_2(s') & \text{for all transitions labelled } a \\ \tilde{K}_a = \left(\displaystyle\sum_{s_j \in \mathcal{S}_1} \tilde{q}_1(s_j \xrightarrow{a} s_k) \tilde{\pi}_1(s_j) \right) / \tilde{\pi}_1(s_k) & \text{for all } s_k \in \mathcal{S}_1, \end{cases} \tag{2}$$

where $\tilde{\pi}_1$ and $\tilde{\pi}_2$ are the steady-state distributions of $\tilde{\mathbf{M}}_1$ and $\tilde{\mathbf{M}}_2^c = \tilde{\mathbf{M}}_2\{a \leftarrow \tilde{K}_a\}$, respectively. The solution of the system of Eqs. (2) may be a hard task and numerical approaches may be required. However, in many relevant cases the described technique may be applied straightforwardly. For example, for reversible models, i.e., when $\mathbf{M}_1$ satisfies the condition:

$$\forall s, s' \in \mathcal{S}_1 \quad q(s', s)\pi_1(s') = \pi_1(s) q(s, s') \tag{3}$$

$\tilde{K}_a$ can be derived from the analysis of $\mathbf{M}_1$ which yields $\tilde{\mathbf{M}}_2^c = \tilde{\mathbf{M}}_2\{x_a \leftarrow K_a\}$. Note that K_a, differently from $\tilde{K}_a$, can be derived solely from Eq. (3) which refers to the rates of $\mathbf{M}_1$. In any case, once K_a has been determined, it becomes possible to compute $\pi_2^c(\tilde{\mathbf{M}}_2^c)$, i.e., the steady-state distribution of $\tilde{\mathbf{M}}_2^c$. Finally, we modify $\tilde{\mathbf{M}}_1$ by replacing all the rates of the transitions $s \xrightarrow{a} s'$ with $\tilde{q}_1(s \xrightarrow{a} s')$ according to the first equation of (2). Changes in the rates of the active transitions in $\tilde{\mathbf{M}}_1$ follow the idea that if the self-loops are added to states with low equilibrium probability, then $\tilde{q}_1(s \xrightarrow{a} s') \simeq q_1(s \xrightarrow{a} s')$. The bigger is the value of $\epsilon = q_1(s \xrightarrow{a} s') - \tilde{q}_1(s \xrightarrow{a} s')$, the worse is the approximation.

In general, if $\mathbf{M}_1 \oplus_L \mathbf{M}_2$ we defined $\tilde{\mathbf{M}}_i = \langle \mathcal{S}_i, Act, \tilde{\rightarrow}_i \rangle$ with $i = 1, 2$ as

$$\tilde{\rightarrow}_1 = \rightarrow_1 \setminus \left(\bigcup_{a \in L} \{s \xrightarrow{a,\lambda} s' : s \in \mathcal{S}_1\} \right) \cup \left(\bigcup_{a \in L} \{s \xrightarrow{a, \tilde{q}_1(s \xrightarrow{a} s')} s' : s \in \mathcal{S}_1\} \right)$$

$$\tilde{\rightarrow}_2 = \rightarrow_2 \cup \left(\bigcup_{a \in L} \{s \xrightarrow{a,x} s : s \in \mathcal{S}_2^{\neg a}\} \right).$$

The pseudo-code algorithm to transform $\mathbf{M}_1 \oplus_L \mathbf{M}_2$ into the approximation $\tilde{\mathbf{M}}_1 \oplus_L \tilde{\mathbf{M}}_2$ is depicted in Algorithm 1. The algorithm assumes that the reversed rate of the active transitions is independent of their forward rate. If this condition is not satisfied, steps from 2 to 11 should be replaced by the numerical solution of system of Eqs. (2).

Algorithm 1 The pseudo-code to transform $\mathbf{M}_1 \oplus_L \mathbf{M}_2$ into $\tilde{\mathbf{M}}_1 \oplus_L \tilde{\mathbf{M}}_2$

Start with: $\mathbf{M}_1 = \langle \mathcal{S}_1, Act_1, \to_1 \rangle$ such that $\forall s, s' \in \mathcal{S}_1 \quad \frac{q(s',s)\pi_2(s')}{\pi_2(s)} = q(s,s')$, $\mathbf{M}_2 = \langle \mathcal{S}_2, Act_2, \to_2 \rangle$, $\mathbf{M}_1 \oplus_{\{a_1 \dots a_N\}} \mathbf{M}_2$ is irreducible and $\{a_1 \dots a_N\} \in \mathcal{A}_1 \cap \mathcal{P}_2$;
Calculate $\pi_1(\mathbf{M}_1)$ and K_{a_i};
// Calculate $\tilde{\to}_2$
for all $i = 1 \dots N$ **do**
  Calculate $\mathcal{S}_2^{\neg a_i}$;
  Set $\to_2 = \to_2 \cup \{s \xrightarrow{a_i, x} s : s \in \mathcal{S}_2^{\neg a_i}\}$;
end for
// Calculate $\mathbf{M}^c$
for all $i = 1 \dots N$ **do**
  Calculate $\mathbf{M}_2^c = \mathbf{M}_2\{x_{a_i} \leftarrow K_{a_i}\}$;
end for
Calculate $\pi_2^c(\mathbf{M}_2^c)$;
for all $s \in \mathcal{S}_1$ **do**
  Set $\tilde{q}_1(s \xrightarrow{a} s') = \left(1 - \sum_{s_2 \in \mathcal{S}_2^{\neg a}} \pi_2^c(s_2)\right) q_1(s \xrightarrow{a} s')$;
end for
// Calculate $\tilde{\to}_1$;
for all $i = 1 \dots N$ **do**
  Set $\to_1 = \to_1 \setminus \{s \xrightarrow{a_i, \lambda} s' : s \in \mathcal{S}_1\} \cup \{s \xrightarrow{a_i, \tilde{q}_1(s \xrightarrow{a_i} s')} s' : s \in \mathcal{S}_1\}$;
end for
return $\tilde{\mathbf{M}}_1 = \langle \mathcal{S}_1, Act_1, \to_1 \rangle$, $\tilde{\mathbf{M}}_2 = \langle \mathcal{S}_2, Act_2, \to_2 \rangle$ and $\tilde{\mathbf{M}}_1 \oplus_{\{a_1 \dots a_N\}} \tilde{\mathbf{M}}_2$.

3.2. Different reversed incoming flow into states

In this section, we propose a method to approximate models in which Condition 2 is not satisfied, i.e., when the reversed incoming flow of an active label is not constant for all the states.

Consider two LMAs $\mathbf{M}_1$ and $\mathbf{M}_2$ synchronising only on label a, with $a \in \mathcal{A}_1 \cap \mathcal{P}_2$. Let $\mathcal{S}_1$ be the state-space of $\mathbf{M}_1$, then we define for all $s \in \mathcal{S}_1$:

$$K_a(s) = \sum_{s' \in \mathcal{S}_1} \frac{\pi_1(s')}{\pi_1(s)} q_1(s' \xrightarrow{a} s). \tag{4}$$

Eq. (4) is the total reverse flow into state s with respect to a transition labelled a. If $K_a(s)$ is independent of s, then GRCAT Condition 2 would be satisfied, otherwise we define:

$$\tilde{K}_a = \sum_{s \in \mathcal{S}_1} K_a(s)\pi_1(s) = \sum_{s \in \mathcal{S}_1} \sum_{s' \in \mathcal{S}_1} \pi_1(s') q_1(s' \xrightarrow{a} s). \tag{5}$$

We can see $\tilde{K}_a$ as the weighted average reversed flux incoming to the states of $\mathbf{M}_1$ due to transitions labelled a. We aim at obtaining $\tilde{\mathbf{M}}_1$ in a way such that $\tilde{\pi}_1 = \pi_1$. Let us define the following sets:

- $\mathcal{S}_1^{a,<\tilde{K}_a} = \{s \in \mathcal{S}_1 : K_a(s) < \tilde{K}_a\}$, i.e., the set of states of $\mathbf{M}_1$ whose incoming reversed flux due to active transitions labelled a is lower than the average;
- $\mathcal{S}_1^{a,>\tilde{K}_a} = \{s \in \mathcal{S}_1 : K_a(s) > \tilde{K}_a\}$, i.e., the set of states of $\mathbf{M}_1$ whose incoming reversed flux due to active transitions labelled a is higher that the average;
- $\mathcal{S}_1^{a,=\tilde{K}_a} = \{s \in \mathcal{S}_1 : K_a(s) = \tilde{K}_a\}$, i.e., the set of states of $\mathbf{M}_1$ whose incoming reversed flux due to active transitions labelled a is exactly $\tilde{K}_a$.

