



UNIVERSITÀ  
DEGLI STUDI  
DI UDINE

## Università degli studi di Udine

### Proportional lumpability and proportional bisimilarity

*Original*

*Availability:*

This version is available <http://hdl.handle.net/11390/1208602> since 2021-07-30T14:58:26Z

*Publisher:*

*Published*

DOI:10.1007/s00236-021-00404-y

*Terms of use:*

The institutional repository of the University of Udine (<http://air.uniud.it>) is provided by ARIC services. The aim is to enable open access to all the world.

*Publisher copyright*

(Article begins on next page)



# Proportional lumpability and proportional bisimilarity

Andrea Marin<sup>1</sup> · Carla Piazza<sup>2</sup> · Sabina Rossi<sup>1</sup>

Received: 5 August 2020 / Accepted: 24 May 2021  
© The Author(s) 2021

## Abstract

In this paper, we deal with the lumpability approach to cope with the state space explosion problem inherent to the computation of the stationary performance indices of large stochastic models. The lumpability method is based on a state aggregation technique and applies to Markov chains exhibiting some structural regularity. Moreover, it allows one to efficiently compute the exact values of the stationary performance indices when the model is actually lumpable. The notion of quasi-lumpability is based on the idea that a Markov chain can be altered by relatively small perturbations of the transition rates in such a way that the new resulting Markov chain is lumpable. In this case, only upper and lower bounds on the performance indices can be derived. Here, we introduce a novel notion of quasi-lumpability, named *proportional lumpability*, which extends the original definition of lumpability but, differently from the general definition of quasi-lumpability, it allows one to derive exact stationary performance indices for the original process. We then introduce the notion of *proportional bisimilarity* for the terms of the performance process algebra PEPA. Proportional bisimilarity induces a proportional lumpability on the underlying continuous-time Markov chains. Finally, we prove some compositionality results and show the applicability of our theory through examples.

**Keywords** CTMCs · Aggregation techniques · Lumpability · Process algebra · Bisimilarity

## 1 Introduction

In the context of performance evaluation of computer systems, continuous-time Markov chains (CTMCs) constitute the underlying semantics model of a plethora of modelling formalism such as Stochastic Petri nets [30], Stochastic Automata Networks (SAN) [31], queuing networks [7] and a class of Markovian process algebras (MPAs), e.g. [20,21]. Usually, one

---

✉ Sabina Rossi  
sabina.rossi@unive.it

Andrea Marin  
marin@unive.it

Carla Piazza  
carla.piazza@uniud.it

<sup>1</sup> Università Ca' Foscari Venezia, Venezia, Italy

<sup>2</sup> Università di Udine, Udine, Italy

is interested in computing the stationary performance indices of the model such as throughput, expected response time, resource utilization and so on. This requires the preliminary computation of the stationary probability distribution of the CTMC underlying the model.

Although the use of high-level modelling formalism highly simplifies the specification of quantitative models by exploiting the compositional properties and the hierarchical approach, the stochastic process underlying even a very compact model may have a number of states that makes its analysis a difficult, even computationally impossible, task. In order to study models with a large state space without using approximations or resorting to simulation, we can attempt to reduce the state space of the underlying Markov chain by aggregating states with equivalent behaviours (according to a notion of equivalence that may vary).

In this paper, we deal with the lumpability approach to cope with the state space explosion problem inherent to the computation of the stationary performance indices of large stochastic models. The lumpability method is based on a state aggregation technique and applies to Markov chains exhibiting some structural regularity. Moreover, it allows one to efficiently compute the exact values of the performance indices when the model is actually lumpable. In the literature, several notions of lumping have been introduced. Interestingly, it has been shown that for Markovian process algebras there is a strong connection between the idea of *bisimulation* and that of *ordinary lumping* (see e.g. [21]). However, it is well known that not all Markov chains are lumpable, if we exclude the trivial aggregation. In fact, only a small percentage of Markov chains arising in real-life applications is expected to be non-trivially lumpable. The notion of *quasi-lumpability* is based on the idea that a Markov chain can be altered by relatively small perturbations of the transition rates in such a way that the new resulting Markov chain is lumpable. In this case, only upper and lower bounds on the performance indices can be derived [18,19]. Here, we face the problem of relaxing the conditions of ordinary lumpability while allowing one to derive the exact stationary performance indices for the original process.

*Contribution* This article is an extended and revised version of the work presented at the International Conference on Formal Modeling and Analysis of Timed Systems (FORMATS 2019) [26]. We introduce a novel notion of quasi-lumpability, named *proportional lumpability*, which extends the original definition of lumpability but, differently from the general definition of quasi-lumpability, it allows one to derive exact stationary performance indices for the original process when the stationary probabilities of the perturbed one are known. Then, we study this notion in the context of a Markovian process algebra. We consider the Performance Evaluation Process Algebra (PEPA) [21] and introduce the concept of *proportional bisimilarity* over PEPA components. Proportional bisimilarity induces a proportional lumpability on the underlying Markov chains. With respect to [26], here we present various characterizations of both proportional lumpability and proportional bisimilarity that provide more insights on their definitions. Moreover, we prove some compositionality results and show the applicability of our theory through examples.

*Related work* At the stochastic process level of abstraction, several approaches, both exact and approximated, have been proposed to cope with the state space explosion problem.

We begin our survey by first considering exact aggregation methods and then we consider approximate approaches.

In [22, Ch. 6], the authors introduce the notion of *ordinary lumping* of states in a discrete-time Markov chain (DTMC), but the concept can be straightforwardly extended to CTMCs. In ordinary lumping, the states of the Markov chain are clustered according to some structural properties of the transition rate matrix so that a CTMC with a smaller number of states can be defined. Since the computational efforts of the analysis of this latter chain are lower than

those required by the original one, lumping can be an effective way for studying large Markov chains both for transient and stationary properties.

The results on lumping can be extended to cope with higher-level Markovian formalisms including Stochastic Petri Nets (SPNs) and Markovian process algebras. For example, a structural-based approach to lumping for SPNs is studied in [3,5], where structural symmetries of the net are exploited to derive a lumped underlying CTMC in an efficient way. These works share with our results the consequence of reducing the computational complexity of the stationary performance indices of the model. However, in our case, we propose the idea of lumping a perturbed Markov chain from which we can derive the stationary properties of the original chain.

In the context of Markovian process algebras, structural process properties are studied in [6,9–11,21,27] for state space reduction purposes by means of equivalence relations inspired by bisimulation. In [12], the spectrum of bisimilarities for DTMCs and CTMCs without actions is presented. In [9–11,21], action-labeled CTMCs are considered, i.e. CTMCs where transitions are labeled with both actions and rates. In particular, in [21] the notion of strong bisimilarity is considered, while in [9–11] a weak variant is studied aiming at reducing sequences of internal  $\tau$ -actions to a single internal  $\tau$ -action that preserves the average duration of the sequence. If two components are equivalent, it is possible to replace one of them (that with more states) with the other without affecting the behaviour of the remaining parts of the system. The notion of *strong equivalence* introduced in [21] for processes expressed as terms of the Performance Evaluation Process Algebra (PEPA) always induces a lumping of the CTMC underlying a PEPA process, although in general the opposite is not true. With respect to these works, our results on PEPA models generalize the notion of strong equivalence and allow us to deal with underlying Markov chains that are not directly lumpable. On the other hand, the conditions on the cooperation between components that preserve proportional lumpability are stricter than those required by strong equivalence.

We now analyze approximate aggregation methods. In [18], the notion of quasi-lumpable Markov chain is introduced. The idea is that a quasi-lumpable Markov chain is one that can be made lumpable by a relatively small perturbation of the transition rates. In [18, 19], a technique for the computation of bounds based on the Courtois and Semal's method is presented. Another approach based on approximated aggregation of states is presented in [1]. Although the authors consider transient analysis of DTMCs, their results can be extended to the stationary analysis of CTMCs. Interestingly, they provide explicit bounds on the errors introduced when computing the state probabilities. With respect to our work, we can retrieve the exact stationary distribution, but, differently from [1], we do not consider transient analysis.

The notion of quasi-lumpability in the context of a Markovian process algebra has been studied in [29], where the authors introduce the concept of approximate strong equivalence for PEPA components and propose a partitioning strategy that involves the use of a clustering algorithm that minimizes an upper bound for approximate strong equivalence. Differently from approximate strong equivalence, which is not preserved under union, we define proportional bisimilarity as the union of all proportional bisimulations and provide an efficient algorithm, based on classical lumpability ones, to compute it. We then exploit the results on proportional lumpability to derive the exact stationary distribution of the underlying Markov chain.

*Structure of the paper* The paper is structured as follows: In Sect. 2, we review the theoretical background on continuous-time Markov chains and recall the concept of lumpability. The notions of quasi-lumpability and proportional lumpability are introduced and illustrated

through an example. In Sect. 3, we recall the Performance Evaluation Process Algebra (PEPA) [21] that is an algebraic calculus enhanced with stochastic timing information that may be used to calculate performance measures as well as prove functional system properties. The notions of quasi-lumpable bisimulation and proportional bisimilarity are defined following the lines of [26]. Moreover, some compositionality results for proportional bisimilarity are proved and two examples are discussed. In Sect. 4, we compare the notion of proportional lumpability with others forms of lumpability. Section 5 concludes the paper.

## 2 CTMCs and proportional lumpability

In this section, we review the theoretical background on continuous-time Markov chains and the concept of lumpability. Then, we introduce and characterize proportional lumpability.

### 2.1 Continuous-time Markov chains

A continuous-time Markov chain (CTMC) is a stochastic process  $X(t)$  for  $t \in \mathbb{R}^+$  taking values into a discrete state space  $\mathcal{S}$  such that the *Markov property* holds, i.e. the conditional (on both past and present states) probability distribution of its future behaviour is independent of its past evolution until the present state:

$$\begin{aligned} \text{Prob}(X(t_{n+1}) = s_{n+1} \mid X(t_1) = s_1, X(t_2) = s_2, \dots, X(t_n) = s_n) = \\ \text{Prob}(X(t_{n+1}) = s_{n+1} \mid X(t_n) = s_n). \end{aligned}$$

$X(t)$  is *stationary* if the collection of random variables  $(X(t_1), X(t_2), \dots, X(t_n))$  is distributed as  $(X(t_1 + \tau), X(t_2 + \tau), \dots, X(t_n + \tau))$  for all  $t_1, t_2, \dots, t_n, \tau \in \mathbb{R}^+$ .

A CTMC  $X(t)$  is said to be *time-homogeneous* if the conditional probability  $\text{Prob}(X(t + \tau) = s \mid X(t) = s')$  does not depend upon  $t$ , and is *irreducible* if every state in  $\mathcal{S}$  can be reached from every other state. A state in a Markov chain is called *recurrent* if the probability that the process will eventually return to the same state is one. A recurrent state is called *positive-recurrent* if the expected number of steps until the process returns to it is finite. A CTMC is *ergodic* if it is irreducible and all its states are positive-recurrent. In the case of finite Markov chains, irreducibility is sufficient for ergodicity. Henceforth, we assume the ergodicity of the CTMCs that we study.

An ergodic CTMC possesses an *equilibrium* (or *steady-state*) *distribution*, which is the *unique* collection of positive real numbers  $\pi(s)$  with  $s \in \mathcal{S}$  such that

$$\lim_{t \rightarrow \infty} \text{Prob}(X(t) = s \mid X(0) = s') = \pi(s).$$

Notice that the above equation for  $\pi(s)$  is independent of  $s'$ . We denote by  $q(s, s')$  the transition rate out of state  $s$  to state  $s'$ , with  $s \neq s'$ , and by  $q(s)$  the sum of all transition rates out of state  $s$  to any other state in the chain. A state  $s$  for which  $q(s) = \infty$  is called an instantaneous state since when entered it is instantaneously left. Whereas such states are theoretically possible, we shall assume throughout that  $0 < q(s) < \infty$  for all state  $s$ . The infinitesimal generator matrix  $\mathbf{Q}$  of a CTMC  $X(t)$  with state space  $\mathcal{S}$  is the  $|\mathcal{S}| \times |\mathcal{S}|$  matrix whose off-diagonal elements are the  $q(s, s')$ 's and whose diagonal elements are the negative sum of the extra diagonal elements of each row, i.e.  $q(s, s) = -\sum_{s' \in \mathcal{S}, s' \neq s} q(s, s')$ . For the sake of simplicity, we use  $q(s, s')$  to denote the components of matrix  $\mathbf{Q}$ .

Any non-trivial vector of positive real numbers  $\mu$  satisfying the system of global balance equations (GBEs)

$$\mu Q = 0$$

is called *invariant measure* of the CTMC. For an irreducible CTMC  $X(t)$ , if  $\mu_1$  and  $\mu_2$  are two invariant measures of  $X(t)$ , then there exists a constant  $k > 0$  such that  $\mu_1 = k\mu_2$ . If the CTMC is ergodic, then there exists a unique invariant measure  $\pi$  whose components sum to unity, i.e.  $\sum_{s \in S} \pi(s) = 1$ . In this case  $\pi$  is the *equilibrium* or *steady-state distribution* of the CTMC.

## 2.2 Lumpability

In the context of performance and reliability analysis, the notion of *lumpability* provides a model simplification technique that can be used for generating an aggregated Markov chain that is smaller than the original one but allows one to determine exact results for the original process.

The concept of lumpability can be formalized in terms of equivalence relations over the state space of the Markov chain. Any such equivalence induces a *partition* on the state space of the Markov chain and aggregation is achieved by clustering equivalent states into macro-states, thus reducing the overall state space. If the partition can be shown to satisfy the so-called *ordinary lumpability condition* [4,22], then the equilibrium solution of the aggregated process may be used to derive an exact solution of the original one.

Ordinary lumpability has been introduced in [22] and further studied in [2,13,28,33].

**Definition 1** (Ordinary lumpability) Let  $X(t)$  be a CTMC with state space  $S$  and  $\sim$  be an equivalence relation over  $S$ . We say that  $X(t)$  is *ordinarily lumpable* with respect to  $\sim$  (resp.,  $\sim$  is a *ordinary lumpability* for  $X(t)$ ) if  $\sim$  induces a partition on the state space of  $X(t)$  such that for any equivalence classes  $S_i, S_j \in S/\sim$  with  $S_i \neq S_j$  and  $s, s' \in S_i$ ,

$$\sum_{s'' \in S_j} q(s, s'') = \sum_{s'' \in S_j} q(s', s'').$$

Thus, an equivalence relation over the state space of a Markov chain is an ordinary lumpability if it induces a partition into equivalence classes such that for any two states within an equivalence class their aggregated transition rates to any other class are the same. Notice that every Markov chain is ordinarily lumpable with respect to the identity relation, and also with respect to the trivial relation having only one equivalence class.