We now derive $\tilde{\mathbf{M}}_1$ from $\mathbf{M}_1$ by modifying the rates in the following way for each state $s \in \mathcal{S}$:

- ($s \in \mathcal{S}_1^{a,<\tilde{K}_a}$): Add a self-loop transition labelled a with rate $\tilde{q}_1(s \xrightarrow{a} s) = \tilde{K}_a - K_a(s)$.
- ($s \in \mathcal{S}_1^{a,>\tilde{K}_a}$): Rate $q_1(s' \xrightarrow{a} s)$ is replaced by

$$\tilde{q}_1(s' \xrightarrow{a} s) = \frac{\tilde{K}_a}{K_a(s)} q(s' \xrightarrow{a} s). \tag{6}$$

Intuitively, Eq. (6) means that the reversed incoming flow to state s is reduced to $\tilde{K}_a$ by slowing down all the synchronising transitions labelled a with the same proportion $\tilde{K}_a/K_a(s)$. To keep the same steady-state distribution, a new non-synchronising transition $s' \xrightarrow{c} s$ with $c \neq a$ is added. The new rate is described by the following expression:

$$q_1^\star(s' \xrightarrow{c} s) = q_1(s' \xrightarrow{a} s) - \tilde{q}_1(s' \xrightarrow{a} s). \tag{7}$$

- $(s \in \mathcal{S}_1^{a,=\tilde{K}_a})$: The rates of the transition in the subset of states $\mathcal{S}_1^{a,=\tilde{K}_a}$ remain the same.

Note that the steady-state probabilities of the chain underlying $\tilde{\mathbf{M}}_1$ are identical to those of $\mathbf{M}_1$. However, the key idea of the previous modifications is to make the sum of the reversed rates incoming to each state of $\tilde{\mathbf{M}}_1$ constant, in order to satisfy Eq. (1). Informally, this is achieved by augmenting the sum by self-loops when the states have an incoming reversed flow which is lower than expected. Conversely, when the reversed flow of the incoming transitions is higher than $\tilde{K}_a$, the desired value is achieved by splitting the transitions into a synchronising one and a non-synchronising one. The rates are opportunely assigned so that the sum of the forward rates remains unchanged. In general, if $\mathbf{M}_1 \oplus_L \mathbf{M}_2$, such that $\forall a \in L$ it holds that $a \in \mathcal{A}_1 \cap \mathcal{P}_2$ we defined $\tilde{\mathbf{M}}_1 = \langle \mathcal{S}_1 Act, \tilde{\rightarrow}_1\rangle$

$$\begin{aligned}\tilde{\rightarrow}_1 = \rightarrow_1 &\cup \left(\bigcup_{a\in L}\{s \xrightarrow{a,r_1} s : s \in \mathcal{S}_1^{a,<\tilde{K}_a}\}\right) \setminus \left(\bigcup_{a\in L}\{s' \xrightarrow{a,\lambda} s : s \in \mathcal{S}_1^{a,>\tilde{K}_a}\}\right)\\ &\cup \left(\bigcup_{a\in L}\{s' \xrightarrow{a,r_2} s : s \in \mathcal{S}_1^{a,>\tilde{K}_a}\}\right) \cup \{s \xrightarrow{c,r_3} s' : s \in \mathcal{S}_1^{a,>\tilde{K}_a}, c \notin L\}\end{aligned} \tag{8}$$

where $r_1 = \tilde{K}_a - K_a(s)$, $r_2 = \tilde{q}_1(s' \xrightarrow{a} s)$, $r_3 = q_1^\star(s' \xrightarrow{c} s)$.

As explained earlier, this transformation does not change the steady-state probabilities of $\mathbf{M}_1$, as the following proposition states.

Proposition 1. *Let $\tilde{\mathbf{Q}}$ and $\mathbf{Q}$ be the two generator matrices of the two automata $\tilde{\mathbf{M}} = \langle \mathcal{S}_1, Act, \tilde{\rightarrow}\rangle$ and $\mathbf{M} = \langle \mathcal{S}_1, Act, \rightarrow\rangle$, where $\tilde{\rightarrow}$ has been derived as described in Eq. (8). Then $\tilde{\mathbf{Q}} = \mathbf{Q}$.*

Proof. For $s \in \mathcal{S}_1$ and $s' \neq s$, the entries of the non-diagonal elements of the generator matrix of $\tilde{\mathbf{M}}$ are the following:

$$\begin{aligned}\tilde{q}(s \to s') = &\sum_{a\in Act} q(s \xrightarrow{a} s')\mathbf{1}[s' \in \mathcal{S}_1^{a,<\tilde{K}_a} \vee s \in \mathcal{S}_1^{a=\tilde{K}_a}] + \sum_{a\in Act} \tilde{q}(s \xrightarrow{a} s')\mathbf{1}[s' \in \mathcal{S}_1^{s\in\mathcal{S}_1^{a>\tilde{K}_a}}]\\ &+ \sum_{a\in Act} q^\star(s \xrightarrow{a} s')\mathbf{1}[s' \in \mathcal{S}_1^{a>\tilde{K}_a}]\end{aligned}$$

where $\mathbf{1}[\cdot]$ is the indicator function. By Eq. (7) we derive:

$$\begin{aligned}\tilde{q}(s \to s') = &\sum_{a\in Act} q(s \xrightarrow{a} s')\mathbf{1}[s' \in \mathcal{S}_1^{a,<\tilde{K}_a} \vee s \in \mathcal{S}_1^{a=\tilde{K}_a}] + \sum_{a\in Act} \tilde{q}(s \xrightarrow{a} s')\mathbf{1}[s' \in \mathcal{S}_1^{s\in\mathcal{S}_1^{a>\tilde{K}_a}}]\\ &+ \sum_{a\in Act} (q(s' \xrightarrow{a} s) - \tilde{q}(s' \xrightarrow{a} s))\mathbf{1}[s' \in \mathcal{S}_1^{a>\tilde{K}_a}].\end{aligned}$$

The result follows by observing that the subsets $\mathcal{S}_1^{a,<\tilde{K}_a}$, $\mathcal{S}_1^{a=\tilde{K}_a}$, $\mathcal{S}_1^{a>\tilde{K}_a}$ form a partition of the state space $\mathcal{S}_1$. □

The pseudo-code algorithm to transform $\mathbf{M}_1 \oplus_L \mathbf{M}_2$ into the approximation $\mathbf{M}_1 \oplus_L \tilde{\mathbf{M}}_2$ is depicted in Algorithm 2.

4. Applications

In this section, we investigate stochastic models consisting of two or three components whose interaction is known not to yield a product-form stationary distribution.

4.1. Simple network with blocking

We study an example where it is possible to apply Algorithm 1. Let us consider the tandem of queues depicted by Fig. 1(A). Customers arrive at the first queue according to a Poisson process with rate λ. Customer's service time is exponentially distributed with rate μ_1, and the discipline of service is first-come-first-serve (FCFS). At a job completion, the customer enters the second queue, if there is space in its buffer, or waits in the first queue according to a Repetitive Service (RS) policy (see, e.g., [12] for details of this blocking discipline). In the second queue, the service time is distributed according to an exponential random variable with rate μ_2. Fig. 1(B) shows the automata $\mathbf{M}_1$ and $\mathbf{M}_2$ underlying the two queues in isolation.

Algorithm 2 The pseudo-code to transform $\mathbf{M}_1 \oplus_L \mathbf{M}_2$ into $\tilde{\mathbf{M}}_1 \oplus_L \tilde{\mathbf{M}}_2$

Start with: $\mathbf{M}_1 = \langle \mathcal{S}_1, Act_1, \to_1 \rangle$, $\mathbf{M}_2 = \langle \mathcal{S}_2, Act_2, \to_2 \rangle$, $\mathbf{M}_1 \oplus_{\{a_1 \ldots a_N\}} \mathbf{M}_2$ and $\{a_1 \ldots a_N\} \in \mathcal{A}_1 \cap \mathcal{P}_2$;
Calculate $\pi_1(\mathbf{M}_1)$;
// Calculate $\tilde{\to}_1$
for all $i = 1 \ldots N$ **do**
  Calculate $K_{a_i}(s) = \sum_{s' \in \mathcal{S}_1} \frac{\pi_1(s')}{\pi_1(s)} q_1(s' \xrightarrow{a_i} s)$;
  Calculate $\tilde{K}_{a_i} = \sum_{s \in \mathcal{S}_1} K_{a_i}(s)$;
  Calculate $\mathcal{S}_1^{a_i, <\tilde{K}_{a_i}}$;
  Calculate $\mathcal{S}_1^{a_i, >\tilde{K}_{a_i}}$;
  for all $s \in \mathcal{S}_1$ **do**
    Select $c \in Act_1 \backslash (Act_1 \cup L)$
    Calculate $\tilde{q}_1(s' \xrightarrow{a_i} s) = \frac{\tilde{K}_{a_i}}{K_{a_i}(s)} q(s' \xrightarrow{a_i} s)$;
    Calculate $q_1^\star(s' \xrightarrow{c} s) = q_1(s' \xrightarrow{a_i} s) - \tilde{q}_1(s' \xrightarrow{a_i} s)$;
  end for
  Set

$$\begin{aligned} \to_1 \quad = \quad & \to_1 \cup (\{s \xrightarrow{a_i, \tilde{K}_{a_i} - K_{a_i}(s)} s : s \in \mathcal{S}_1^{a_i, <\tilde{K}_{a_i}}\}) \backslash (\{s' \xrightarrow{a_i, \lambda} s : s \in \mathcal{S}_1^{a_i, >\tilde{K}_{a_i}}\}) \\ & \cup (\{s' \xrightarrow{a_i, \tilde{q}_1(s' \xrightarrow{a_i} s)} s : s \in \mathcal{S}_1^{a_i, >\tilde{K}_{a_i}}\}) \\ & \cup \{s \xrightarrow{c, q_1^\star(s' \xrightarrow{c} s)} s' : s \in \mathcal{S}_1^{a, >\tilde{K}_a}, c \neq a_1 \ldots a_N\}; \end{aligned}$$

end for
return $\tilde{\mathbf{M}}_1 = \langle \mathcal{S}_1, Act_1, \to_1 \rangle$ and $\tilde{\mathbf{M}}_1 \oplus_{\{a_1 \ldots a_N\}} \mathbf{M}_2$.

Observe that GRCAT conditions are not satisfied since state B_2 of $\mathbf{M}_2$ does not exhibit an outgoing passive transition labelled a. According to our approach, we have $\mathcal{S}_2^{\neg a} = \{B_2\}$. To compute $\tilde{\pi}_2$ we must know the value of K_a. We exploit the fact that $\mathbf{M}_1$ is reversible, i.e., satisfies Eq. (3) and therefore $K_a = \lambda$. Hence, $\mathbf{M}_2^c = \mathbf{M}_2\{x_a \leftarrow K_a\}$ is defined straightforwardly without the need of solving system (2). The stationary probability $\tilde{\pi}_2$ can be finally computed to obtain the approximation $\tilde{\mu}_1$ for the definition of $\mathbf{M}_1$. The resulting models are shown in Fig. 1(C).

4.2. Network with Coxian service and state-dependent arrivals

We consider a tandem of queues as depicted by Fig. 2. The first queue of the tandem has a FCFS discipline with Coxian service time distribution consisting of two exponential stages with possibly different rates. In this example, we assume, as a simplifying hypothesis, that the first exponential stage has rate $\mu_1 + \mu_2$, while the second μ_2. The probability for a customer to join the second stage after a job completion is $\mu_1/(\mu_1 + \mu_2)$ while with probability $\mu_2/(\mu_1 + \mu_2)$ the customer immediately enters the second queue. Customers arrive from the outside according to a Poisson process with rate λ. Arrivals when the second phase is busy are discarded. The state-space of this queue in isolation is depicted by Fig. 3(A). After being served, customers enter a standard./M/1 queue with exponential service time distribution with rate γ. The latter queue is modelled using passive transitions labelled a from state n to state $n+1$ (birth transitions), and non-synchronising active ones with rate γ from state $n+1$ to n (death transitions), with $n \geq 0$.

Since the second queue satisfies GRCAT conditions, only the first queue needs to be changed in order to obtain a product-form network. For the first queue, we derive the expression of the steady-state probability distribution by solving the following system of global balance equations (GBEs):

$$\begin{aligned} &\pi_1(0,0)\lambda = \pi_1(1,0)\mu_2 + \pi_1(1,1)\mu_2 \\ &\pi_1(n,0)[\lambda + \mu_1 + \mu_2] = \pi_1(n+1,0)\mu_2 + \pi_1(n+1,1)\mu_2 + \pi_1(n-1,0)\lambda \\ &\pi_1(n,1)\mu_2 = \pi_1(n,0)\mu_1 \quad n > 0. \end{aligned}$$

By solving the recurrence relation we obtain the following analytical solution:

$$\begin{aligned} &\pi_1(1,0) = \pi_1(0,0)\frac{\lambda}{\mu_1 + \mu_2} \\ &\pi_1(n,0) = \pi_1(0,0)\left(\frac{\lambda}{\mu_1 + \mu_2}\right)^n \\ &\pi_1(n,1) = \pi_1(0,0)\left(\frac{\lambda}{\mu_1 + \mu_2}\right)^n \frac{\mu_1}{\mu_2}. \end{aligned}$$

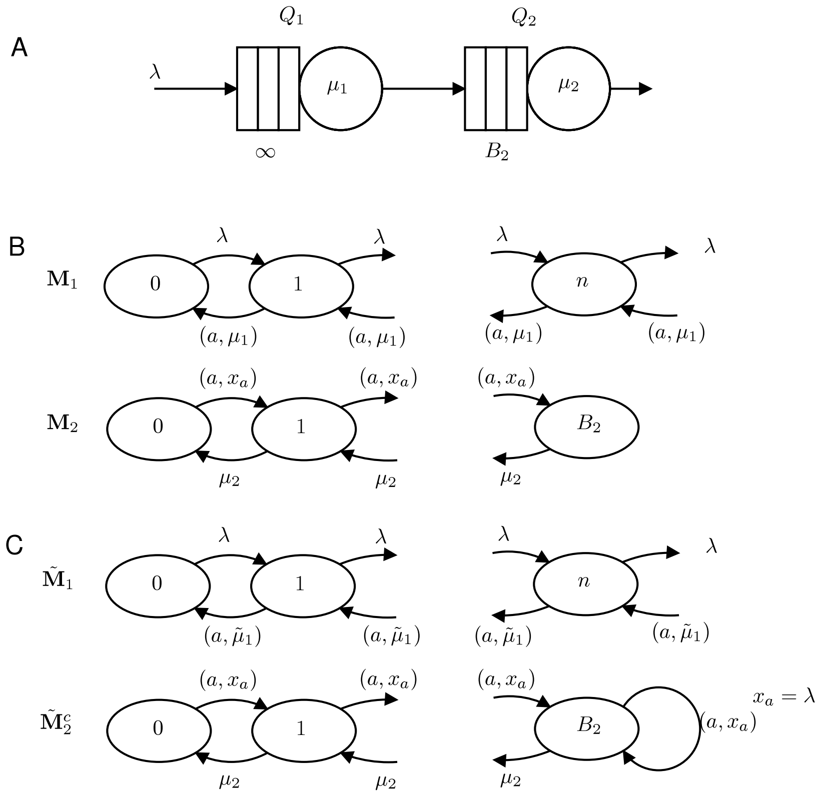


Fig. 1. Network composed of a tandem of queues. The first queue is an M/M/1 and the second queue is an./M/1 with finite capacity B_2. Figure (A) Network representation. Figure (B) State space diagram of the two queues as LMAs $\mathbf{M}_1$, $\mathbf{M}_2$. Figure (C) State space diagram of the LMAs $\tilde{\mathbf{M}}_1$, $\tilde{\mathbf{M}}_2$ of the approximated queues. If the variable x_a is substituted with λ in $\tilde{\mathbf{M}}_2$ then $\tilde{\mathbf{M}}_2^c$ is derived.

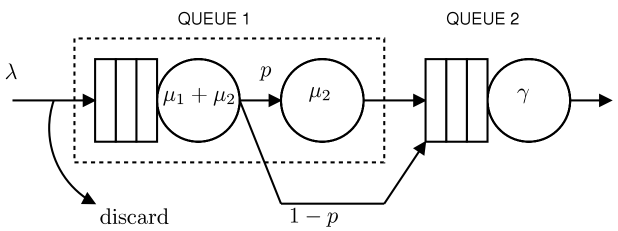


Fig. 2. Queue representation of network with Coxian service and state-dependent arrivals considered in Section 4.2.