In [22], the authors prove that for an equivalence relation  $\sim$  over the state space of a Markov chain  $X(t)$ , the aggregated process is a Markov chain for every initial distribution if, and only if,  $\sim$  is a ordinary lumpability for  $X(t)$ . Moreover, the transition rate between two aggregated states  $S_i, S_j \in S/\sim$  is equal to  $\sum_{s' \in S_j} q(s, s')$  for any  $s \in S_i$ .

**Proposition 1** (Aggregated process for ordinary lumpability) Let  $X(t)$  be a CTMC with state space  $S$ , infinitesimal generator  $Q$  and equilibrium distribution  $\pi$ . Let  $\sim$  be a ordinary lumpability for  $X(t)$  and  $\tilde{X}(t)$  be the aggregated process with state space  $S/\sim$  and infinitesimal generator  $\tilde{Q}$  defined by: for any equivalence classes  $S_i, S_j \in S/\sim$  with  $S_i \neq S_j$

$$\tilde{q}(S_i, S_j) = \sum_{s' \in S_j} q(s, s')$$

for any  $s \in S_i$ . Then the equilibrium distribution  $\tilde{\pi}$  of  $\tilde{X}(t)$  is such that for any equivalence class  $S \in \mathcal{S}/\sim$ ,

$$\tilde{\pi}(S) = \sum_{s \in S} \pi(s).$$

In general, a non-trivial lumpable partition might not exist. In [18] the notion of *quasi-lumpability* has been introduced to characterize those Markov chains that can be made lumpable by a relatively small perturbation to the transition rates.

**Definition 2** (Quasi-lumpability) Let  $X(t)$  be a CTMC with state space  $\mathcal{S}$  and  $\sim$  be an equivalence relation over  $\mathcal{S}$ . We say that  $X(t)$  is *quasi-lumpable* with respect to  $\sim$  and a bound  $\epsilon \geq 0$  (resp.,  $\sim$  is a *quasi-lumpability* for  $X(t)$  with respect to  $\epsilon$ ) if  $\sim$  induces a partition on the state space of  $X(t)$  such that for any equivalence classes  $S_i, S_j \in \mathcal{S}/\sim$  with  $S_i \neq S_j$  and  $s, s' \in S_i$ ,

$$\left| \sum_{s'' \in S_j} q(s, s'') - \sum_{s'' \in S_j} q(s', s'') \right| \leq \epsilon.$$

The notion of quasi-lumpability coincides with the concept of *near-lumpability* presented in [13]. Techniques for computing bounds to the steady state probabilities of quasi-lumpable Markov chains have been studied in [16,18,19,32].

### 2.3 Proportional lumpability

In [26], we introduced a novel notion of lumpability, named *proportional lumpability* that as the notion of quasi-lumpability, extends the original definition of ordinary lumpability but, differently from the general definition of quasi-lumpability, it allows us to derive an exact solution of the original process.

**Definition 3** (Proportional lumpability) Let  $X(t)$  be a CTMC with state space  $\mathcal{S}$  and  $\sim$  be an equivalence relation over  $\mathcal{S}$ . We say that  $X(t)$  is *proportionally lumpable* with respect to  $\sim$  (resp.,  $\sim$  is a *proportional lumpability* for  $X(t)$ ) if there exists a function  $\kappa$  from  $\mathcal{S}$  to  $\mathbb{R}^+$  such that  $\sim$  induces a partition on the state space of  $X(t)$  satisfying the property that for any equivalence classes  $S_i, S_j \in \mathcal{S}/\sim$  with  $S_i \neq S_j$  and  $s, s' \in S_i$ ,

$$\frac{\sum_{s'' \in S_j} q(s, s'')}{\kappa(s)} = \frac{\sum_{s'' \in S_j} q(s', s'')}{\kappa(s')}.$$

We say that  $X(t)$  is  $\kappa$ -*proportionally lumpable* with respect to  $\sim$  (resp.,  $\sim$  is a  $\kappa$ -*proportional lumpability* for  $X(t)$ ) if  $X(t)$  is *proportionally lumpable* with respect to  $\sim$  and function  $\kappa$ .

The following proposition proves that proportional lumpability allows one to compute an exact solution for the original model when the stationary probabilities of the perturbed one are known.

**Proposition 2** (Aggregated process for proportional lumpability) *Let  $X(t)$  be a CTMC with state space  $\mathcal{S}$ , infinitesimal generator  $\mathbf{Q}$  and equilibrium distribution  $\pi$ . Let  $\kappa$  be a function from  $\mathcal{S}$  to  $\mathbb{R}^+$ ,  $\sim$  be a  $\kappa$ -proportional lumpability for  $X(t)$  and  $\tilde{X}(t)$  be the aggregated process*

with state space  $\mathcal{S}/\sim$  and infinitesimal generator  $\tilde{\mathbf{Q}}$  defined by: for any equivalence classes  $S_i, S_j \in \mathcal{S}/\sim$  with  $S_i \neq S_j$ ,

$$\tilde{q}(S_i, S_j) = \frac{\sum_{s' \in S_j} q(s, s')}{\kappa(s)}$$

for any  $s \in S_i$ . Then the function  $\tilde{\mu}$  defined by letting for any equivalence class  $S \in \mathcal{S}/\sim$ ,

$$\tilde{\mu}(S) = \sum_{s \in S} \pi(s)\kappa(s), \tag{1}$$

is an invariant measure of  $\tilde{X}(t)$ .

**Proof** Let  $\tilde{X}(t)$  be the aggregated process defined as above. In order for  $\tilde{\mu}$  to be an invariant measure, it must satisfy the global balance equations for all  $S \in \mathcal{S}/\sim$ , which is

$$\tilde{\mu}(S) \sum_{\substack{S' \in \mathcal{S}/\sim \\ S' \neq S}} \tilde{q}(S, S') = \sum_{\substack{S' \in \mathcal{S}/\sim \\ S' \neq S}} \tilde{\mu}(S')\tilde{q}(S', S). \tag{2}$$

The proof follows by substituting the definitions of  $\tilde{q}$  and  $\tilde{\mu}$  given above. Indeed, the left-hand side of Eq. (2) can be written as follows, where  $s''$  is an arbitrary state in  $S$ :

$$\left( \sum_{s \in S} \pi(s)\kappa(s) \right) \left( \sum_{\substack{S' \in \mathcal{S}/\sim \\ S' \neq S}} \frac{\sum_{s' \in S'} q(s'', s')}{\kappa(s'')} \right) = \left( \sum_{s \in S} \sum_{\substack{S' \in \mathcal{S}/\sim \\ S' \neq S}} \pi(s)\kappa(s) \frac{\sum_{s' \in S'} q(s'', s')}{\kappa(s'')} \right),$$

and since  $\frac{q(s'', s')}{\kappa(s'')}$  is the same, by proportional lumpability, for all  $s'' \in S$ , we can rewrite this expression as

$$\left( \sum_{s \in S} \sum_{\substack{S' \in \mathcal{S}/\sim \\ S' \neq S}} \pi(s)\kappa(s) \frac{\sum_{s' \in S'} q(s, s')}{\kappa(s)} \right) = \sum_{s \in S} \pi(s) \sum_{\substack{S' \in \mathcal{S}/\sim \\ S' \neq S}} \sum_{s' \in S'} q(s, s').$$

The right-hand side of Eq. (2) can be written as: for all  $s'' \in S'$ ,

$$\sum_{\substack{S' \in \mathcal{S}/\sim \\ S' \neq S}} \left( \sum_{s' \in S'} \pi(s')\kappa(s') \right) \frac{\sum_{s \in S} q(s'', s)}{\kappa(s'')},$$

and since  $\frac{q(s'', s)}{\kappa(s'')}$  is the same, by proportional lumpability, for all  $s'' \in S'$ , we can rewrite this expression as

$$\begin{aligned} \sum_{\substack{S' \in \mathcal{S}/\sim \\ S' \neq S}} \left( \sum_{s' \in S'} \pi(s')\kappa(s') \frac{\sum_{s \in S} q(s', s)}{\kappa(s')} \right) &= \sum_{\substack{S' \in \mathcal{S}/\sim \\ S' \neq S}} \sum_{s' \in S'} \pi(s') \sum_{s \in S} q(s', s) \\ &= \sum_{s \in S} \sum_{\substack{S' \in \mathcal{S}/\sim \\ S' \neq S}} \sum_{s' \in S'} \pi(s')q(s', s). \end{aligned}$$



Therefore, Eq. (2) can be rewritten as:

$$\sum_{s \in \mathcal{S}} \pi(s) \sum_{s' \in \mathcal{S}, s' \neq s} q(s, s') = \sum_{s \in \mathcal{S}} \sum_{s' \in \mathcal{S}, s' \neq s} \pi(s')q(s', s)$$

which is true by the general conservation law that states that, for any closed boundary, the effective flow inward must equal the effective flow outward.  $\square$

The following Definition 4 introduces a way to perturb a proportionally lumpable CTMC in order to obtain a ordinarily lumpable one. In contrast to previous perturbation-based approaches, Proposition 3 gives a way to compute the stationary probabilities of the original chain given those of the perturbed lumpable one. In the cases in which lumpability allows for an efficient computation of the detailed state probabilities, Proposition 3 gives an efficient way to compute the detailed probabilities also of the original process.

We show how to compute the equilibrium distribution of a proportionally lumpable CTMC  $X(t)$  from the equilibrium distribution of a perturbed chain  $X'(t)$  defined as follows.

**Definition 4** (Perturbed Markov chains) Let  $X(t)$  be a CTMC with state space  $\mathcal{S}$  and infinitesimal generator  $\mathbf{Q}$ . Let  $\kappa$  be a function from  $\mathcal{S}$  to  $\mathbb{R}^+$ . We say that a CTMC  $X'(t)$  with infinitesimal generator  $\mathbf{Q}'$  is a perturbation of  $X(t)$  with respect to  $\kappa$  if  $X'(t)$  is obtained from  $X(t)$  by perturbing its rates such that for all  $s, s' \in \mathcal{S}$  with  $s \neq s'$ ,

$$q'(s, s') = \frac{q(s, s')}{\kappa(s)}.$$

**Proposition 3** (Equilibrium distribution for proportional lumpability) *Let  $X(t)$  be a CTMC with state space  $\mathcal{S}$ , infinitesimal generator  $\mathbf{Q}$  and equilibrium distribution  $\pi$ . Let  $\kappa$  be a function from  $\mathcal{S}$  to  $\mathbb{R}^+$ . Then, for any perturbation  $X'(t)$  of the original chain  $X(t)$  with respect to  $\kappa$  according to Definition 4 with infinitesimal generator  $\mathbf{Q}'$  and equilibrium distribution  $\pi'$ , the equilibrium distribution  $\pi$  of  $X(t)$  satisfies the following property: let  $K = \sum_{s \in \mathcal{S}} \pi'(s)/\kappa(s)$  then, for all  $s \in \mathcal{S}$*

$$\pi(s) = \frac{\pi'(s)}{K \kappa(s)}.$$

**Proof** Recall that, for ergodic CTMCs, the equilibrium distribution is the only distribution that satisfies the system of global balance equations. The proof is carried out as follows: we exploit the fact that  $\pi'$  is the only distribution that satisfies the global balance equations for  $X'(t)$  and show that this implies that  $\pi$  satisfies the global balance equations for  $X(t)$ .

For all  $s \in \mathcal{S}$ , the corresponding global balance equation is

$$\pi(s) \sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} q(s, s') = \sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} \pi(s')q(s', s). \tag{3}$$

We now replace the guess of  $\pi(s)$  given above. Indeed, the left-hand side of Eq. (3) can be written as follows:

$$\frac{\pi'(s)}{K \kappa(s)} \left( \sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} q(s, s') \right) = \frac{\pi'(s)}{K} \left( \sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} \frac{q(s, s')}{\kappa(s)} \right) = \frac{\pi'(s)}{K} \sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} q'(s, s'),$$

where the last equality follows from Definition 4.

The right-hand side of Eq. (3) can be written as follows:

$$\sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} \frac{\pi'(s')}{K \kappa(s')} q(s', s) = \frac{1}{K} \sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} \pi'(s') \frac{q(s', s)}{\kappa(s')} = \frac{1}{K} \sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} \pi'(s') q'(s', s).$$

Hence, global balance equations (3) of  $X(t)$  are satisfied by  $\pi$  since they are equivalent to the global balance equations of  $X'(t)$ , i.e. for all  $s \in \mathcal{S}$

$$\pi'(s) \sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} q'(s, s') = \sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} \pi'(s') q'(s', s).$$

We now prove that the normalizing condition also holds: i.e.  $\sum_{s \in \mathcal{S}} \pi(s) = 1$ . The proof follows trivially from the fact that  $K = \sum_{s \in \mathcal{S}} \pi'(s)/\kappa(s)$ , in fact:

$$\sum_{s \in \mathcal{S}} \pi(s) = \sum_{s \in \mathcal{S}} \frac{\pi'(s)}{K \kappa(s)} = \frac{1}{K} \sum_{s \in \mathcal{S}} \frac{\pi'(s)}{\kappa(s)} = \frac{1}{K} K = 1.$$

□

Understanding which class of models fulfills the conditions of proportional lumpability may be not straightforward. One possible approach is similar to that introduced in the context of the analysis of product-form models. These are models that, similarly to lumpable ones, allow for an efficient numerical or analytical study of the stationary properties of a CTMC. Intuitively, in order to relax the conditions for the product-form, the authors propose several methods to modify the CTMC transition rates of a product-form model in such a way that its solution can be used to efficiently derive the stationary distribution (sometimes called semi-product-form) of the modified model (see, e.g. [8,15,34]). This has led to important applications such as [36].

In our case, we can follow the same lines and use Definition 3 to understand the modifications that we can apply to an ordinary lumpable model to obtain a proportionally lumpable one as shown in Example 1.

**Example 1** Consider a set of  $N$  independent identical CTMCs  $X'_i(t)$ ,  $1 \leq i \leq N$  on the state space  $\mathcal{S}_0$  that represents the computational phases of a single CTMC, then the process  $X'(t) = (X'_1(t), \dots, X'_N(t))$  is still a CTMC and is known to be ordinary lumpable. Let  $s = (s_1, \dots, s_N) \in \mathcal{S}_0^N$  be a state of the joint process and denote  $|s|_x$  the number of components equal to  $x \in \mathcal{S}_0$ . Then, the ordinary lumping relation  $\sim$  is defined as follows:

$$\forall s_1, s_2 \in \mathcal{S}_0^N, \quad s_1 \sim s_2 \iff |s_1|_x = |s_2|_x \quad \forall x \in \mathcal{S}_0.$$

Now, suppose that state  $y \in \mathcal{S}_0$  denotes a critical situation for the process (e.g. some emergency task cannot be taken when the process is in that phase) and hence we aim at reducing the probability of the situation in which all the processes are simultaneously in state  $y$ . We will observe the state of  $M < N$  processes, and if all these are in state  $y$ , we can increase the speed of all  $M$  servers by a factor  $k > 1$ , e.g. by frequency scaling. As a consequence, our system consumes more energy but will reduce the residence time in a state in which at least  $M$  processes are in the critical state. Without loss of generality, we define the state  $s$  in such a way that the  $M$  observed servers are the first. Let  $X(t)$  be the CTMC obtained by this modification.

We can immediately see that  $\sim$  is not a lumping for  $X(t)$  because states  $(y, \dots, y, s_{M+1}, \dots, s_N)$ ,  $s_j \in \mathcal{S}_0$ ,  $j = M + 1, \dots, N$ , do not have the same behaviour as the states

in which we count the same number of  $y$  but not all in the first  $M$  positions. However, the conditions of Definition 3 are trivially satisfied because the speed scaling factor is the same for all the  $N$  processes and hence we can solve the model and apply Proposition 3 to compute the stationary state probabilities of  $X(t)$  given those of  $X'(t)$ .

With Proposition 4, we provide a novel useful characterization of proportional lumpability. Hereafter, for a given equivalence relation  $\sim$  over the state space of a CTMC, we denote by  $q_{\sim}(s)$  the sum of all transition rates from the state  $s$  to any state  $t$  such that  $s \not\sim t$ , i.e. for all  $s \in \mathcal{S}$ ,

$$q_{\sim}(s) = \sum_{t \not\sim s} q(s, t).$$

**Proposition 4** (Characterization of proportionally lumpable CTMCs) *Let  $X(t)$  be an ergodic CTMC with state space  $\mathcal{S}$  and  $\sim$  be an equivalence relation over  $\mathcal{S}$ . The relation  $\sim$  is a proportional lumpability for  $X(t)$  if and only if for any equivalence classes  $S_i, S_j \in \mathcal{S}/\sim$  with  $S_i \neq S_j$  and  $s, s' \in S_i$ ,*

- either  $q_{\sim}(s) = q_{\sim}(s') = 0$
- or  $q_{\sim}(s), q_{\sim}(s') > 0$  and

$$\frac{\sum_{s'' \in S_j} q(s, s'')}{q_{\sim}(s)} = \frac{\sum_{s'' \in S_j} q(s', s'')}{q_{\sim}(s')}.$$

**Proof**  $\Rightarrow$ ) Suppose that  $\sim$  is a  $\kappa$ -proportional lumpability for a function  $\kappa : \mathcal{S} \rightarrow \mathbb{R}^+$ , i.e. for any equivalence classes  $S_i, S_j \in \mathcal{S}/\sim$  with  $S_i \neq S_j$  and  $s, s' \in S_i$ ,

$$\frac{\sum_{s'' \in S_j} q(s, s'')}{\kappa(s)} = \frac{\sum_{s'' \in S_j} q(s', s'')}{\kappa(s')}. \tag{4}$$

By summing the left and right terms of the above equation over all  $S_j \in \mathcal{S}/\sim$  with  $S_i \neq S_j$  we get

$$\sum_{\substack{S_j \in \mathcal{S}/\sim \\ i \neq j}} \frac{\sum_{s'' \in S_j} q(s, s'')}{\kappa(s)} = \sum_{\substack{S_j \in \mathcal{S}/\sim \\ i \neq j}} \frac{\sum_{s'' \in S_j} q(s', s'')}{\kappa(s')},$$

that can be rewritten as

$$\frac{\sum_{\substack{S_j \in \mathcal{S}/\sim \\ i \neq j}} \left( \sum_{s'' \in S_j} q(s, s'') \right)}{\kappa(s)} = \frac{\sum_{\substack{S_j \in \mathcal{S}/\sim \\ i \neq j}} \left( \sum_{s'' \in S_j} q(s', s'') \right)}{\kappa(s')},$$

that is:

$$\frac{q_{\sim}(s)}{\kappa(s)} = \frac{q_{\sim}(s')}{\kappa(s')}.$$

Notice that by definition of proportional lumpability  $q_{\sim}(s) = 0$  if and only if  $q_{\sim}(s') = 0$ . If  $q_{\sim}(s) = q_{\sim}(s') = 0$  then the property is trivially satisfied. Suppose that  $q_{\sim}(s), q_{\sim}(s') > 0$ , then we get

$$\frac{\kappa(s)}{q_{\sim}(s)} = \frac{\kappa(s')}{q_{\sim}(s')}. \tag{5}$$

Now by multiplying the left and right terms of Eqs. (4) and (5) we have that for any equivalence classes  $S_i, S_j \in \mathcal{S}/\sim$  with  $S_i \neq S_j$  and  $s, s' \in S_i$ ,

$$\frac{\kappa(s)}{q_{\sim}(s)} \frac{\sum_{s'' \in S_j} q(s, s'')}{\kappa(s)} = \frac{\kappa(s')}{q_{\sim}(s')} \frac{\sum_{s'' \in S_j} q(s', s'')}{\kappa(s')},$$

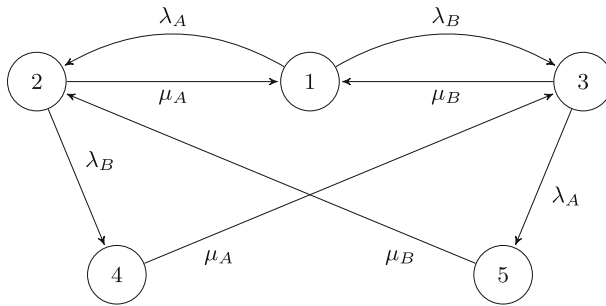


Fig. 1 Two processor system

and, by simplifying, we get the thesis.

⇐) Suppose that  $\sim$  is an equivalence relation such that for any equivalence classes  $S_i, S_j \in \mathcal{S}/\sim$  with  $S_i \neq S_j$  and  $s, s' \in S_i$ ,

- either  $q_{\sim}(s) = q_{\sim}(s') = 0$
- or  $q_{\sim}(s), q_{\sim}(s') > 0$  and

$$\frac{\sum_{s'' \in S_j} q(s, s'')}{q_{\sim}(s)} = \frac{\sum_{s'' \in S_j} q(s', s'')}{q_{\sim}(s')}.$$

Then, by Definition 3,  $\sim$  is a proportional lumpability with respect to the function  $\kappa : \mathcal{S} \rightarrow \mathbb{R}^+$  such that for all state  $s \in \mathcal{S}$ ,  $\kappa(s) = q_{\sim}(s)$  if  $q_{\sim}(s) > 0$  otherwise  $\kappa(s)$  is an arbitrary positive real number. □

**Example 2** Consider a system with multiple CPUs and a two-level memory: the first level is private for each processor (e.g. a cache) and the second level is a shared memory that can be accessed in a mutually exclusive way. The CPUs alternate a phase in which they use their private memory and a phase in which they request the access to the shared memory and, once this is granted, they use it. At the end of this phase, the shared memory lock is released. The duration of the first phase is exponentially distributed with mean  $\lambda_P^{-1}$  for processor  $P$ . The common memory access duration is also assumed to be exponentially distributed with mean  $\mu_P^{-1}$  for processor  $P$ .

For the sake of simplicity, let us analyse a two-processor version of the systems, where processors are denoted by  $A$  and  $B$ . Assume that the processors have different timing characteristics: the private and common memory accesses of  $A$  are governed by two exponential distributions with parameters  $\lambda_A$  and  $\mu_A$ , respectively, while the private and common memory accesses of  $B$  are governed by two exponential distributions with parameters  $\lambda_B$  and  $\mu_B$ , respectively. The CTMC describing the behaviour of this two-processor system is depicted in Fig. 1 and has five states that can be interpreted as follows:

- State 1:  $A$  and  $B$  both executing in their private memories;
- State 2:  $B$  executing in private memory, and  $A$  accessing common memory;
- State 3:  $A$  executing in private memory, and  $B$  accessing common memory;
- State 4:  $A$  accessing common memory,  $B$  waiting for common memory;
- State 5:  $B$  accessing common memory,  $A$  waiting for common memory.

In general, for arbitrary rates  $\lambda_A, \mu_A, \lambda_B$  and  $\mu_B$ , the CTMC is not lumpable. However, suppose that the rates are related as follows:

$$\lambda_A = k_1 \lambda \quad \mu_A = k_1^{-1} \mu \quad \lambda_B = k_2 \lambda \quad \mu_B = k_2^{-1} \mu$$

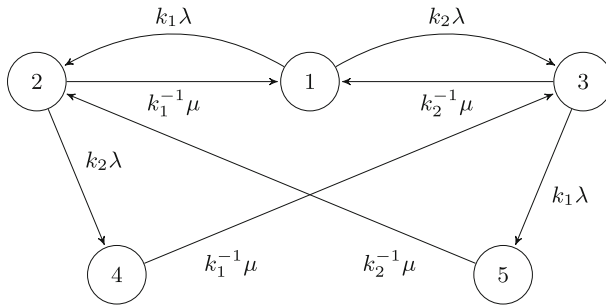


Fig. 2 Two processor system with proportional factors

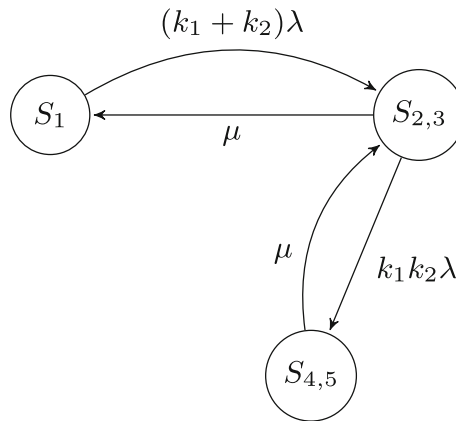


Fig. 3 Two processor reduced system

for  $\lambda, \mu, k_1, k_2 \in \mathbb{R}^+$ . In this case, the CTMC appears as represented in Fig. 2. We can observe that it is proportionally lumpable with respect to the equivalence classes  $S_1 = \{1\}$ ,  $S_{2,3} = \{2, 3\}$  and  $S_{4,5} = \{4, 5\}$  and the function  $\kappa$  defined by:  $\kappa(1) = 1$ ,  $\kappa(2) = k_1^{-1}$ ,  $\kappa(3) = k_2^{-1}$ ,  $\kappa(4) = k_1^{-1}$ ,  $\kappa(5) = k_2^{-1}$ . Thus, we can analyse the reduced chain represented in Fig. 3 and, by Propositions 2 and 3, we can compute the exact solution of the original model.

**Example 3** We consider a simple buffer in which messages are added according to a Poisson process with rate  $\lambda$  and that is cleared at exponentially spaced instants. The mean time between successive clearances is  $n\mu^{-1}$ , where  $n$  denotes the number of items in the buffer. The buffer has capacity  $M$  and, when full, arrivals are lost.

The behaviour of the buffer is described by the CTMC depicted in Fig. 4.

If we consider the function  $\kappa$  such that  $\kappa(B_0) = 1$  and  $\kappa(B_n) = 1/n$  for all  $n$  with  $0 < n \leq M$ , then it is easy to prove that  $B_0$  and  $B'_0$  are proportionally lumpable as well as  $B'_i$  and  $B'_1$  for  $1 \leq i \leq M$ , where  $B'_0$  and  $B'_1$  are depicted in Fig. 5.

From the equilibrium distribution of the reduced system, we can then compute the equilibrium distribution of the original system by applying Proposition 3.

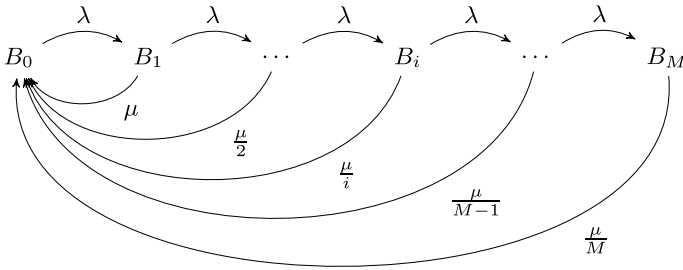


Fig. 4 Original buffer system

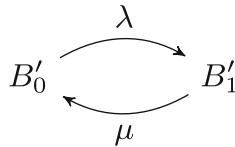


Fig. 5 The buffer reduced system

### 3 PEPA and proportional bisimilarity

In this section, we recall the Performance Evaluation Process Algebra (PEPA) [21] and introduce the notion of *proportional bisimilarity*. We then study the compositionality properties of proportional bisimilarity.

#### 3.1 The process algebra PEPA

PEPA is an algebraic calculus enhanced with stochastic timing information that may be used to calculate performance measures as well as prove functional system properties.

The basic elements of PEPA are *components* and *activities*. Each activity is represented by a pair  $(\alpha, r)$  where  $\alpha$  is a label, or *action type*, and  $r$  is its *activity rate*, that is the parameter of an exponential distribution determining its duration. We assume that there is a countable set,  $\mathcal{A}$ , of possible action types, including a distinguished type,  $\tau$ , which can be regarded as the *unknown* type. Activities of this type will be private to the component in which they occur. Activity rates may be any positive real number, or the distinguished symbol  $\top$  that should be read as *unspecified*.

The PEPA language provides a small set of combinators. These allow language terms to be constructed defining the behaviour of components, via the activities they undertake and the interactions between them. The syntax for PEPA terms is defined by the grammar:

$$\begin{aligned}
 S &::= (\alpha, r).S \mid S + S \mid A \\
 P &::= P \boxtimes_L P \mid P/L \mid S
 \end{aligned}$$

where  $L \subseteq \mathcal{A} \setminus \{\tau\}$ ,  $S$  denotes a *sequential component*, while  $P$  denotes a *model component* that executes in parallel. We assume that there is a countable set of *constants*,  $A$ . We write  $\mathcal{C}$  for the set of all possible components.