By normalising, we find the value of $\pi_1(0, 0)$ as follows:

$$1 = \sum_{i=0}^{\infty} \pi_1(0, 0) \left[\frac{\lambda}{\mu_1 + \mu_2}\right]^i + \sum_{i=1}^{\infty} \pi_1(0, 0) \left[\frac{\lambda}{\mu_1 + \mu_2}\right]^i \frac{\mu_1}{\mu_2}$$

$$\pi_1(0, 0) = \frac{\mu_2(\mu_1 + \mu_2 - \lambda)}{(\mu_1 + \mu_2)\mu_2 + \lambda\mu_1}$$

with $\lambda < \mu_1 + \mu_2$. Once the steady-state distribution is known, the computation of the sum of the reversed rates of the transitions incoming to states $(n, 0)$ $(n \geq 0)$ is straightforward:

$$K_a(n, 0) = \frac{\pi_1(n + 1, 0)\mu_2 + \pi_1(n + 1, 0)\mu_2}{\pi_1(n, 0)} = \lambda. \tag{9}$$

We notice that although the GRCAT condition expressed by Eq. (1) is satisfied for states $(n, 0)$ with $n \neq 0$, states $(n, 1)$ have no active incoming transitions.

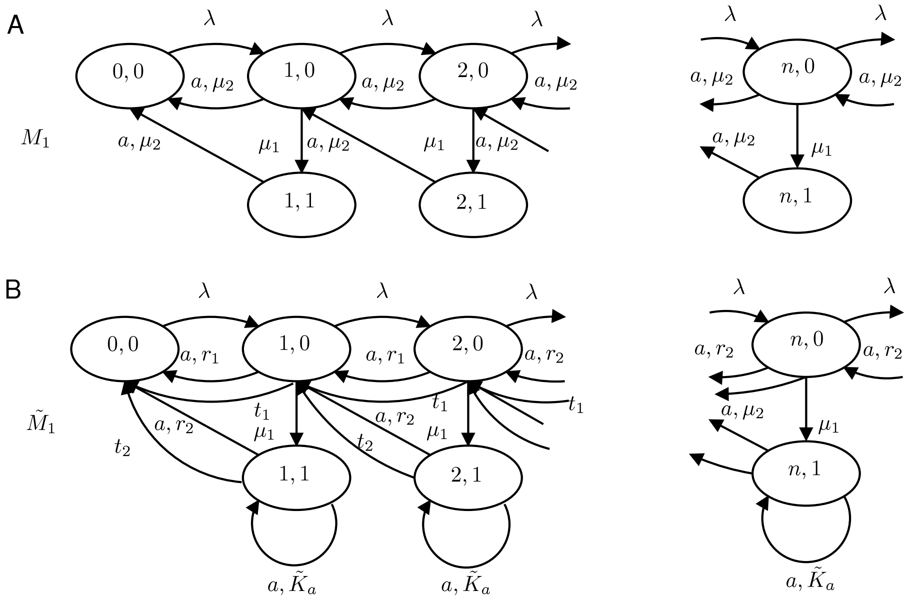


Fig. 3. Transition diagram of the queue with Coxian service time and state-dependent arrivals in the network considered in Section 4.2. (A) LMA representation of the original model. (B) LMA representation of the approximated model.

We therefore derive:

$$\tilde{K}_a = \sum_{n=0}^{\infty} \pi_1(n, 0) K_a(n, 0) + \sum_{n=1}^{\infty} \pi_1(n, 1) 0 = \frac{\lambda \mu_2(\mu_1 + \mu_2)}{\mu_1 \lambda + \mu_2(\mu_1 + \mu_2)}. \tag{10}$$

We observe that $\tilde{K}_a < K_a(n, 0)$ as $\sum_{n=0}^{\infty} \pi_1(n, 0) < 1$ since $\pi_1(n, 1) > 0$. This implies that $\mathcal{S}_1^{a,<\tilde{K}_a} = \{(n, 1) : n \geq 1\}$ and $\mathcal{S}_1^{a,>\tilde{K}_a} = \{(n, 0) : n \geq 0\}$.

Proposition 2. *Let $\mathbf{M}_1 = \langle \mathcal{S}_1, Act, \rightarrow \rangle$ be the automaton described in Fig. 3 (A). The transition relation $\tilde{\rightarrow}_1$ of the approximated automaton $\tilde{\mathbf{M}}_1 = \langle \mathcal{S}_1, Act_1, \tilde{\rightarrow}_1 \rangle$ is depicted in Fig. 3 (B) where r_1, r_2, t_1, t_2 have the following values:*

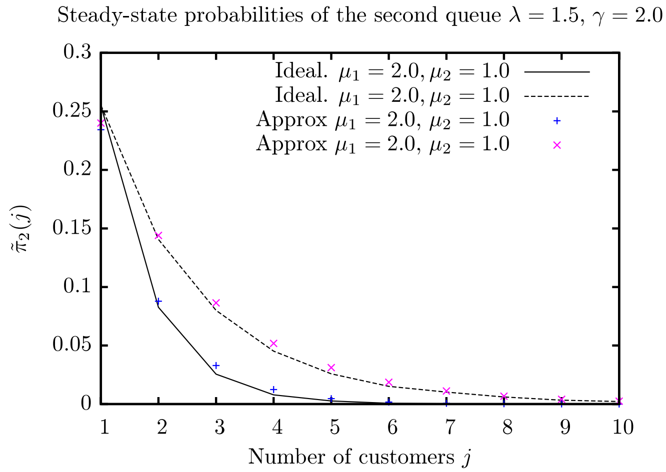
$$r_1 = r_2 = \frac{\mu_2^2(\mu_1 + \mu_2)}{\mu_1 \lambda + \mu_2(\mu_1 + \mu_2)} \qquad t_1 = t_2 = \frac{\mu_1 \mu_2 \lambda}{\mu_1 \lambda + \mu_2(\mu_1 + \mu_2)}.$$

Proof. By Eq. (6) we obtain r_1, r_2 and by Eq. (7) we obtain t_1, t_2. □

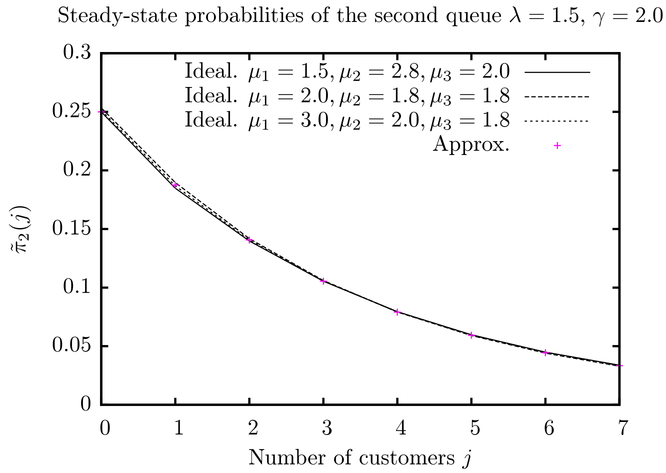
Observe that, if $\mu_1 \to 0$, then the first queue tends to assume the behaviour of a M/M/1 queue, and hence the product-form is satisfied without the need of approximations. According to this observation, we point out that $r_1 = r_2 = \mu_2$ and $t_1 = t_2 = 0$ as $\mu_1 \to 0$. Fig. 4(a) shows the steady-state distributions of the number of customers obtained with two sets of parameters in the exponential queue for the original tandem model and the approximating models obtained with the technique explained in Section 3.2. The exact values are obtained by the analysis of the truncation of the joint process since the load factor of the queues is kept low.

We now take a closer look at the behaviour of the whole network and its approximation. When the first queue is in state $(n, 1)$ the arriving customers are discarded. We define *the effective arrival rate*, λ_e, as the rate at which customers enter the queue. Clearly, it holds that $\lambda_e \leq \lambda$. In the original network, a customer that enters the first queue definitely enters also the second queue. In the approximated model of the network, some customers may leave the system before entering the second queue, and there may be some extra external arrivals directly at the second queue. The external arrivals are introduced by the cooperation with self-loops at states $(n, 1)$-see Fig. 3(B). The following lemma is important because it shows that, at steady-state, the effective arrival rate at the first queue is indeed $\tilde{K}_a$.

Lemma 1. *The effective arrival rate of the queue with state space described in Fig. 3(A) is $\tilde{K}_a$.*



(a) Steady-state distribution of the second queue of the network with Coxian service and state-dependent arrivals as studied in Section 4.2.



(b) Steady-state distribution of the second queue of the network with Coxian service and LCFS discipline as studied in Section 4.3.