**Structural Operational Semantics** The structural operational semantics of PEPA is given in Table 1. Component  $(\alpha, r).P$  carries out the activity  $(\alpha, r)$  of type  $\alpha$  at rate  $r$  and subsequently behaves as  $P$ . When  $a = (\alpha, r)$ , component  $(\alpha, r).P$  may be written as  $a.P$ . Component

**Table 1** Operational semantics for PEPA components

$\frac{}{(\alpha, r).P \xrightarrow{(\alpha, r)} P}$	$\frac{P \xrightarrow{(\alpha, r)} P'}{A \xrightarrow{(\alpha, r)} P'} \quad (A \stackrel{def}{=} P)$
$\frac{P \xrightarrow{(\alpha, r)} P'}{P + Q \xrightarrow{(\alpha, r)} P'}$	$\frac{Q \xrightarrow{(\alpha, r)} Q'}{P + Q \xrightarrow{(\alpha, r)} Q'}$
$\frac{P \xrightarrow{(\alpha, r)} P'}{P/L \xrightarrow{(\alpha, r)} P'/L} \quad (\alpha \notin L)$	$\frac{P \xrightarrow{(\alpha, r)} P'}{P/L \xrightarrow{(\tau, r)} P'/L} \quad (\alpha \in L)$
$\frac{P \xrightarrow{(\alpha, r)} P'}{P \boxtimes_L Q \xrightarrow{(\alpha, r)} P' \boxtimes_L Q} \quad (\alpha \notin L)$	$\frac{Q \xrightarrow{(\alpha, r)} Q'}{P \boxtimes_L Q \xrightarrow{(\alpha, r)} P \boxtimes_L Q'} \quad (\alpha \notin L)$
$\frac{P \xrightarrow{(\alpha, r_1)} P' \quad Q \xrightarrow{(\alpha, r_2)} Q'}{P \boxtimes_L Q \xrightarrow{(\alpha, R)} P' \boxtimes_L Q'}$	$R = \frac{r_1}{r_\alpha(P)} \frac{r_2}{r_\alpha(Q)} \min(r_\alpha(P), r_\alpha(Q)) \quad (\alpha \in L)$

$P + Q$  represents a system that may behave either as  $P$  or as  $Q$ .  $P + Q$  enables all the current activities of both  $P$  and  $Q$ . Component  $P/L$  behaves as  $P$  except that any activity of type within the set  $L$  are *hidden*, i.e., they are relabeled with the unobservable type  $\tau$ . The meaning of a constant  $A$  is given by a defining equation such as  $A \stackrel{def}{=} P$  that gives the constant  $A$  the behaviour of the component  $P$ . The cooperation combinator  $\boxtimes_L$  is in fact an indexed family of combinators, one for each possible set of action types,  $L \subseteq \mathcal{A} \setminus \{\tau\}$ . The *cooperation set*  $L$  defines the action types on which the components must synchronize or *cooperate* (the unknown action type,  $\tau$ , may not appear in any cooperation set). It is assumed that each component proceeds independently with any activity whose type does not occur in the cooperation set  $L$  (*individual activities*). However, activities with action types in the set  $L$  require the simultaneous involvement of both components (*shared activities*). These shared activities will only be enabled in  $P \boxtimes_L Q$  when they are enabled in both  $P$  and  $Q$ . The shared activity will have the same action type as the two contributing activities and a rate reflecting the rate of the slower participant [21]. If an activity has an unspecified rate in a component then the component is passive with respect to that action type. In this case, the rate of the shared activity will be completely determined by the other component. In general, the rate of a shared activity will reflect the capacity of each component to carry out activities of that type. For a given process  $P$  and action type  $\alpha$ , the *apparent rate* of  $\alpha$  in  $P$ , denoted  $r_\alpha(P)$ , is the sum of the rates of the  $\alpha$  activities enabled in  $P$ .

The semantics of each term  $P$  in PEPA is given via a labeled *multi-transition system*, named *derivation graph* of  $P$  ( $\mathcal{D}(P)$ ) obtained by applying the semantic rules exhaustively and where the multiplicities of arcs are significant. The set of reachable states of a model  $P$  is termed the *derivative set* of  $P$ , denoted by  $ds(P)$ , and constitutes the set of nodes of  $\mathcal{D}(P)$ . We denote by  $\mathcal{A}(P)$  the set of all the *current action types* of  $P$ , i.e. the set of action types that the component  $P$  may next engage in. We denote by  $Act(P)$  the multiset of all the *current activities* of  $P$ . For any component  $P$ , the *exit rate* from  $P$  will be the sum of the activity rates of all the activities enabled in  $P$ , i.e.  $q(P) = \sum_{a \in Act(P)} r_a$ , with  $r_a$  being the rate of activity  $a$ . If  $P$  enables more than one activity,  $|Act(P)| > 1$ , then the dynamic behaviour of the model is determined by a race condition. This has the effect of replacing the nondeterministic branching of the pure process algebra with probabilistic branching. The

probability that a particular activity completes is given by the ratio of the activity rate to the exit rate from  $P$ .

*Underlying Stochastic Process* In [21], it is proved that for any finite PEPA model  $P \stackrel{\text{def}}{=} P_0$  with  $ds(P) = \{P_0, \dots, P_n\}$ , if we define the stochastic process  $X(t)$  for  $t \geq 0$ , such that  $X(t) = P_i$  indicates that the system behaves as component  $P_i$  at time  $t$ , then  $X(t)$  is a continuous-time Markov chain.

The *transition rate* between two components  $P_i$  and  $P_j$ , denoted  $q(P_i, P_j)$ , is the sum of the activity rates labeling arcs that connect the node corresponding to  $P_i$  to the node corresponding to  $P_j$  in  $\mathcal{D}(P)$ , i.e.

$$q(P_i, P_j) = \sum_{a \in \text{Act}(P_i|P_j)} r_a$$

where  $P_i \neq P_j$  and  $\text{Act}(P_i|P_j) = \{a \in \text{Act}(P_i) \mid P_i \xrightarrow{a} P_j\}$ . Clearly, if  $P_j$  is not a one-step derivative of  $P_i$ ,  $q(P_i, P_j) = 0$ . The  $q(P_i, P_j)$  (also denoted  $q_{ij}$ ), are the off-diagonal elements of the infinitesimal generator matrix of the Markov chain,  $\mathbf{Q}$ . Diagonal elements are formed as the negative sum of the non-diagonal elements of each row. We use the following notation:  $q(P_i) = \sum_{j \neq i} q(P_i, P_j)$  and  $q_{ii} = -q(P_i)$ . For any finite and irreducible PEPA model  $P$ , the steady-state distribution  $\pi(\cdot)$  exists and it may be found by solving the normalization equation and the global balance equations:

$$\sum_{P_i \in ds(P)} \pi(P_i) = 1 \quad \wedge \quad \pi \mathbf{Q} = \mathbf{0}.$$

The *conditional transition rate* from  $P_i$  to  $P_j$  via an action type  $\alpha$  is denoted  $q(P_i, P_j, \alpha)$ . This is the sum of the activity rates labeling arcs connecting the corresponding nodes in the derivation graph with label  $\alpha$ . It is the rate at which a system behaving as component  $P_i$  evolves to behaving as component  $P_j$  as the result of completing a type  $\alpha$  activity. The *total conditional transition rate* from  $P$  to  $S \subseteq ds(P)$ , denoted  $q[P, S, \alpha]$ , is defined as

$$q[P, S, \alpha] = \sum_{P' \in S} q(P, P', \alpha).$$

### 3.2 Proportional bisimilarity

In a process algebra, actions, rather than states, play the role of capturing the observable behaviour of a system model. This leads to a formally defined notion of equivalence in which components are regarded as equal if, under observation, they appear to perform exactly the same actions.

Let us first recall the notion of *strong equivalence* for PEPA components introduced in [21]. Two PEPA components are *strongly equivalent* if there is an equivalence relation between them such that, for any action type  $\alpha$ , the total conditional transition rates from those components to any equivalence class, via activities of this type, are the same.

**Definition 5** (Strong equivalence) An equivalence relation over PEPA components,  $\mathcal{R} \subseteq \mathcal{C} \times \mathcal{C}$ , is a *strong equivalence* if whenever  $(P, Q) \in \mathcal{R}$  then for all  $\alpha \in \mathcal{A}$  and for all  $S \in \mathcal{C}/\mathcal{R}$ ,

$$q[P, S, \alpha] = q[Q, S, \alpha].$$



We are interested in the relation that is the largest strong equivalence, which is the union of all strong equivalences. With abuse of notation, we still call such equivalence *strong equivalence* and denote it by  $\cong$ .

**Definition 6** (Strong equivalence  $\cong$ ) We say that two PEPA components  $P$  and  $Q$  are *strongly equivalent*, written  $P \cong Q$ , if  $(P, Q) \in \mathcal{R}$  for some strong equivalence  $\mathcal{R}$ , i.e.

$$\cong = \bigcup \{ \mathcal{R} \mid \mathcal{R} \text{ is a strong equivalence} \}.$$

$\cong$  is the largest strong equivalence over PEPA components.

Strong equivalence induces a partition on the derivative set  $ds(P)$  of  $P$  into equivalence classes that is an *ordinary lumpability* for the underlying Markov chain. As a consequence, the aggregated process satisfies the property that the steady-state probability of each aggregated macro-state is equal to the sum of the steady-state probabilities of the corresponding equivalent states in the initial CTMC.

In [29] the notion of *approximate strong equivalence* for PEPA components inducing a *quasi-lumpability* on the underlying CTMCs [18] has been introduced in order to relax the conditions of strong equivalence. Two PEPA components are *approximately strongly equivalent* if there is an equivalence relation between them such that, for any action type  $\alpha$ , the total conditional transition rates from those components to any equivalence class, via activities of this type, are equal after a small perturbation of the system.

**Definition 7** (Approximate strong equivalence) An equivalence relation over PEPA components,  $\mathcal{R} \subseteq \mathcal{C} \times \mathcal{C}$ , is an *approximate strong equivalence with respect to  $\epsilon$*  with  $\epsilon \geq 0$  if whenever  $(P, Q) \in \mathcal{R}$  then for all  $\alpha \in \mathcal{A}$  and for all  $S \in \mathcal{C}/\mathcal{R}$ ,

$$|q[P, S, \alpha] - q[Q, S, \alpha]| \leq \epsilon, \quad \epsilon \geq 0.$$

It is easy to prove that an approximate strong equivalence over the state space of a PEPA component  $P$  induces a quasi-lumpability on the state space of the Markov chain underlying  $P$ . However approximate strong equivalence is not preserved under union. In [29], a partitioning strategy for PEPA components that involves the use of a clustering algorithm that minimizes an upper bound for approximate strong equivalence is proposed.

In this paper, we follow a different approach and introduce the notion of *proportional bisimulation* to relax the notion of strong equivalence.

**Definition 8** (Proportional bisimulation) An equivalence relation over PEPA components,  $\mathcal{R} \subseteq \mathcal{C} \times \mathcal{C}$ , is a *proportional bisimulation* if there exists a function  $\kappa$  from  $\mathcal{C}$  to  $\mathbb{R}^+$  such that whenever  $(P, Q) \in \mathcal{R}$  then for all  $\alpha \in \mathcal{A}$  and for all  $S \in \mathcal{C}/\mathcal{R}$ ,

$$\frac{q[P, S, \alpha]}{\kappa(P)} = \frac{q[Q, S, \alpha]}{\kappa(Q)}.$$

The following proposition provides a characterization of proportional bisimulation in terms of the exit rate from a process  $P$ .

**Proposition 5** (Characterization of proportional bisimulation) *An equivalence relation over PEPA components,  $\mathcal{R} \subseteq \mathcal{C} \times \mathcal{C}$ , is a proportional bisimulation if and only if whenever  $(P, Q) \in \mathcal{R}$  then*

- either  $q(P) = q(Q) = 0$

– or  $q(P), q(Q) > 0$  and for all  $\alpha \in \mathcal{A}$  and for all  $S \in \mathcal{C}/\mathcal{R}$

$$\frac{q[P, S, \alpha]}{q(P)} = \frac{q[Q, S, \alpha]}{q(Q)}.$$

**Proof**  $\Rightarrow$ ) Suppose that  $\mathcal{R}$  is a proportional bisimulation for a function  $\kappa$  from  $\mathcal{C}$  to  $\mathbb{R}^+$ , i.e. for any  $(P, Q) \in \mathcal{R}$  then for all  $\alpha \in \mathcal{A}$  and for all  $S \in \mathcal{C}/\mathcal{R}$ ,

$$\frac{q[P, S, \alpha]}{\kappa(P)} = \frac{q[Q, S, \alpha]}{\kappa(Q)}. \tag{6}$$

If  $q(P) = 0$  then for all  $\alpha \in \mathcal{A}$  and for all  $S \in \mathcal{C}/\mathcal{R}$ ,  $q[P, S, \alpha] = 0$  then also  $q[Q, S, \alpha] = 0$ , i.e.  $q(Q) = 0$  and, viceversa, if  $q(Q) = 0$  then also  $q(P) = 0$ . Assume that  $q(P), q(Q) > 0$ . By summing the left and right terms of the above equation over all  $\alpha \in \mathcal{A}$  and all  $S \in \mathcal{C}/\mathcal{R}$  we get

$$\frac{\sum_{\alpha \in \mathcal{A}} (\sum_{S \in \mathcal{C}/\mathcal{R}} q[P, S, \alpha])}{\kappa(P)} = \frac{\sum_{\alpha \in \mathcal{A}} (\sum_{S \in \mathcal{C}/\mathcal{R}} q[Q, S, \alpha])}{\kappa(Q)}$$

that can be written as

$$\frac{q(P)}{\kappa(P)} = \frac{q(Q)}{\kappa(Q)}.$$

Since, by hypothesis,  $q(P), q(Q) > 0$ , we get

$$\frac{\kappa(P)}{q(P)} = \frac{\kappa(Q)}{q(Q)}. \tag{7}$$

Now by multiplying the left and right terms of Eqs. (6) and (7) we have that for any  $\alpha \in \mathcal{A}$  and for any equivalence class  $S \in \mathcal{C}/\mathcal{R}$

$$\frac{q[P, S, \alpha]}{\kappa(P)} \cdot \frac{\kappa(P)}{q(P)} = \frac{q[Q, S, \alpha]}{\kappa(Q)} \cdot \frac{\kappa(Q)}{q(Q)},$$

and, by simplifying, we get the thesis.

$\Leftarrow$ ) Suppose that  $\mathcal{R} \subseteq \mathcal{C} \times \mathcal{C}$  is an equivalence relation such that whenever  $(P, Q) \in \mathcal{R}$  then for all  $\alpha \in \mathcal{A}$  and for all  $S \in \mathcal{C}/\mathcal{R}$

- either  $q(P) = q(Q) = 0$
- or  $q(P), q(Q) > 0$  and

$$\frac{q[P, S, \alpha]}{q(P)} = \frac{q[Q, S, \alpha]}{q(Q)}.$$

Then, by Definition 8 of proportional bisimulation,  $\mathcal{R}$  is a proportional bisimulation with respect to the function  $\kappa$  from  $\mathcal{C}$  to  $\mathbb{R}^+$  such that  $\kappa(P) = q(P)$  whenever  $q(P) > 0$ , and  $\kappa(P)$  is an arbitrary positive real number otherwise.  $\square$

Notice that, similarly to the conditions for strong equivalence in PEPA, the conditions for proportional bisimulations require us to consider also the transitions within an equivalence class. This is necessary for ensuring the compositionality properties studied at the end of this section.

We are interested in the relation that is the largest proportional bisimulation, formed by the union of all proportional bisimulations.

The following proposition states that the transitive closure of a union of proportional bisimulations generates a proportional bisimulation.