Fig. 4. (a)—Steady-state distribution of the second queue of the network with Coxian service and state-dependent arrivals as studied in Section 4.2. The label (*Ideal*) denotes the original model and the label (*Approx.*) denotes the approximated model. (b)—Steady-state distribution of the second queue of the network with Coxian service and LCFS discipline as studied in Section 4.3. The label (*Ideal*) denotes the original model and the label (*Approx.*) denotes the approximated model.

Proof. We apply the PASTA property [13], i.e., customers arriving according to a Poisson process see time averages. Mathematically, the effective arrival rate λ_e is defined as

$$\lambda_e = \lambda \sum_{i=0}^{\infty} \pi_1(n, 0).$$

By Eq. (9) it holds that $K_a(n, 0) = \lambda$. Observing that the definition of λ_e is identical to Eq. (10) we obtain the result. □

In the approximated model, $\tilde{K}_a$ is the solution of the traffic equation or the rate we substitute in the LMA $\mathbf{M}_2^c = \mathbf{M}_2\{x_a \leftarrow \tilde{K}_a\}$. This implies that $\tilde{K}_a$ is the arrival rate at the second queue. We conclude that the effective arrival rates at the first and the second queue in the approximated network are the same. Table 1 shows a comparison of the mean number of customers in the second queue between the original and the approximating model. Since Little's law [14] can be applied to the second queue, one may derive from the average number of customers also the mean response time, considering that the throughput in stability is $\tilde{K}_a$.

4.3. Network with Coxian service and LCFS discipline

We consider another tandem of queues in which the first node has Coxian service time and LCFS scheduling discipline as depicted in Fig. 5. Pre-emption is assumed and when a job is pre-empted the time spent in service is lost. Note that, this

Table 1
Comparison of mean number of customers in the second queue of the network with Coxian service and state dependent arrivals as studied in Section 4.2. The table shows the comparison of the mean number of customers in the second queue for the exact and the approximated network.

Parameters	Exact N	Approx. N
$\lambda = 0.1, \mu_1 = 0.15, \mu_2 = 0.1, \gamma = 0.4$	0.177053	0.189708
$\lambda = 0.1, \mu_1 = 0.35, \mu_2 = 0.1, \gamma = 0.4$	0.153409	0.191314
$\lambda = 0.1, \mu_1 = 0.35, \mu_2 = 0.1, \gamma = 0.9$	0.0645057	0.0777937
$\lambda = 0.1, \mu_1 = 0.02, \mu_2 = 0.7, \gamma = 0.9$	0.124247	0.100385
$\lambda = 0.1, \mu_1 = 0.2, \mu_2 = 0.2, \gamma = 0.9$	0.0948834	0.0969089

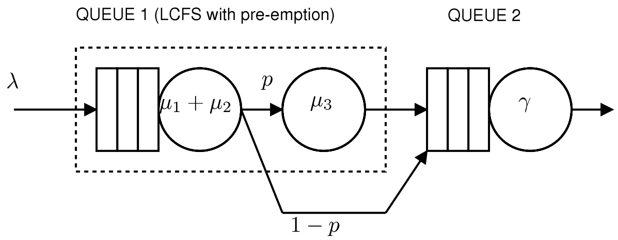


Fig. 5. Queue representation of the network with Coxian service time distribution and LCFS scheduling discipline as studied in Section 4.3.

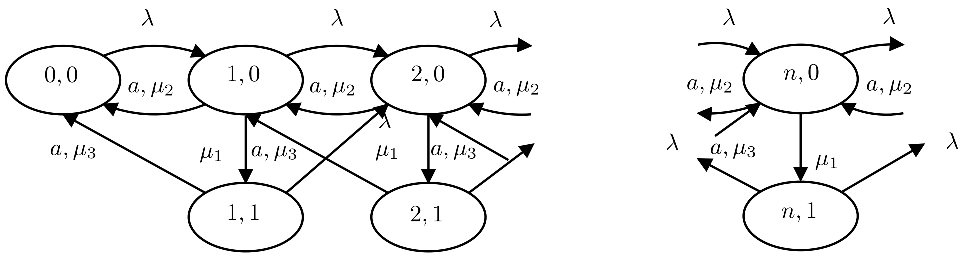


Fig. 6. Transition diagram of the queue with Coxian service time distribution and LCFS scheduling discipline (example analysed in Section 4.3).

queue is *not* the same as in [2]. In fact, it can be shown that the discipline is *not* symmetric [4], which is a necessary and sufficient condition for this type of queue to yield a product-form solution.

After a job completion at the first queue, customers enter another simple exponential queue whose service rate is γ. The state space for the first queue is depicted in Fig. 6. The GBEs for this queue are the following:

$$\pi_1(0, 0)\lambda = \pi_1(1, 0)\mu_2 + \pi_1(1, 1)\mu_3 \tag{11}$$

$$\pi_1(1, 0)[\lambda + \mu_1 + \mu_2] = \pi_1(2, 0)\mu_2 + \pi_1(2, 1)\mu_3 + \pi_1(0, 0)\lambda \tag{12}$$

$$\pi_1(1, 1)[\mu_3 + \lambda] = \pi_1(1, 0)\mu_1 \tag{13}$$

$$\pi_1(n, 0)[\lambda + \mu_1 + \mu_2] = \pi_1(n + 1, 0)\mu_2 + \pi_1(n + 1, 1)\mu_3 + \pi_1(n - 1, 0)\lambda + \pi_1(n - 1, 1)\lambda \quad \text{for } n > 0 \tag{14}$$

$$\pi_1(n, 1)[\mu_3 + \lambda] = \pi_1(n, 0)\mu_1. \tag{15}$$

For this model, the steady-state distribution of the Coxian queue can be computed by matrix geometrics techniques [15].

It is possible to verify that the sum of the reversed rates of the transitions incoming to states $(n, 0)$ with $n \geq 0$ are not constant (and active transitions labelled a incoming to states $(n, 1)$ are missing). Note that once the first queue is numerically solved the algorithm presented in Section 3.2 can be straightforwardly applied. $\tilde{K}_a = \sum_{i=0}^{\infty} \pi_1(n, 0)K_a$ can be numerically computed, and the approximated LMA $\tilde{\mathbf{M}}_1$ can be found.

In this example, we wish to focus the attention on what happens to the synchronisation of the first and the second queue. Adding self-loops to states $(n, 1)$ with $n > 0$, from the point of view of the exponential queue, means that an external arrival stream of customers is present when the first queue is in one of these states. On the other hand, splitting the synchronising transitions labelled a entering into states $(n, 0)$, $n \geq 0$, causes some customers to leave the system after a job completion instead of entering the second queue. An interesting problem is to compare the arrival rate at the second queue in the approximated model with that of the original one. In the latter case it is obvious that since there is no customer loss, or customer generation, the rate (in equilibrium) must be λ. The following lemmas show that the approximated model preserves the arrival rate at the second queue, i.e., $\tilde{K}_a = \lambda$.

Table 2
The table shows the comparison of mean number of customers in the second queue of the exact and approximate network with Coxian service and LCFS discipline as studied in Section 4.3. It is assumed $\lambda = 0.31$.

Parameters	Exact N	Approx. N
$\mu_1 = 1.5, \mu_2 = 2.8, \mu_3 = 2.0, \gamma = 2.0$	1.8372	1.7717
$\mu_1 = 0.01, \mu_2 = 0.8, \mu_3 = 0.4, \gamma = 0.9$	0.525	0.520
$\mu_1 = 0.001, \mu_2 = 0.35, \mu_3 = 0.4, \gamma = 2.9$	0.116	0.116
$\mu_1 = 0.01, \mu_2 = 0.4, \mu_3 = 0.8, \gamma = 1.2$	0.348	0.345
$\mu_1 = 0.9, \mu_2 = 0.9, \mu_3 = 0.4, \gamma = 1.2$	0.362	0.330

Lemma 2. *The steady-state probabilities π_1 of the Coxian queue depicted by Fig. 6 fulfil the following property:*

$$\begin{aligned} &\pi_1(i+1,0)\mu_2 + \pi_1(i+1,1)\mu_3 = \lambda[\pi_1(i,0) + \pi_1(i,1)] \\ &\pi_1(i+1,0)(\mu_1+\mu_2) = \lambda[\pi_1(i,0) + \pi_1(i,1) + \pi_1(i+1,1)] \end{aligned} \tag{16}$$

for all $i \geq 1$.