**Proposition 6** (Transitive closure of a union of proportional bisimulations) *Let  $I$  be a set of indices and  $\mathcal{R}_i$  be a proportional bisimulation for all  $i \in I$ . Then the transitive closure of their union,  $\mathcal{R} = (\cup_{i \in I} \mathcal{R}_i)^+$ , is also a proportional bisimulation.*

**Proof** From the fact that each  $\mathcal{R}_i$  is an equivalence relation, it follows, by the definition of  $\mathcal{R}$ , that  $\mathcal{R}$  is also an equivalence relation.

Let  $\mathcal{C}/\mathcal{R}$  and  $\mathcal{C}/\mathcal{R}_i$  be the sets of equivalence classes for  $\mathcal{R}$  and each  $\mathcal{R}_i$ , respectively. By definition,  $(P, Q) \in \mathcal{R}_i$  implies that  $(P, Q) \in \mathcal{R}$ , and so any equivalence class  $S_j^i \in \mathcal{C}/\mathcal{R}_i$  is wholly contained within some equivalence class  $S_k \in \mathcal{C}/\mathcal{R}$ . Moreover, there is some set  $I_k^i$  of indices such that  $S_k = \cup_{j \in I_k^i} S_j^i$ .

Let us denote by  $\mathcal{R}^n$  the relation  $(\cup_{i \in I} \mathcal{R}_i)^n$  defined by:  $(\cup_{i \in I} \mathcal{R}_i)^1 = (\cup_{i \in I} \mathcal{R}_i)$  and  $(\cup_{i \in I} \mathcal{R}_i)^n = \{(P, Q) \mid \text{there exists } C \in \mathcal{C} \text{ such that } (P, C) \in \mathcal{R}_i \text{ for some } i \in I \text{ and } (C, Q) \in \mathcal{R}^{n-1}\}$ . Let  $(P, Q) \in \mathcal{R}$ , then  $(P, Q) \in (\cup_{i \in I} \mathcal{R}_i)^n$  for some  $n > 0$ .

In order to prove that  $\mathcal{R}$  is a proportional bisimulation, we prove that for each  $\mathcal{R}^n$  if  $(P, Q) \in \mathcal{R}^n$  then for all  $\alpha \in \mathcal{A}$  and for all  $S_k \in \mathcal{C}/\mathcal{R}$

- either  $q(P) = q(Q) = 0$
- or  $q(P), q(Q) > 0$  and

$$\frac{q[P, S_k, \alpha]}{q(P)} = \frac{q[Q, S_k, \alpha]}{q(Q)}.$$

We proceed by induction on  $n$ .

*Base case* Let  $n = 1$ . Then  $(P, Q) \in \mathcal{R}^1$  implies that  $(P, Q) \in \mathcal{R}_i$  for some  $i \in I$ . Since  $\mathcal{R}_i$  is a proportional bisimulation, for all  $\alpha \in \mathcal{A}$  and for all  $S_j^i \in \mathcal{C}/\mathcal{R}_i$

- either  $q(P) = q(Q) = 0$
- or  $q(P), q(Q) > 0$  and

$$\frac{q[P, S_j^i, \alpha]}{q(P)} = \frac{q[Q, S_j^i, \alpha]}{q(Q)}.$$

Assume that  $q(P), q(Q) > 0$ . Let  $S_k \in \mathcal{C}/\mathcal{R}$ . Since  $S_k = \cup_{j \in I_k^i} S_j^i$  for some set  $I_k^i$  we have

$$\frac{q[P, S_k, \alpha]}{q(P)} = \frac{\sum_{j \in I_k^i} q[P, S_j^i, \alpha]}{q(P)} = \frac{\sum_{j \in I_k^i} q[Q, S_j^i, \alpha]}{q(Q)} = \frac{q[Q, S_k, \alpha]}{q(Q)}.$$

*Inductive step* Let  $n > 1$ . We assume that for all  $\mathcal{R}^m$  with  $m < n$ , whenever  $(P, Q) \in \mathcal{R}^m$  then for all  $\alpha \in \mathcal{A}$  and for all  $S_k \in \mathcal{C}/\mathcal{R}$  it holds

$$\frac{q[P, S_k, \alpha]}{q(P)} = \frac{q[Q, S_k, \alpha]}{q(Q)}.$$

If  $(P, Q) \in \mathcal{R}^n$  then there exists  $C \in \mathcal{C}$  such that  $(P, C) \in \mathcal{R}_i$  for some  $i \in I$  and  $(C, Q) \in \mathcal{R}^{n-1}$ . Let  $S_k \in \mathcal{C}/\mathcal{R}$  and  $\alpha \in \mathcal{A}$ .

Assume that  $q(P) = q(C) = 0$ . From the fact that  $q(C) = 0$ , by induction hypothesis we have that also  $q(Q) = 0$  and hence  $q(P) = q(Q) = 0$ .

Assume now that  $q(P), q(C) > 0$ , then by induction hypothesis also  $q(Q) > 0$ . Since each  $S_k = \cup_{j \in I_k^i} S_j^i$  for some set  $I_k^i$ , by the same argument as above,

$$\frac{q[P, S_k, \alpha]}{q(P)} = \frac{q[C, S_k, \alpha]}{q(C)},$$

moreover, by the induction hypothesis,

$$\frac{q[C, S_k, \alpha]}{q(C)} = \frac{q[Q, S_k, \alpha]}{q(Q)},$$

and then we can conclude

$$\frac{q[P, S_k, \alpha]}{q(P)} = \frac{q[Q, S_k, \alpha]}{q(Q)}$$

as required. Hence  $\mathcal{R}$  is a proportional bisimulation. □

Based on the above result we can define the maximal proportional bisimulation as the union of all proportional bisimulations.

**Definition 9** (Proportional bisimilarity) Two PEPA components  $P$  and  $Q$  are *proportionally bisimilar*, written  $P \approx_p Q$ , if  $(P, Q) \in \mathcal{R}$  for some proportional bisimulation  $\mathcal{R}$ , i.e.

$$\approx_p = \bigcup \{ \mathcal{R} \mid \mathcal{R} \text{ is a proportional bisimulation} \}.$$

$\approx_p$  is called *proportional bisimilarity* and it is the largest proportional bisimulation over PEPA components.

The relation  $\approx_p$  partitions the set of components  $\mathcal{C}$ , and it is easy to see that if restricted to the derivative set of any component  $P$ , the relation partitions this set. Let  $ds(P)/\approx_p$  denote the set of equivalence classes generated in this way. It is easy to prove the following result.

**Proposition 7** (Proportional bisimilarity vs proportional lumpability) *For any PEPA component  $P$ ,  $ds(P)/\approx_p$  induces a proportional lumpability on the state space of the Markov chain corresponding to  $P$  whenever it is irreducible.*

**Proof** Let  $S_i, S_j \in ds(P)/\approx_p$  such that  $i \neq j$  and consider two elements  $P, P' \in S_i$ . By Definition 8, there exists a function  $\kappa$  from  $\mathcal{C}$  to  $\mathbb{R}^+$  such that

$$\frac{\sum_{\alpha \in \mathcal{A}} q[P, S_j, \alpha]}{\kappa(P)} = \frac{\sum_{\alpha \in \mathcal{A}} q[P', S_j, \alpha]}{\kappa(P')}$$

that can be rewritten as

$$\frac{\sum_{P'' \in S_j} q(P, P'')}{\kappa(P)} = \frac{\sum_{P'' \in S_j} q(P', P'')}{\kappa(P')}.$$

Hence, by Definition 3, it follows immediately that the partition  $ds(P)/\approx_p$  induces a proportional lumpability on the state space of the Markov chain corresponding to  $P$ . □

Moreover, the relation  $\approx_p$  can be efficiently computed exploiting classical lumpability algorithms [17].

**Proposition 8** (Proportional bisimilarity algorithm) *The quotient  $ds(P)/\approx_p$  can be computed in time  $O(m \log n)$ , where  $m$  is the number of edges and  $n$  the number of states of  $\mathcal{D}(P)$ .*

**Proof** As a consequence of Proposition 5, the quotient  $ds(P)/\approx_p$  can be computed as follows. Let  $Disc(P)$  be the multi-transition system obtained from  $\mathcal{D}(P)$  by replacing in each transition  $Q \xrightarrow{(\alpha, r)} Q'$  the label  $(\alpha, r)$  with  $(\alpha, \frac{r}{q(Q)})$ . The algorithm described in [2] on input graph  $Disc(P)$  and initial partition  $\{ds(P)\}$  returns the quotient  $ds(P)/\approx_p$ . □

*Properties of proportional bisimilarity* We prove some compositionality properties of proportional bisimilarity  $\approx_p$ . First we show that proportional bisimilarity is preserved by prefix and hiding operators.

**Proposition 9** (Preservation by prefix and hiding) *Let  $P_1$  and  $P_2$  be two PEPA components. If  $P_1 \approx_p P_2$  then for all activity  $a = (\alpha, r)$  and for all  $L \subseteq \mathcal{A} \setminus \{\tau\}$  it holds:*

1.  $a.P_1 \approx_p a.P_2$ ,
2.  $P_1/L \approx_p P_2/L$ .

**Proof** Since  $P_1 \approx_p P_2$  then, by Proposition 5, for all  $\alpha \in \mathcal{A}$  and for all  $S \in \mathcal{C}/\approx_p$

- either  $q(P_1) = q(P_2) = 0$
- or  $q(P_1), q(P_2) > 0$  and

$$\frac{q[P_1, S, \alpha]}{q(P_1)} = \frac{q[P_2, S, \alpha]}{q(P_2)}.$$

1. Let  $a = (\alpha, r)$  and

$$\mathcal{R} = \{(a.P_1, a.P_2) \mid P_1 \approx_p P_2\}.$$

We prove that  $\mathcal{R}^+ = \mathcal{R} \cup Id$  is a proportional lumpability. Indeed,  $q(a.P_1) = q(a.P_2) = r$  and then it follows trivially that for all  $S \in \mathcal{C}/\mathcal{R}^+$

$$\frac{q[a.P_1, S, \alpha]}{q(a.P_1)} = \frac{q[a.P_2, S, \alpha]}{q(a.P_2)},$$

i.e.,  $a.P_1 \approx_p a.P_2$ .

2. Let  $L \subseteq \mathcal{A} \setminus \{\tau\}$  and

$$\mathcal{R} = \{(P_1/L, P_2/L) \mid P_1 \approx_p P_2\}.$$

We prove that  $\mathcal{R}^+ = \mathcal{R} \cup Id$  is a proportional lumpability. First, it is easy to see that  $q(P_1/L) = q(P_1)$  and  $q(P_2/L) = q(P_2)$ . Moreover for all  $\alpha \neq L$ , by the hypothesis that  $P_1 \approx_p P_2$ , we have

$$\frac{q[P_1/L, S, \alpha]}{q(P_1/L)} = \frac{q[P_1, S, \alpha]}{q(P_1)} = \frac{q[P_2, S, \alpha]}{q(P_2)} = \frac{q[P_2/L, S, \alpha]}{q(P_2/L)}.$$

Now, consider the unknown type  $\tau$ . By the operational semantics for PEPA components presented in Table 1 and the hypothesis that  $P_1 \approx_p P_2$ , we have

$$\begin{aligned} \frac{q[P_1/L, S, \tau]}{q(P_1/L)} &= \frac{q[P_1, S, \tau] + \sum_{\alpha \in L} q[P_1, S, \alpha]}{q(P_1)} \\ &= \frac{q[P_2, S, \tau] + \sum_{\alpha \in L} q[P_2, S, \alpha]}{q(P_2)} \\ &= \frac{q[P_2/L, S, \tau]}{q(P_2/L)}, \end{aligned}$$

i.e.  $P_1/L \approx_p P_2/L$ .

□

Preservation by choice does not hold in general.

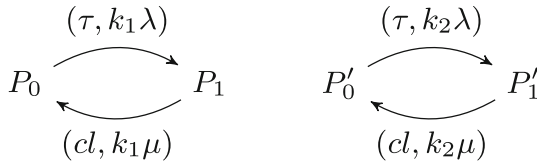


Fig. 6 Two simple proportionally bisimilar processes

**Example 4** Consider the simple proportionally bisimilar PEPA processes depicted in Fig. 6 where  $k_1$  and  $k_2$  are constants with  $k_1 \neq k_2$ . It is easy to see that  $P_0 \approx_p P'_0$  and  $P_1 \approx_p P'_1$ . However, for an arbitrary PEPA model  $Q$ , we cannot always infer that  $P_0 + Q \approx_p P'_0 + Q$ . Indeed, let  $S = \{P_1, P'_1\}$ ,  $q(Q) = \rho$  and  $q[Q, S, \tau] = 0$  then

$$\frac{q[P_0 + Q, S, \tau]}{q(P_0 + Q)} = \frac{k_1 \lambda}{k_1 \lambda + k_1 \mu + \rho}$$

and

$$\frac{q[P'_0 + Q, S, \tau]}{q(P'_0 + Q)} = \frac{k_2 \lambda}{k_2 \lambda + k_2 \mu + \rho},$$

and, in general

$$\frac{k_1 \lambda}{k_1 \lambda + k_1 \mu + \rho} \neq \frac{k_2 \lambda}{k_2 \lambda + k_2 \mu + \rho}.$$

The following proposition shows that proportional bisimilarity is preserved under the choice operator when the equivalent components satisfy some condition on the exit rates.

**Proposition 10** (Preservation by choice) *Let  $P_1, P_2$  and  $Q$  be PEPA components. If  $P_1 \approx_p P_2$  and  $q(P_1) = q(P_2) = kq(Q)$  where  $k > 0$  is a constant then  $P_1 + Q \approx_p P_2 + Q$ .*

**Proof** Since  $P_1 \approx_p P_2$  then, by Proposition 5, for all  $\alpha \in \mathcal{A}$  and for all  $S \in \mathcal{C}/\approx_p$

- either  $q(P_1) = q(P_2) = 0$
- or  $q(P_1), q(P_2) > 0$  and

$$\frac{q[P_1, S, \alpha]}{q(P_1)} = \frac{q[P_2, S, \alpha]}{q(P_2)}.$$

If  $q(P_1) = q(P_2) = q(Q) = 0$ , then we have the thesis. In the other case, since by hypothesis  $q(P_1) = q(P_2) = kq(Q)$  we have  $q(P_1 + Q) = \frac{k+1}{k}q(P_1) = \frac{k+1}{k}q(P_2) = q(P_2 + Q)$ . Let  $h = \frac{k+1}{k}$ . Hence,

$$\begin{aligned} \frac{q[P_1 + Q, S, \alpha]}{q(P_1 + Q)} &= \frac{q[P_1, S, \alpha] + q[Q, S, \alpha]}{hq(P_1)} \\ &= \frac{q[P_2, S, \alpha] + q[Q, S, \alpha]}{hq(P_2)} \\ &= \frac{q[P_2 + Q, S, \alpha]}{q(P_2 + Q)}. \end{aligned}$$

Thus we conclude that  $P_1 + Q \approx_p P_2 + Q$ . □

Notice that the requirement in the above proposition that  $q(P_1) = q(P_2)$  is not required for their derivatives and hence does not entail that they are strongly equivalent.

In order to prove that, under some conditions, proportional bisimilarity is preserved by cooperation, we first provide a further useful characterization of proportional bisimulation.

**Proposition 11** (Second characterization of proportional bisimulation) *An equivalence relation over PEPA components,  $\mathcal{R} \subseteq \mathcal{C} \times \mathcal{C}$ , is a proportional bisimulation if and only if whenever  $(P, Q) \in \mathcal{R}$  then*

- either  $q(P) = q(Q) = 0$
- or  $q(P), q(Q) > 0$  and for all  $\alpha \in \mathcal{A}$  and for all  $S \in \mathcal{C}/\mathcal{R}$ 
  - either  $r_\alpha(P) = r_\alpha(Q) = 0$
  - or  $r_\alpha(P), r_\alpha(Q) > 0$  and

$$\frac{r_\alpha(P)}{q(P)} = \frac{r_\alpha(Q)}{q(Q)} \quad \text{and} \quad \frac{q[P, S, \alpha]}{r_\alpha(P)} = \frac{q[Q, S, \alpha]}{r_\alpha(Q)}.$$

**Proof**  $\Rightarrow$ ) By Proposition 5, we have that, since  $\mathcal{R}$  is a proportional bisimulation, whenever  $(P, Q) \in \mathcal{R}$  it holds that

- either  $q(P) = q(Q) = 0$
- or  $q(P), q(Q) > 0$  and for all  $\alpha \in \mathcal{A}$  and for all  $S \in \mathcal{C}/\mathcal{R}$

$$\frac{q[P, S, \alpha]}{q(P)} = \frac{q[Q, S, \alpha]}{q(Q)}.$$

Let  $(P, Q) \in \mathcal{R}$  with  $q(P), q(Q) > 0$ . From the last equation we have that for each  $\alpha \in \mathcal{A}$  it holds  $r_\alpha(P) = \sum_{S \in \mathcal{C}/\mathcal{R}} q[P, S, \alpha] = 0$  if and only if  $r_\alpha(Q) = \sum_{S \in \mathcal{C}/\mathcal{R}} q[Q, S, \alpha] = 0$ . So we have to consider the case  $r_\alpha(P), r_\alpha(Q) > 0$  and we have to prove that for all  $S \in \mathcal{C}/\mathcal{R}$

$$\frac{r_\alpha(P)}{q(P)} = \frac{r_\alpha(Q)}{q(Q)} \quad \text{and} \quad \frac{q[P, S, \alpha]}{r_\alpha(P)} = \frac{q[Q, S, \alpha]}{r_\alpha(Q)}.$$

Since for all  $S \in \mathcal{C}/\mathcal{R}$

$$\frac{q[P, S, \alpha]}{q(P)} = \frac{q[Q, S, \alpha]}{q(Q)}$$

by summing over all the classes of  $\mathcal{C}/\mathcal{R}$  we get

$$\frac{r_\alpha(P)}{q(P)} = \frac{\sum_{S \in \mathcal{C}/\mathcal{R}} q[P, S, \alpha]}{q(P)} = \frac{\sum_{S \in \mathcal{C}/\mathcal{R}} q[Q, S, \alpha]}{q(Q)} = \frac{r_\alpha(Q)}{q(Q)}.$$

Now since for all  $S \in \mathcal{C}/\mathcal{R}$

$$\frac{q[P, S, \alpha]}{q(P)} = \frac{q[Q, S, \alpha]}{q(Q)} \quad \text{and} \quad \frac{r_\alpha(P)}{q(P)} = \frac{r_\alpha(Q)}{q(Q)}$$

by multiplying the terms of the first equation by the inverse of the terms of the second equation we get

$$\frac{q[P, S, \alpha]}{r_\alpha(P)} = \frac{q[Q, S, \alpha]}{r_\alpha(Q)}.$$

$\Leftarrow$ ) We have to prove that  $\mathcal{R}$  is a proportional bisimulation. We can equivalently prove that  $\mathcal{R}$  satisfies Proposition 5. In particular, we have to prove that if  $q(P), q(Q) > 0$ , then for each  $\alpha \in \mathcal{A}$  and for each  $S \in \mathcal{C}/\mathcal{R}$  it holds that

$$\frac{q[P, S, \alpha]}{q(P)} = \frac{q[Q, S, \alpha]}{q(Q)}.$$

If  $r_\alpha(P) = r_\alpha(Q) = 0$ , then we get the thesis. Let us now consider the case  $r_\alpha(P), r_\alpha(Q) > 0$ . In this case from

$$\frac{r_\alpha(P)}{q(P)} = \frac{r_\alpha(Q)}{q(Q)} \quad \text{and} \quad \frac{q[P, S, \alpha]}{r_\alpha(P)} = \frac{q[Q, S, \alpha]}{r_\alpha(Q)}$$

by multiplying the terms of the two equations we get the thesis. □

By applying the above characterization to  $\approx_p$ , that is the largest proportional bisimulation, we obtain the following result.

**Corollary 1** *Let  $P$  and  $Q$  be PEPA components. It holds that  $P \approx_p Q$  if and only if:*

- either  $q(P) = q(Q) = 0$
- or  $q(P), q(Q) > 0$  and for each  $\alpha \in \mathcal{A}$  and for each  $S \in \mathcal{C}/\approx_p$ :
  - either  $r_\alpha(P) = r_\alpha(Q) = 0$
  - or  $r_\alpha(P), r_\alpha(Q) > 0$  and

$$\frac{r_\alpha(P)}{q(P)} = \frac{r_\alpha(Q)}{q(Q)} \quad \text{and} \quad \frac{q[P, S, \alpha]}{r_\alpha(P)} = \frac{q[Q, S, \alpha]}{r_\alpha(Q)}.$$

Moreover, the following necessary condition can be used to easily prove that two components are not proportionally bisimilar.

**Corollary 2** (Necessary conditions for proportional bisimilarity) *Let  $P$  and  $Q$  be two components. If  $P \approx_p Q$  then:*

- either  $q(P) = q(Q) = 0$
- or  $q(P), q(Q) > 0$  and  $\mathcal{A}(P) = \mathcal{A}(Q)$  and for each  $\alpha \in \mathcal{A}(P)$  it holds that

$$\frac{r_\alpha(P)}{q(P)} = \frac{r_\alpha(Q)}{q(Q)}.$$

As preservation by choice, also preservation by cooperation does not hold in general.

**Example 5** Consider again the simple proportionally bisimilar PEPA processes depicted in Figure 6 where  $k_1$  and  $k_2$  are constants with  $k_1 \neq k_2$ ,  $P_0 \approx_p P'_0$  and  $P_1 \approx_p P'_1$ . For an arbitrary PEPA model  $Q$ , we cannot always infer that  $P_0 \boxtimes_L Q \approx_p P'_0 \boxtimes_L Q$ . Indeed, let  $S = \{P_1, P'_1\}$ ,  $q(Q) = \rho$ ,  $q[Q, S, \tau] = 0$  and  $L \subseteq \mathcal{A} \setminus \{\tau\}$  such that both  $P_0$  and  $P'_0$  do not synchronize with  $Q$ . Then

$$\frac{q[P_0 \boxtimes_L Q, S, \tau]}{q(P_0 \boxtimes_L Q)} = \frac{k_1 \lambda}{k_1 \lambda + k_1 \mu + \rho}$$

and

$$\frac{q[P'_0 \boxtimes_L Q, S, \tau]}{q(P'_0 \boxtimes_L Q)} = \frac{k_2 \lambda}{k_2 \lambda + k_2 \mu + \rho},$$

and, in general

$$\frac{k_1 \lambda}{k_1 \lambda + k_1 \mu + \rho} \neq \frac{k_2 \lambda}{k_2 \lambda + k_2 \mu + \rho}.$$

The following proposition proves that, under some conditions, proportional bisimulation is preserved by cooperation.



**Proposition 12** (Preservation by cooperation) *Let  $P_1, P_2$  and  $Q$  be PEPA components and  $L \subseteq \mathcal{A} \setminus \{\tau\}$ . If*

1.  $P_1 \approx_p P_2$ ,
2. for all  $P'_1 \in ds(P_1), Q' \in ds(Q)$  it holds  $(\mathcal{A}(P'_1) \setminus L) \cap (\mathcal{A}(Q') \setminus L) = \emptyset$ ,
3. for all  $P'_1 \in ds(P_1), P'_2 \in ds(P_2), Q' \in ds(Q)$ , if  $P'_1 \approx_p P'_2$ , then:
  - either  $q(P'_1 \underset{L}{\boxtimes} Q') = q(P'_2 \underset{L}{\boxtimes} Q') = 0$
  - or  $q(P'_1 \underset{L}{\boxtimes} Q'), q(P'_2 \underset{L}{\boxtimes} Q') > 0$  and for all  $\alpha \in \mathcal{A}$

$$\frac{r_\alpha(P'_1 \underset{L}{\boxtimes} Q')}{q(P'_1 \underset{L}{\boxtimes} Q')} = \frac{r_\alpha(P'_2 \underset{L}{\boxtimes} Q')}{q(P'_2 \underset{L}{\boxtimes} Q')}$$

then  $P_1 \underset{L}{\boxtimes} Q \approx_p P_2 \underset{L}{\boxtimes} Q$ .

**Proof** We first observe that since  $P_1 \approx_p P_2$  then for each  $P'_1 \in ds(P_1)$  there exists at least one  $P'_2 \in ds(P_2)$  such that  $P'_1 \approx_p P'_2$ . As a consequence, hypothesis 3 implies that for all  $A, B \in ds(P_1) \cup ds(P_2), Q' \in ds(Q)$  if  $A \approx_p B$ , then:

- either  $q(A \underset{L}{\boxtimes} Q') = q(B \underset{L}{\boxtimes} Q') = 0$
- or  $q(A \underset{L}{\boxtimes} Q'), q(B \underset{L}{\boxtimes} Q') > 0$  and for all  $\alpha \in \mathcal{A}$

$$\frac{r_\alpha(A \underset{L}{\boxtimes} Q')}{q(A \underset{L}{\boxtimes} Q')} = \frac{r_\alpha(B \underset{L}{\boxtimes} Q')}{q(B \underset{L}{\boxtimes} Q')}.$$

This is stronger than hypothesis 3 since it just requires  $A, B \in ds(P_1) \cup ds(P_2)$  instead of  $A \in ds(P_1)$  and  $B \in ds(P_2)$ . Consider the relation

$$\mathcal{R} = \{(A \underset{L}{\boxtimes} Q', B \underset{L}{\boxtimes} Q') \mid A \approx_p B \text{ and } A, B \in ds(P_1) \cup ds(P_2) \text{ and } Q' \in ds(Q)\}.$$

We extend this relation to a relation  $\mathcal{R}^\dagger$  over all components, where  $\mathcal{R}^\dagger = \mathcal{R} \cup Id$ . We will show that  $\mathcal{R}^\dagger$  is a proportional bisimulation. Since  $\approx_p$  is an equivalence relation,  $\mathcal{R}^\dagger$  is an equivalence relation. We will prove that  $\mathcal{R}^\dagger$  satisfies Proposition 11.

In particular, hypothesis 3 in its stronger version ensures that we only have to prove that if  $(A \underset{L}{\boxtimes} Q', B \underset{L}{\boxtimes} Q') \in \mathcal{R}$ , then

- either  $q(A \underset{L}{\boxtimes} Q') = q(B \underset{L}{\boxtimes} Q') = 0$
- or  $q(A \underset{L}{\boxtimes} Q'), q(B \underset{L}{\boxtimes} Q') > 0$  and for all  $\alpha \in \mathcal{A}$  and for all  $T \in \mathcal{C}/\mathcal{R}^\dagger$ 
  - either  $r_\alpha(A \underset{L}{\boxtimes} Q') = r_\alpha(B \underset{L}{\boxtimes} Q') = 0$
  - or  $r_\alpha(A \underset{L}{\boxtimes} Q'), r_\alpha(B \underset{L}{\boxtimes} Q') > 0$  and

$$\frac{q[A \underset{L}{\boxtimes} Q', T, \alpha]}{r_\alpha(A \underset{L}{\boxtimes} Q')} = \frac{q[B \underset{L}{\boxtimes} Q', T, \alpha]}{r_\alpha(B \underset{L}{\boxtimes} Q')}$$

If  $q(A \underset{L}{\boxtimes} Q') = 0$  then, from the hypothesis that  $A \approx_p B$ , it is easy to prove that also  $q(B \underset{L}{\boxtimes} Q') = 0$ .

Let us consider the case  $q(A \underset{L}{\boxtimes} Q'), q(B \underset{L}{\boxtimes} Q') > 0$ . This means that both  $A \underset{L}{\boxtimes} Q'$  and  $B \underset{L}{\boxtimes} Q'$  have at least one derivative. Notice that any derivative of a cooperation of components will have the form of a cooperation of components. Thus we only consider the

equivalence classes  $T \in \mathcal{C}/\mathcal{R}^\dagger$  such that there is some element  $C \boxtimes_L Q'' \in T$  with  $C \in S$  and  $Q'' \in ds(Q)$ . Then, for some  $S \in \mathcal{C}/\approx_p$ ,

$$T = \{C \boxtimes_L Q'' \mid C \in S \text{ and } C \in ds(P_1) \cup ds(P_2) \text{ and } Q'' \in ds(Q)\}.$$

Moreover, we always have  $q[A \boxtimes_L Q', T, \alpha] = q[A \boxtimes_L Q', T', \alpha]$  and  $q[B \boxtimes_L Q', T, \alpha] = q[B \boxtimes_L Q', T', \alpha]$ , where  $T' = \{C \boxtimes_L Q'' \mid C \in S\}$ . Thus we may denote each such  $T$  as  $T_{(S, Q')}$ . For any equivalence class  $T \in \mathcal{C}/\mathcal{R}^\dagger$  that is not of this form, for all  $\alpha \in \mathcal{A}$ ,  $q[A \boxtimes_L Q, T, \alpha] = 0 = q[B \boxtimes_L Q, T, \alpha]$  and the thesis trivially holds.

If  $r_\alpha(A \boxtimes_L Q') = 0$ , then  $\alpha \notin \mathcal{A}(A \boxtimes_L Q')$ . Hence,  $\alpha \notin \mathcal{A}(B \boxtimes_L Q')$  and  $r_\alpha(B \boxtimes_L Q') = 0$ . On the other hand if  $r_\alpha(A \boxtimes_L Q'), r_\alpha(B \boxtimes_L Q') > 0$ , three cases are possible.