Proof. Eq. (16) is proven by induction. For $i = 1$, we take the GBE for state $(1, 0)$ given by Eq. (12) and replace expression $\pi_1(1, 0)\mu_1$ using Eq. (13). We obtain:

$$\pi_1(1,0)[\lambda + \mu_2] + \pi_1(1,1)[\mu_3 + \lambda] = \pi_1(2,0)\mu_2 + \pi_1(2,1)\mu_3 + \pi_1(0,0)\lambda.$$

By rearranging the terms we obtain:

$$\lambda[\pi_1(1,0) + \pi_1(1,1)] + \underbrace{\pi_1(1,0)\mu_2} + \underbrace{\pi_1(1,1)\mu_3} = \pi_1(2,0)\mu_2 + \pi_1(2,1)\mu_3 + \underbrace{\pi_1(0,0)\lambda}.$$

After observing that the under-braced terms correspond to the GBE of state $(0, 0)$ of Eq. (11) the proof for $i = 1$ is concluded.
For $i = n > 1$ we write done the GBE (14) for state $(n, 0)$:

$$\pi_1(n,0)[\lambda + \mu_1 + \mu_2] = \pi_1(n+1,0)\mu_2 + \pi_1(n+1,1)\mu_3 + \pi_1(n-1,0)\lambda + \pi_1(n-1,1)\lambda.$$

By substituting in the left-hand side $\pi_1(n, 0)\mu_1$ with the expression given by Eq. (15) and applying the induction step the result follows. Eq. (16) can be derived by substituting expression $\mu_3\pi_1(i, 1)$ with the left-hand side of GBE (13) in Eq. (16). This concludes the proof. □

Lemma 3. *The average reversed incoming flow to a state due to an action labelled a equals the rate of arrivals, $\tilde{K}_a = \lambda$, in the Coxian queue depicted in Fig. 6 with LCFS pre-emptive scheduling discipline.*

Proof. By Eq. (5) we must compute the following series:

$$\sum_{i=0}^{\infty}(\pi_1(i+1,0)\mu_2 + \pi_1(i+1,1)\mu_3).$$

Using Eq. (16) we straightforwardly have:

$$\pi_1(1,0)\mu_2 + \pi_1(1,1)\mu_3 + \sum_{i=1}^{\infty}\lambda(\pi_1(i,0) + \pi_1(i,1)).$$

Observe that by GBE (11) this may be conveniently rewritten as:

$$\lambda\left[\pi_1(0,0) + \sum_{i=1}^{\infty}(\pi_1(i,0) + \pi_1(i,1))\right] = \lambda. \quad \square$$

The approximate model is obtained from the original one by forcing the total reversed rate in each state to be $\tilde{K}_a$ without changing its steady-state distribution. $\tilde{K}_a$ is the arrival rate seen by the exponential queue which is at the end of the tandem.

Fig. 4(B) shows a comparison of the steady-state probability distribution of the number of customers in the exponential queue (recall that the Coxian model steady-state probabilities are not affected by the modifications). Table 2 illustrates the comparison of the average number of customers in the exponential queue in the original model (obtained as solution of a truncation of the infinitesimal generator of the joint process) and in the approximation.

4.4. Network with collisions

Model description

In this section we study a queueing network in which customer collisions may occur. Collision will cause the elimination of the customers involved. The topology is depicted in Fig. 7. Customers arrive at nodes 1 and 2 according to independent

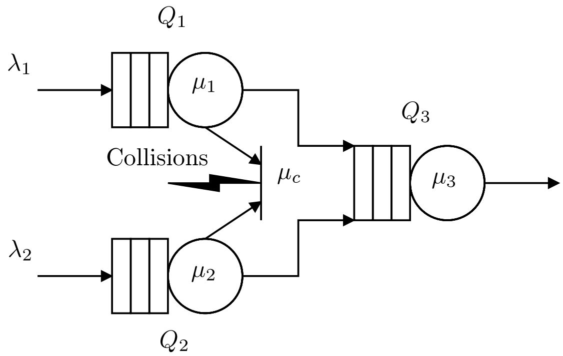


Fig. 7. Network with collisions studied in Section 4.4.

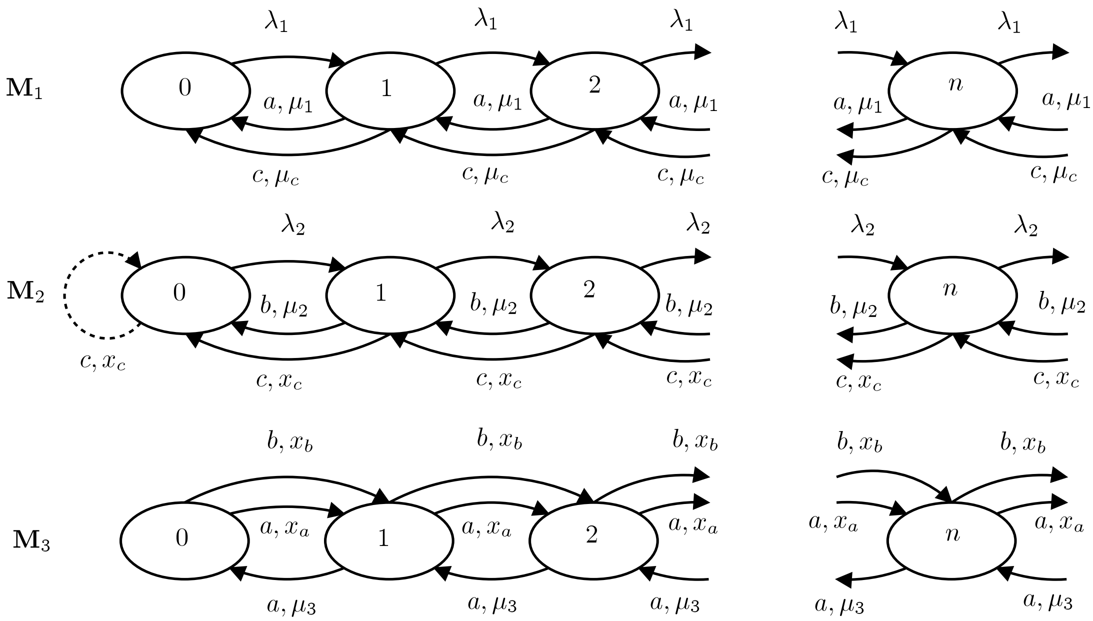


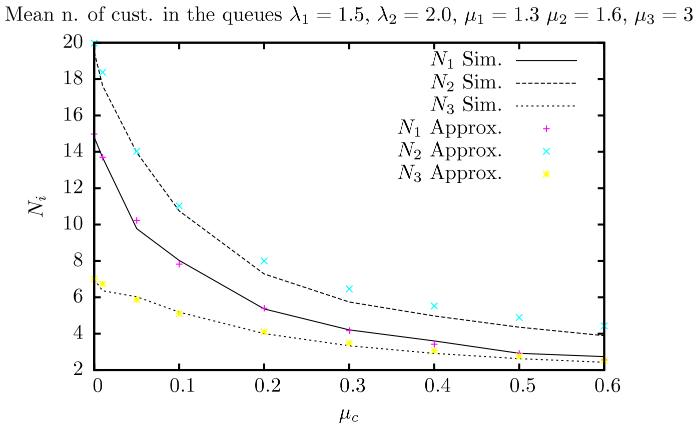
Fig. 8. Transition diagram as LMA of each single queue of network with collisions, studied in Section 4.4.

Poisson processes with rates λ_1 and λ_2, respectively. Service times at nodes 1 and 2 are independent and exponentially distributed with rates μ_1 and μ_2, respectively. After a job completion at one the first two queues, customers enter a third node whose service time is independent and exponentially distributed with rate μ_3. When the first two queues are both non-empty, collisions between the customers occur with rate μ_c. The collision simultaneously destroys one customer from each of these queues.

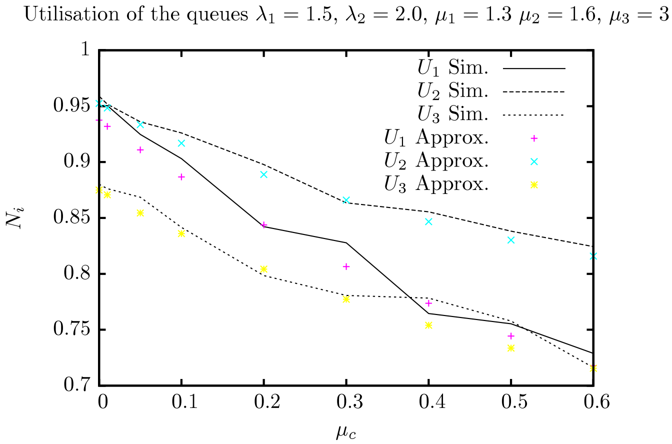
Model analysis

Fig. 8 illustrates the transition diagram of each queue (ignore the dashed self-loop). Observe that Condition 1 is not satisfied since state 0 in $\mathbf{M}_2$,corresponding to queue Q_2, has no passive outgoing transition labelled c. According to the algorithm given in Section 3.1 a passive self-loop in state 0 of $\mathbf{M}_2$ must be added to obtain $\tilde{\mathbf{M}}_2$. Observe that each queue is a birth-and-death process if each LMA is closed. This ensures that the reversed rates of the labelled death transitions are constant. Specifically, in $\tilde{\mathbf{M}}_1$ we have that $\tilde{K}_c = \lambda_1\tilde{\mu}_c/(\mu_1 + \tilde{\mu}_c)$ and in $\tilde{\mathbf{M}}_2$ we have $\mathcal{S}_2^{\neg c} = \{0\}$ and $1 - \sum_{s\in\mathcal{S}_2^{\neg c}} \tilde{\pi}_2(s) = \lambda_2/(\mu_2 + \tilde{K}_c)$. Therefore, the system of Eqs. (2) in this example is:

$$\begin{cases} \tilde{K}_c = \lambda_1 \dfrac{\tilde{\mu}_c}{\mu_1 + \tilde{\mu}_c} \\ \tilde{\mu}_c = \dfrac{\lambda_2}{\mu_2 + \tilde{K}_c}\mu_c \end{cases}$$



(a) Network with collisions as analysed in Section 4.4. The graph shows the expected number of customers.



(b) Network with collisions as analysed in Section 4.4. The graph shows the utilisation.

Fig. 9. (a) Average number of customers in the queues of the network with collisions as studied in Section 4.4. (b) Utilisation of the nodes of network with collisions as in the example considered in Section 4.4. The results for the original model are obtained by simulation. The results for the approximated model are based on the analytical solution of the product-form.

whose solution can be obtained symbolically:

$$\tilde{\mu}_c = \frac{-\mu_1\mu_2 + \lambda_2\mu_c + \sqrt{\Delta}}{2(\lambda_1 + \mu_2)}$$
$$\tilde{K}_c = \lambda_1 - \frac{2\lambda_1\mu_1(\lambda_1 + \mu_2)}{2\lambda_1\mu_1 + \mu_1\mu_2 + \lambda_2\mu_c + \sqrt{\Delta}},$$

where $\Delta = 4\lambda_2\mu_1(\lambda_1 + \mu_2)\mu_c + (\mu_1\mu_2 - \lambda_2\mu_c)^2$. Observe that, if $\mu_c \to 0$ then the model is equivalent to a Jackson's network in product-form, and coherently the approximation gives $\tilde{\mu}_c = 0$. Given the product-form joint-model $\tilde{\mathbf{M}}_1 \oplus_{\{c\}} \tilde{\mathbf{M}}_2$, we observe that the reversed rates of the transitions labelled a and b are constant and that $\mathbf{M}_3$ satisfies Condition 1 of GRCAT. Therefore the product-form of $(\tilde{\mathbf{M}}_1 \oplus_{\{c\}} \tilde{\mathbf{M}}_2) \oplus_{\{a,b\}} \mathbf{M}_3$ is obtained without applying the approximating algorithms. We have $\tilde{K}_a = \lambda\mu_1/(\mu_1 + \tilde{\mu}_c)$ and $\tilde{K}_b = \lambda_2\mu_2/(\mu_2 + \tilde{K}_c)$. The steady-state distribution is:

$$\pi(n_1, n_2, n_3) \approx \tilde{\pi}(n_1, n_2, n_3)$$
$$= \left(1 - \frac{\lambda_1}{\mu_1 + \tilde{\mu}_c}\right)\left(1 - \frac{\lambda_2}{\mu_2 + \tilde{K}_c}\right)\left(1 - \frac{\tilde{K}_a + \tilde{K}_b}{\mu_3}\right)\left(\frac{\lambda_1}{\mu_1 + \tilde{\mu}_c}\right)^{n_1} \cdot \left(\frac{\lambda_2}{\mu_2 + \tilde{K}_c}\right)^{n_2}\left(\frac{\tilde{K}_a + \tilde{K}_b}{\mu_3}\right)^{n_3}$$

where n_i denotes the number of customers in node $i = 1, 2, 3$. Fig. 9 shows the comparison between the average number of customers and the utilisation in the approximated model and in the original one. For this latter case, the results have been obtained by simulation with confidence intervals of 95% and maximum interval width of 0.5 and 0.08 for the graphs of Fig. 9(A) and (B), respectively.

5. Conclusion

In this paper, we have investigated approximations of non-product-form models with product-form ones. We have taken a practical approach and considered several examples. The method presented in Sections 3.1 and 3.2 exploits the formulation of GRCAT given in [9]. The main strengths of the proposed approach are:

- Modularity: the perturbations may be interpreted for each component in isolation;
- Generality: the algorithms are defined in terms of cooperating LMAs;
- Symbolic analysis: the method allows the modeller to derive an approximation of the steady-state distribution both numerically and analytically.

With our algorithms it is possible to study open models with infinite state-spaces and perform sensitivity analysis efficiently.

We have considered four case studies, and we have analysed the results of the approximations. The estimates of our approximation are showed to be relatively accurate with respect to the exact solution, in our case studies. One can expect that stronger perturbations in the models will give worse approximations (see the example of Section 4.4). As future research efforts are concerned, we shall investigate the problem of the analytical definition of bounds for the errors on the performance measures introduced by the proposed approach.

Appendix. Formal treatment of the theoretical background

Definition of LMA can be found in Section 2. LMAs can be seen as CTMCs with labelled transitions. The definition of LMA has been inspired by PEPA [16]. The definition of two interacting automata can be found below.

Definition 2. Let $\mathbf{M}_1 = \langle \mathcal{S}_1, Act_1, \rightarrow_1 \rangle$ and $\mathbf{M}_2 = \langle \mathcal{S}_2, Act_2, \rightarrow_2 \rangle$ be two LMAs.

The *interacting* LMA $\mathbf{M}_1 \oplus_L \mathbf{M}_2 = \langle \mathcal{S}, Act, \rightarrow, \rangle$ with $L \subseteq Act_1 \cap Act_2$ is a new automata defined as follows:

- $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$.
- $Act = Act_1 \cup Act_2$.
- $\rightarrow$ is the smallest relation defined by the rules below:

$$\frac{s_1 \xrightarrow{a,\lambda}_1 s_1' \quad s_2 \xrightarrow{a,x_a}_2 s_2'}{(s_1, s_2) \xrightarrow{a,\lambda} (s_1', s_2')}(a \in L) \qquad \frac{s_1 \xrightarrow{a,r}_1 s_1'}{(s_1, s_2) \xrightarrow{a,r} (s_1', s_2)}(a \notin L).$$

The symmetric rules are omitted.

The use of variables is needed to denote that a passive transition occurs with an unknown rate, since it depends on the transitions of other automata. If an automaton does not contain any passive transition, then the underlying model description is a time-homogeneous CTMC. On this basis we justify the following definitions.

Definition 3 (*Open and Closed Automata*)**.** We distinguish the following classes of automata:

1. An LMA $\mathbf{M} = \langle \mathcal{S}, Act, \rightarrow \rangle$ is called *open* if there exists a label $a \in Act$ and a state $s \in \mathcal{S}$ such that a passively enabled in s, i.e., $\exists s' \in \mathcal{S}$ such that $s \xrightarrow{a,x_a} s'$ and $s \neq s'$.
2. An LMA $\mathbf{M} = \langle \mathcal{S}, Act, \rightarrow \rangle$ is called *closed* if it is not open.

We exclude the possibility of two passive transitions with the same label outgoing from the same state. The automata that enjoy this property are called *well-formed*.

Definition 4 (*Well-formed Automata*)**.** LMA $\mathbf{M} = \langle \mathcal{S}, Act, \rightarrow \rangle$ is *well-formed* if:

1. Given a label $a \in Act$ then all the transitions labelled a are either active or passive. Hence, we can say that label a is active or passive for the automaton:

$$\mathcal{A}(\mathbf{M}) \cap \mathcal{P}(\mathbf{M}) = \emptyset.$$

2. If a is a passive label, then for every state s of the automaton there exists exactly one transition labelled a outgoing from s:

$$\forall s \exists s' \in \mathcal{S} \quad \text{such that } s \xrightarrow{a,x_a} s'$$
$$\forall s, s', s'' \in \mathcal{S}, \quad s \xrightarrow{a,x_a} s' \wedge s \xrightarrow{a,x_a} s'' \Longrightarrow s' = s''.$$

We introduce the notion of irreducible LMA.