- a.  $\alpha \in \mathcal{A}(A) \setminus L = \mathcal{A}(B) \setminus L$ . In this case, by hypothesis 2,  $\alpha \notin \mathcal{A}(Q')$ .
- b.  $\alpha \in \mathcal{A}(Q') \setminus L$ . In this case, by hypothesis 2,  $\alpha \notin \mathcal{A}(A) = \mathcal{A}(B)$ .
- c.  $\alpha \in L$ .

*Case a* Let  $\alpha \in \mathcal{A}(A) \setminus L$ . We have  $\alpha \notin \mathcal{A}(Q')$ . Only  $A$  can complete activities of type  $\alpha$  and so for all  $T_{(S, Q')} \in \mathcal{C}/\mathcal{R}^\dagger$ ,

$$q[A \boxtimes_L Q', T_{(S, Q')}, \alpha] = \sum_{C \in S} q(A, C, \alpha) = q[A, S, \alpha].$$

The action type  $\alpha$  is also an individual action type for  $B$  in  $B \boxtimes_L Q'$  and by similar reasoning  $q[B \boxtimes_L Q', T_{(S, Q')}, \alpha] = q[B, S, \alpha]$ . In particular,  $r_\alpha(A \boxtimes_L Q') = r_\alpha(A)$  and  $r_\alpha(B \boxtimes_L Q') = r_\alpha(B)$ . By definition of  $\mathcal{R}$  we have  $A \approx_p B$ , and then by Proposition 11,

$$\frac{q[A, S, \alpha]}{r_\alpha(A)} = \frac{q[B, S, \alpha]}{r_\alpha(B)}.$$

Therefore it follows that

$$\frac{q[A \boxtimes_L Q', T_{(S, Q')}, \alpha]}{r_\alpha(A \boxtimes_L Q')} = \frac{q[A, S, \alpha]}{r_\alpha(A)} = \frac{q[B, S, \alpha]}{r_\alpha(B)} = \frac{q[B \boxtimes_L Q', T_{(S, Q')}, \alpha]}{r_\alpha(B \boxtimes_L Q')}.$$

*Case b* Let  $\alpha \in \mathcal{A}(Q') \setminus L$ . In this case only  $Q'$  can complete activities of type  $\alpha$  so  $A \boxtimes_L Q' \xrightarrow{(\alpha, r)} A \boxtimes_L Q''$  for some  $Q''$  and similarly  $B \boxtimes_L Q' \xrightarrow{(\alpha, r)} B \boxtimes_L Q''$ . Therefore  $r_\alpha(A \boxtimes_L Q') = r_\alpha(B \boxtimes_L Q') = r_\alpha(Q')$ . By definition of  $\mathcal{R}$ ,  $A \boxtimes_L Q''$  and  $B \boxtimes_L Q''$  will lie within the same equivalence class  $T_{(S, Q')}$ , and so, for all  $T_{(S, Q')} \in \mathcal{C}/\mathcal{R}^\dagger$ ,

$$q[A \boxtimes_L Q', T_{(S, Q')}, \alpha] = q[B \boxtimes_L Q', T_{(S, Q')}, \alpha],$$

and hence,

$$\begin{aligned} \frac{q[A \boxtimes_L Q', T_{(S, Q')}, \alpha]}{r_\alpha(A \boxtimes_L Q')} &= \frac{q[A \boxtimes_L Q', T_{(S, Q')}, \alpha]}{r_\alpha(Q')} \\ &= \frac{q[B \boxtimes_L Q', T_{(S, Q')}, \alpha]}{r_\alpha(Q')} \\ &= \frac{q[B \boxtimes_L Q', T_{(S, Q')}, \alpha]}{r_\alpha(B \boxtimes_L Q')}. \end{aligned}$$

*Case c* Let  $\alpha \in L, \alpha \in \mathcal{A}(A \boxtimes_L Q')$ . In this case  $\alpha$  is a shared action type of  $A$  and  $Q'$ . In general for a shared action type  $\alpha$ ,

$$q(P \boxtimes_L Q, P' \boxtimes_L Q', \alpha) = \frac{q(P, P', \alpha)}{r_\alpha(P)} \cdot \frac{q(Q, Q', \alpha)}{r_\alpha(Q)} \cdot \min(r_\alpha(P), r_\alpha(Q))$$

and  $\min(r_\alpha(P), r_\alpha(Q)) = r_\alpha(P \boxtimes_L Q)$ . Since by definition of  $\mathcal{R}$ ,  $A \approx_p B$ , then by Proposition 11, for all  $S \in \mathcal{C}/\approx_p$

$$\frac{q[A, S, \alpha]}{r_\alpha(A)} = \frac{q[B, S, \alpha]}{r_\alpha(B)}.$$

Now we consider  $q[A \boxtimes_L Q', T_{(S, Q'')}, \alpha]$  for arbitrary  $T_{(S, Q'')} \in \mathcal{C}/\mathcal{R}^\dagger$ :

$$\begin{aligned} & \frac{q[A \boxtimes_L Q', T_{(S, Q'')}, \alpha]}{r_\alpha(A \boxtimes_L Q')} \\ &= \frac{\sum_{C \boxtimes_L Q'' \in T_{(S, Q'')}} q(A \boxtimes_L Q', C \boxtimes_L Q'', \alpha)}{r_\alpha(A \boxtimes_L Q')} \\ &= \frac{1}{r_\alpha(A \boxtimes_L Q')} \cdot \frac{q(Q', Q'', \alpha)}{r_\alpha(Q')} \cdot \frac{\sum_{C \in S} q(A, C, \alpha)}{r_\alpha(A)} \cdot r_\alpha(A \boxtimes_L Q') \\ &= \frac{q(Q', Q'', \alpha)}{r_\alpha(Q')} \cdot \frac{q[A, S, \alpha]}{r_\alpha(A)} \\ &= \frac{q(Q', Q'', \alpha)}{r_\alpha(Q')} \cdot \frac{q[B, S, \alpha]}{r_\alpha(B)} \\ &= \frac{1}{r_\alpha(B \boxtimes_L Q')} \cdot \frac{q(Q', Q'', \alpha)}{r_\alpha(Q')} \cdot \frac{\sum_{C \in S} q(B, C, \alpha)}{r_\alpha(B)} \cdot r_\alpha(B \boxtimes_L Q') \\ &= \frac{\sum_{C \boxtimes_L Q'' \in T_{(S, Q'')}} q(B \boxtimes_L Q', C \boxtimes_L Q'', \alpha)}{r_\alpha(B \boxtimes_L Q')} \\ &= \frac{q[B \boxtimes_L Q', T_{(S, Q'')}, \alpha]}{r_\alpha(B \boxtimes_L Q')}. \end{aligned}$$

□

Proposition 12 looks mainly of theoretical interest, since it requires the computation of the derivation graphs of both  $P_1 \boxtimes_L Q$  and  $P_2 \boxtimes_L Q$  in order to check condition 3. However, it allows us to prove the following compositionality result which avoids such computation.

**Corollary 3** *Let  $P_1, P_2$  and  $Q$  be PEPA components and  $L \subseteq \mathcal{A} \setminus \{\tau\}$ . If*

1.  $P_1 \approx_p P_2$ ,
2. for all  $Q' \in ds(Q)$  it holds  $\mathcal{A}(Q') \subseteq L$ ,
3. for all  $P'_i \in ds(P_1) \cup ds(P_2), Q' \in ds(Q)$  and  $\alpha \in \mathcal{A}(P'_i) \cap \mathcal{A}(Q') \cap L$ ,  $\min(r_\alpha(P'_i), r_\alpha(Q')) = r_\alpha(P'_i)$ ,

then  $P_1 \boxtimes_L Q \approx_p P_2 \boxtimes_L Q$ .

**Proof** By Proposition 12, we have to prove that if  $q(P'_1 \bowtie_L Q')$ ,  $q(P'_2 \bowtie_L Q') > 0$  with  $P'_1 \approx_p P'_2$ , then for all  $\alpha \in \mathcal{A}$

$$\frac{r_\alpha(P'_1 \bowtie_L Q')}{q(P'_1 \bowtie_L Q')} = \frac{r_\alpha(P'_2 \bowtie_L Q')}{q(P'_2 \bowtie_L Q')}.$$

We have that either  $\alpha \notin L$  or  $\alpha \in L$ . However, in both cases:

$$\frac{r_\alpha(P'_1 \bowtie_L Q')}{q(P'_1 \bowtie_L Q')} = \frac{r_\alpha(P'_1)}{q(P'_1)} = \frac{r_\alpha(P'_2)}{q(P'_2)} = \frac{r_\alpha(P'_2 \bowtie_L Q')}{q(P'_2 \bowtie_L Q')}.$$

□

The previous results show that, in general, we must restrict the conditions on cooperation between processes in order to preserve proportional bisimilarity. Indeed, Corollary 3 states that proportional bisimilarity is preserved if the transition rate of the cooperation is determined by process  $P$  and not by environment  $Q$ .

**Example 6** Consider a system consisting of the following components: a set of processes  $Comp_i$  for  $i = 1, \dots, N$  and a shared remote server  $Server$ . Processes  $Comp_1, \dots, Comp_N$  are functionally identical and can perform the computations either locally or may (partially) delegate the remote server  $Server$  to perform the computations.

The process algebraic description of this system is the following:

$$\begin{aligned} Comp_i &\stackrel{\text{def}}{=} (dispatch, k_i \lambda).(compute, w_i \mu).Comp_i + (task, k_i \eta).Comp_i \\ Server &\stackrel{\text{def}}{=} (dispatch, \top).Server \\ System &\stackrel{\text{def}}{=} (Comp_1 \parallel \dots \parallel Comp_N) \underset{\{dispatch\}}{\bowtie} Server \end{aligned}$$

where  $\parallel$  denotes  $\overset{\emptyset}{\bowtie}$ , i.e. the parallel composition without any synchronization. The partial delegation of  $Comp_i$  is modeled by  $(dispatch, k_i \lambda).(compute, w_i \mu)$  where  $k_i$  and  $w_i$  are inversely proportional with proportionality constant  $h$ . Notice that part of the work is done by  $Comp_i$  internally ( $(compute, w_i \mu)$ ) and part is delegated. If  $k_i$  is high, then the internal computation modeled by type  $task$  requires less expected time than the sequence  $dispatch, compute$  that models the remote computation.

Consider for simplicity the system:

$$\begin{aligned} Comp_1 &\stackrel{\text{def}}{=} (dispatch, k_1 \lambda).(compute, w_1 \mu).Comp_1 + (task, k_1 \eta).Comp_1 \\ Comp_2 &\stackrel{\text{def}}{=} (dispatch, k_2 \lambda).(compute, w_2 \mu).Comp_2 + (task, k_2 \eta).Comp_2 \\ Server &\stackrel{\text{def}}{=} (dispatch, \top).Server \\ System_{12} &\stackrel{\text{def}}{=} (Comp_1 \parallel Comp_2) \underset{\{dispatch\}}{\bowtie} Server_{12} \end{aligned}$$

where  $k_1 w_1 = k_2 w_2 = h$ .

Consider the component  $(Comp_1 \parallel Comp_2)$  whose derivation graph is depicted in Figure 7 where:

$$\begin{aligned} S_0 &\stackrel{\text{def}}{=} Comp_1 \parallel Comp_2 \\ S_1 &\stackrel{\text{def}}{=} (compute, w_1 \mu).Comp_1 \parallel Comp_2 \\ S_2 &\stackrel{\text{def}}{=} Comp_1 \parallel (compute, w_2 \mu).Comp_2 \\ S_3 &\stackrel{\text{def}}{=} (compute, w_1 \mu).Comp_1 \parallel (compute, w_2 \mu).Comp_2 \end{aligned}$$

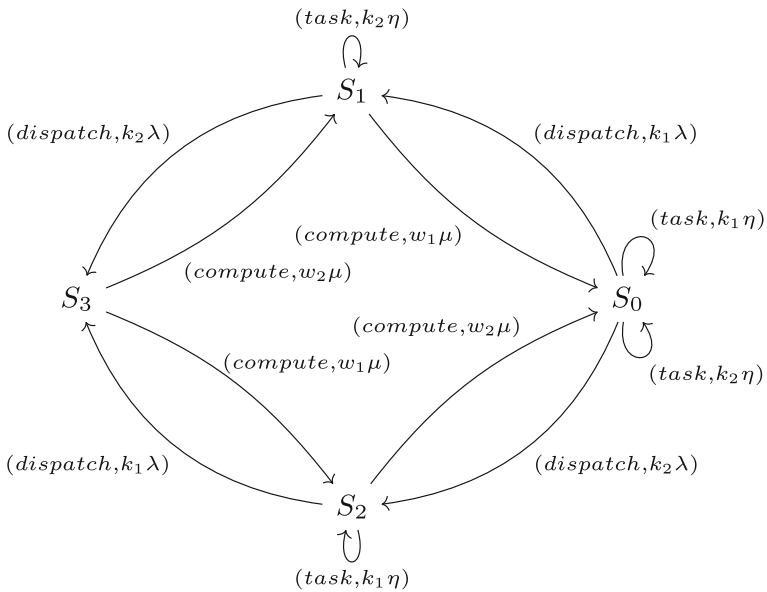


Fig. 7 Derivation graph of  $(Comp_1 || Comp_2)$

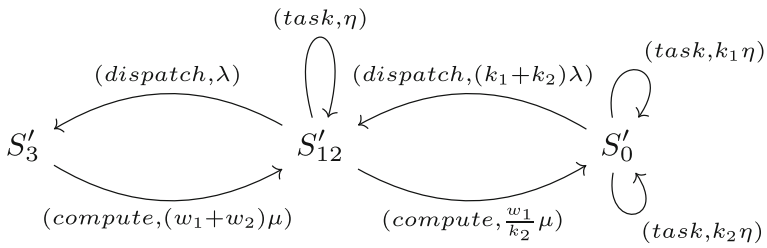


Fig. 8 Derivation graph of  $S'_0$

If we consider the function  $\kappa$  from PEPA components to  $\mathbb{R}^+$  such that  $\kappa(S_1) = k_2$  and  $\kappa(S_2) = k_1$  then it is easy to prove that  $(Comp_1 || Comp_2) \approx_p S'_0$  depicted in Figure 8. This follows from the fact that  $k_1 w_1 = k_2 w_2 = h$  and then  $w_1/k_2 = w_2/k_1$ . Therefore, by Corollary 3, we obtain  $System_{12} \approx_p S'_0 \boxtimes_{dispatch} Server_{12}$ .