Definition 5 (*Reachability Set*)**.** Let $\mathbf{M} = \langle \mathcal{S}, Act, \rightarrow \rangle$ be an LMA.

1. A state s' is said to be *reachable in one step* from s if for some $a \in Act$ and $t \in \mathbb{R}^+ \cup \mathbf{Var}$, $s \xrightarrow{a,t} s'$.
2. A state s_n is said to be *reachable* from s_1 if for some $a_1, \ldots, a_n \in Act$ and $t_1, \ldots, t_n \in \mathbb{R}^+ \cup \mathbf{Var}$ and $s_2, \ldots, s_{n-1} \in \mathcal{S}$ we have $s_1 \xrightarrow{a_1,t_1} s_2 \xrightarrow{a_2,t_2} s_3 \ldots s_{n-1} \xrightarrow{a_n,t_n} s_n$.

We write $\mathsf{Reach}(s)$ the set of all reachable states from s.

Definition 6 (*Irreducible LMA*)**.** An LMA $\mathbf{M} = \langle \mathcal{S}, Act, \to \rangle$ is *irreducible* if for all $s \in \mathcal{S}\mathsf{Reach}(s) = \mathcal{S}$.

We define an operation on an automata which we call *closure*. Informally, it consists in the operation of specifying the rates of the passive transitions of an open automaton $\mathbf{M}$.

Definition 7. Assume $\mathbf{M}$ is a well-formed automaton and $a \in \mathcal{P}(\mathbf{M}) = \{a_1, \ldots, a_N\}$. The *closure of the automaton*, $\mathbf{M}^c_{\{K_{a_1},\ldots,K_{a_n}\}} = \langle \mathcal{S}^c, Act^c, \to^c \rangle \quad K_{a_i} \in \mathbb{R}^+$, is defined as:

- $\mathcal{S}^c = \mathcal{S}$,
- $Act^c = Act$,
- $\to^c = \to \setminus \{(s, a_i, x_{a_i}, s'), a_i \in \mathcal{P}(\mathbf{M}),\ s, s' \in \mathcal{S}\} \cup \{(s, a_i, K_{a_i}, s'), a_i \in \mathcal{P}(\mathbf{M}),\ s, s' \in \mathcal{S}\}$.

For simplicity we shall write $\mathbf{M}\{x_{a_1} \leftarrow K_{a_1}, \ldots, x_{a_1} \leftarrow K_{a_n}\}$ for $\mathbf{M}^c_{\{K_{a_1},\ldots,K_{a_n}\}}$, or when the substitution is clear from the context simply $\mathbf{M}^c$.

The closed automata $\mathbf{M}^c$ is identical to the original automaton $\mathbf{M}$ except that each transition labelled $a_i \in \mathcal{P}(\mathbf{M})$ becomes active with rate K_{a_i}. Note that, in $\mathbf{M}^c$ all the transitions sharing the same label and that were passive in $\mathbf{M}$ have the same rate, and that $\mathbf{M}^c$ has a well-defined underlying CTMC and hence the steady-state analysis can be carried out.

We recall GRCAT in a version [9] which is slightly modified with respect to the original one [8].

Theorem 1 (*GRCAT*)**.** *Let $\mathbf{M}_1$ and $\mathbf{M}_2$ be two well-formed LMAs that cooperate on a finite set of labels $L = \{a_1, \ldots, a_N\}$ such that the state-space $S_1 \times S_2$ of $\mathbf{M}_1 \oplus_L \mathbf{M}_2$ is irreducible.*

If there exists the set of rates $\{K_{a_1}, \ldots, K_{a_n}\}$ which satisfies the following equations:

$$\forall s \in \mathcal{S}_1,\ \forall a_i \in \mathcal{A}(\mathbf{M}_1) \quad \frac{\sum_{s' \in \mathcal{S}_1} q(s' \xrightarrow{a_i} s)\pi_1(s')}{\pi_1(s)} = K_{a_i} \tag{A.1}$$

and

$$\forall s \in \mathcal{S}_2,\ \forall a_i \in \mathcal{A}(\mathbf{M}_2) \quad \frac{\sum_{s' \in \mathcal{S}_2} q(s' \xrightarrow{a_i} s)\pi_2(s')}{\pi_2(s)} = K_{a_i} \tag{A.2}$$

where π_1 and π_2 are the invariant measures of the following closed automata $\mathbf{M}^c_1$ and $\mathbf{M}^c_2$

$$\mathbf{M}^c_1 = \mathbf{M}_1\{a_i \leftarrow K_{a_i} : a_i \in \mathcal{P}(\mathbf{M}_1)\}$$
$$\mathbf{M}^c_2 = \mathbf{M}_2\{a_i \leftarrow K_{a_i} : a_i \in \mathcal{P}(\mathbf{M}_2)\}$$

then the following statements hold:

1. *The invariant measure of $\mathbf{M}_1 \oplus_L \mathbf{M}_2$ has the product-form:*
$$\pi(s_1, s_2) = \pi_1(s_1)\pi_2(s_2) \quad \forall s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2. \tag{A.3}$$
2. *If $\sum_{s_i \in \mathcal{S}_\ell} \pi_\ell(s_i) = 1$, with $\ell = 1, 2$, then π is the steady-state probability distribution of $\mathbf{M}_1 \oplus_L \mathbf{M}_2$.*

Note that Theorem 1 implies that every state of an automaton must have at least one incoming transition for each active label.

Corollary 1. *Let $\mathbf{M}_1$ and $\mathbf{M}_2$ be two well-formed LMAs that cooperate on a finite set of labels $L = \{a_1, \ldots, a_n\}$ such that they satisfy the conditions of Theorem 1. If $a_i \in \mathcal{A}(\mathbf{M}_j)$ then for all $s \in \mathcal{S}_j$ there exists a s' such that $s' \xrightarrow{a_i,\mu} s$ with $\mu > 0$ and $a_i \in L$ and $j = 1, 2$.*

References

[1] R.R. Muntz, Poisson departure processes and queueing networks, Technical Report IBM Research Report RC4145, Yorktown Heights, New York, 1972.
[2] F. Baskett, K.M. Chandy, R.R. Muntz, F.G. Palacios, Open, closed, and mixed networks of queues with different classes of customers, J. ACM 22 (2) (1975) 248–260.
[3] F. Kelly, Reversibility and Stochastic Networks, Wiley, New York, 1979.
[4] K.M. Chandy, A.J. Martin, A characterization of product-form queuing networks, J. ACM 30 (2) (1983) 286–299.
[5] N. van Dijk, Queueing Networks and Product Forms, John Wiley, 1993.
[6] I.F. Akyildiz, Product form approximations for queueing networks with multiple servers and blocking, IEEE Trans. Comput. 38 (1989) 99–114.
[7] P. Buchholz, Product form approximations for communicating Markov processes, Perform. Eval. 67 (9) (2010) 797–815 (special issue): QEST 2008.
[8] P.G. Harrison, Turning back time in Markovian process algebra, Theoret. Comput. Sci. 290 (3) (2003) 1947–1986.
[9] A. Marin, M.G. Vigliotti, A general result for deriving product-form solutions of Markovian models, in: Proc. of First Joint WOSP/SIPEW Int. Conf. on Perf. Eng., ACM, San Josè, CA, USA, 2010, pp. 165–176.
[10] P. Buchholz, Adaptive decomposition and approximation for the analysis of stochastic Petri nets, Perform. Eval. 56 (2004) 23–52.
[11] X. Chao, M. Miyazawa, M. Pinedo, Queueing Networks: Customers, Signals and Product form Solutions, Wiley, 1999.
[12] S. Balsamo, V. De Nitto Personè, R. Onvural, Analysis of Queueing Networks with Blocking, Kluwer Academic Publishers, 2001.
[13] R.W. Wolff, Poisson arrivals see time averages, Oper. Res. 30 (2) (1982) 223–231.
[14] J.D.C. Little, A proof of the queuing formula: $L = \lambda W$, Oper. Res. 9 (3) (1961) 383–387.
[15] M.F. Neuts, Matrix Geometric Solutions in Stochastic Models, John Hopkins, Baltimore, MD, 1981.
[16] J. Hillston, A compositional approach to performance modelling, Ph.D. Thesis, Department of Computer Science, University of Edinburgh, 1994.