### 4 Comparison with other forms of lumpability

In this section, we compare the notion of proportional lumpability with others that share some of the ideas that we propose. Specifically, we first compare the idea of proportional lumpability with that of lumping the embedded Markov chain, then with weak lumpability as discussed in [22,24] and finally with some notions of lumpability defined for the solution

of systems of ordinary differential equations (ODEs). For all these cases, we conclude that proportional lumpability has substantial differences.

*Ordinary lumping of the embedded Markov chain.* One standard approach for computing the stationary probability distribution of an ergodic continuous-time Markov chain  $X(t)$  is by analyzing its embedded Markov chain  $X^E(t)$ . Strictly speaking, the embedded Markov chain is a regular discrete-time Markov chain (DTMC), sometimes referred to as its jump process. Given  $X(t)$  with state space  $\mathcal{S}$ , each element of the one-step transition probability matrix of the corresponding embedded Markov chain is denoted by  $p(s, s')$ , and represents the conditional probability of transitioning from state  $s$  into state  $s'$ . These conditional probabilities may be found by

$$p(s, s') = \frac{q(s, s')}{q(s)}, \quad \text{for } s \neq s'$$

and  $p(s, s) = 0$ . Assuming that  $X^E(t)$  is aperiodic, let  $\pi^*$  be its steady-state distribution. One may derive the distribution  $\pi$  of  $X(t)$  as follows: let  $W = \sum_{s \in \mathcal{S}} \pi^*(s)/q(s)$ , then

$$\pi(s) = \frac{\pi^*(s)}{Wq(s)}.$$

Notice that, in general,  $q_{\sim}(s)$  is different from  $q(s)$ , hence the fact that  $X(t)$  is proportionally lumpable does not imply that the corresponding embedded Markov chain  $X^E(t)$  is lumpable. We may clearly see this from Example 3. The probability of jumping from state  $B_i, i > 0$ , to state  $B_0$  is:

$$p(B_i, B_0) = \frac{\mu/i}{\lambda + \mu/i} = \frac{\mu}{\lambda i + \mu},$$

which depends on  $i$  and hence cannot be the same for all the states  $B_1, \dots, B_M$  as would be required by the ordinary lumping condition applied to the embedded process.

On the other hand, if  $X^E(t)$  is lumpable then  $X(t)$  is proportionally lumpable with respect to function  $\kappa$  from  $\mathcal{S}$  to  $\mathbb{R}^+$  such that  $\kappa(s) = q(s)$  for all  $s \in \mathcal{S}$ . In conclusion, we can say that if  $X(t)$  has an ordinarily lumpable embedded process, then it is also proportionally lumpable but the opposite does not hold.

*Weak lumpability.* This notion of lumpability has been firstly introduced in [22] and then extended to continuous-time Markov chains. Among the works in this field, we will use the necessary condition for weak lumpability of CTMCs as stated in [24]. While weak lumpability is defined for DTMCs and CTMCs that may be not necessarily ergodic, for our purpose we will focus only on ergodic chains, i.e. they admit a unique stationary distribution, the steady-state one.

Let  $\mathbf{v}(t)$  be the probability distribution of  $X(t)$  at time  $t \geq 0$  and let  $\mathbf{v}(0) = \mathbf{v}$ . In weak lumpability, we aim at finding  $\mathbf{v}$  and a partition of the state space  $\sim$  such that the stochastic process obtained by aggregating the states according to  $\sim$  is a CTMC. Clearly, if  $\sim$  is an ordinary lumping, any probability vector  $\mathbf{v}$  satisfies this property. In [22], the authors observe that if  $\mathbf{v}$  and  $\sim$  determine a weak lumping, it must hold that for all  $t \geq 0$ , the probabilities of the states in  $\mathbf{v}(t)$  within the same equivalence class must maintain the same proportion. Notice that, since in our case  $X(t)$  is ergodic, then:

$$\lim_{t \rightarrow \infty} \mathbf{v}(t) = \pi,$$

thus, also  $\pi$  must satisfy the same property of  $\mathbf{v}(t)$ . Let us consider the Markov chain with the following infinitesimal generator taken from [24]:

$$\mathbf{Q} = \begin{bmatrix} -18 & 6 & 12 \\ 0 & -20 & 20 \\ 21 & 3 & -24 \end{bmatrix}.$$

If we choose as probability vector  $\mathbf{v} = (1 - 3a, a, 2a)$  for some  $a$ , and  $S_1 = \{1\}$ ,  $S_2 = \{2, 3\}$ , it can be shown that for all  $t \geq 0$ ,  $\mathbf{v}(t)$  is such that the third component is twice the second. In particular, since  $X(t)$  is ergodic, its steady-state distribution is:

$$\pi = \left( \frac{7}{16}, \frac{3}{16}, \frac{3}{8} \right),$$

and  $3/8 = 2 \cdot 3/16$ . Notice that  $\mathbf{Q}$  cannot be proportionally lumpable for any choice of  $\kappa$  because while state 3 has a positive rate to state 1, state 2 does not. Therefore, weak lumpability does not imply proportional lumpability. To verify if the opposite is true, we consider another example with the following infinitesimal generator:

$$\mathbf{Q} = \begin{bmatrix} -4 & 3 & 1 \\ 4 & -6 & 2 \\ 0 & 5 & -5 \end{bmatrix}$$

We notice that the CTMC is ergodic, and proportionally lumpable with respect to the partition  $S_1 = \{1, 2\}$  and  $S_2 = \{3\}$ , for example we can choose  $\kappa(2) = 2$  and  $\kappa(1) = \kappa(3) = 1$ . Let us prove that this CTMC is not weakly lumpable with respect to this partition. First, we compute the steady-state distribution so that we can find the proportions of the probabilities within the equivalence classes, we have:

$$\pi = \left( \frac{5}{13}, \frac{5}{13}, \frac{3}{13} \right),$$

thus a necessary condition for the chain to be weakly lumpable is that the first two states have the same probabilities in  $\mathbf{v}(t)$  for all  $t$  starting for a certain  $\mathbf{v}$ . We now prove that such a  $\mathbf{v}$  cannot exist. Assume, by contradiction, that  $X(t)$  is weakly lumpable, then the infinitesimal generator of the lumped process is:

$$\tilde{\mathbf{Q}} = \begin{bmatrix} -3/2 & 3/2 \\ 5 & -5 \end{bmatrix},$$

where  $3/2$  is obtained by averaging the rates from class  $S_1$  to class  $S_2$  according to the weights  $1/2$  and  $1/2$  obtained by the reasoning on the steady-state distribution. We now use Corollary 4.2 of [24] that essentially states that, in the aggregated process, the residence time in class  $S_2$  is exponentially distributed if one of the diagonal elements of  $\tilde{\mathbf{Q}}$  is an eigenvalue of the partial block of  $\mathbf{Q}$  associated with that aggregation, in our case:

$$\mathbf{Q}_{S_1} = \begin{bmatrix} -4 & 3 \\ 4 & -6 \end{bmatrix}.$$

The eigenvalues are  $-5 - \sqrt{13}$  and  $-5 + \sqrt{13}$ , neither of which are present in the diagonal of  $\tilde{\mathbf{Q}}$ . In conclusion, we have shown an example of a CTMC that is proportionally lumpable but not weakly lumpable.

It is important to notice that a weakly lumpable CTMC enjoys important properties on the transient phase of the aggregated process, in particular by stating that this process is a CTMC.

In contrast, the proportional lumpability does not guarantee that the aggregated process is still a Markov chain. Instead it states that the Markov chain defined on the aggregated state space according to Definition 3 has a steady-state distribution that is related with that of the original process according to Proposition 3.

*Lumpability on Differential Equations* In [23], systems of linear differential equations obtained by monomolecular reactions are considered. The authors aim to obtain a system of differential equations with fewer variables while preserving some properties of the original one. Intuitively this corresponds to replacing the species occurring in the reactions with fictitious species obtained as combinations of the original ones. As observed by the authors when the fictitious species are obtained considering a partition of the original ones they obtain a lumpability in the sense of Kemeny and Snell. On the other hand, semiproper and improper lumping are obtained when each species is not assigned to a unique class and fictitious species are generic linear combinations of the original ones. In [23], the analysis is extended to near lumpability. A first attempt to extend these works to the case of nonlinear systems of differential equations can be found [25]. In particular, the authors consider systems of the form

$$dy/dt = f(y)$$

where  $y$  is a vector of  $n$  variables. If  $\hat{y}$  is a vector of  $\hat{n} < n$  variables such that

$$\hat{y} = My$$

for an oportune matrix  $M$  of dimension  $\hat{n} \times n$ , then one could try to study the evolution of  $\hat{y}$  defined as

$$d\hat{y}/dt = \hat{f}(\hat{y}) = Mf(y)$$

and infer properties on the evolution of  $y$ . When  $M$  is such that there exists  $\bar{M}$  of dimension  $n \times \hat{n}$  such that  $\bar{M}M = Id$ , we have that

$$y = \bar{M}\hat{y}$$

As proved in [25], a necessary and sufficient condition for  $\hat{f}$  to be properly defined is that

$$Mf(y) = Mf(\bar{M}My)$$

This is a generalization of the necessary and sufficient condition of exact lumpability as presented in [22], where the infinitesimal generator  $Q$  has been replaced by the function  $f$  and the matrix  $U$  representing a probability distribution over each block of the lumpability has been replaced by  $M$  and the matrix  $V$  assigning each node to a block is  $\bar{M}$ . In other terms if we refer to CTMCs the necessary and sufficient condition of [25] is

$$UQ = UQVU$$

where  $U$  is a  $\hat{n} \times n$  matrix,  $V$  is a  $n \times \hat{n}$  matrix, and  $UV = Id$ . This generalizes the notion of exact lumpability since in  $U$  and  $V$  nodes can belong to more than one block, i.e. semiproper and improper lumpings are admissible. In [35], properties of original and lumped systems are analysed, e.g. properties of equilibria, invariant sets and periodic solutions. However, the considered definition of lumping is always the one given in [25] that generalizes exact lumpability as explained above.

For these reasons there is no relationship with our notion of proportional lumpability. Nevertheless, similarly to what has been as done in [25] for differential equations, we can try to express proportional lumpability in term of matrix products. It is easy to prove that an



equivalence relation  $\sim$  is a proportional lumpability for a Markov chain having infinitesimal generator  $\mathbf{Q}$  if and only if there exists a diagonal matrix  $D$  such that

$$V_{\sim} * U_{\sim} * D * \mathbf{Q} * V_{\sim} = D * \mathbf{Q} * V_{\sim}$$

where  $V_{\sim}$  is a Boolean matrix whose columns represent the equivalence classes of  $\sim$  and  $U_{\sim}$  is a matrix whose rows are probability distributions over the classes of  $\sim$  as in [22]. In other terms, lumpability requires that  $\mathbf{Q} * V_{\sim}$  is a fixpoint for the left operator  $V_{\sim} * U_{\sim}$ , while proportional lumpability requires that there exists a diagonal matrix  $D$  such that  $D * \mathbf{Q} * V_{\sim}$  is a fixpoint for the left operator  $V_{\sim} * U_{\sim}$ . Moreover, Proposition 4 ensures that instead of checking the condition over any possible diagonal matrix  $D$ , we can refer to the diagonal matrix  $D_{\sim}$  in which  $D_{\sim}[i, i]$  is the rate from  $i$  to the states that are not equivalent to  $i$ .

Finally, recent results on forward and backward differential equivalences have been proved in [14]. The authors show their equivalences coincide with ordinary (resp., exact) lumpability in the case of differential equations obtained from CTMCs.

## 5 Conclusion

In this paper, we have introduced a novel notion of quasi-lumpability, named *proportional lumpability*, which extends the original definition of lumpability but, differently from the general definition of quasi-lumpability, it allows one to derive exact performance indices at steady-state for the original process. As far as concerns proportional lumpability defined on Markov chains, we have given a characterization that avoids the universal quantification over all possible functions  $\kappa$ . As a consequence, given an equivalence relation we can check in linear time whether it is or not a proportional lumpability. As for the problem of finding the coarsest proportional lumpability that refines a given equivalence relation, even exploiting such characterization, it is not possible to immediately generalize the standard coarsest stable partitioning algorithms available for lumpability. Function  $\kappa$  would change during the computation and the stability conditions that guarantee the correctness of such algorithms would not be preserved. More elaborate extensions of partitioning strategies are under investigation.

Moreover, we have illustrated the concept of proportional bisimilarity for PEPA components that induces a proportional lumpability on the underlying Markov chain. We have provided various characterizations of proportional bisimilarity to give more insights on its definition. In particular, differently from what happens in the context of Markov chains, the first characterization of proportional bisimilarity for PEPA models allowed us to obtain an efficient algorithm for computing the maximum proportional bisimilarity.

Finally, we have investigated the compositionality properties of proportional bisimulation and illustrated the applicability of our theory through examples.

**Acknowledgements** The authors would like to thank Prof. James Worrell for his useful suggestions regarding the characterization of proportional lumpability. Furthermore, we would like to thank the anonymous reviewers for their careful revisions and valuable comments on the paper. This work is partially supported by University of Udine PRID ENCASE and by GNCS INdAM.

**Funding** Open access funding provided by Università Ca' Foscari Venezia within the CRUI-CARE Agreement.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the

article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Abate, A., Brim, L., Češka, M., Kwiatkowska, M.: Adaptive aggregation of Markov chains: Quantitative analysis of chemical reaction networks. In Proceedings of computer aided verification (CAV'15), pp. 195–213. Springer International Publishing, (2015)
2. Alzetta, G., Marin, A., Piazza, C., Rossi, S.: Lumping-based equivalences in Markovian automata: Algorithms and applications to product-form analyses. *Inf. Comput.* **260**, 99–125 (2018)
3. Baair, S., Beccuti, M., Dutheillet, C., Franceschinis, G.: From partially to fully lumped Markov chains in stochastic well formed Petri nets. In Proceedings of Valuetools 2009 conference, pp. 44. ACM, (2009)
4. Baair, S., Beccuti, M., Dutheillet, C., Franceschinis, G., Haddad, S.: Lumping partially symmetrical stochastic models. *Perform. Eval.* **68**(1), 21–44 (2011)
5. Baair, S., Dutheillet, C., Haddad, S., Iliè, J.-M.: On the use of exact lumping in partially symmetrical Well-formed Petri Nets. In Proceedings of International Conference on the Quantitative Evaluation of Systems (QEST'05), pp. 23–32, Torino, Italy. IEEE Comp. Soc (2005)
6. Baier, C., Katoen, J.-P., Hermanns, H., Wolf, V.: Comparative branching-time semantics for Markov chains. *Inf. Comput.* **200**(2), 149–214 (2005)
7. Balsamo, S., Marin, A.: Queueing Networks in Formal methods for performance evaluation, chapter 2, pp. 34–82. M. Bernardo and J. Hillston (Eds), LNCS, Springer, (2007)
8. Balsamo, Simonetta, Marin, Andrea: Product-form solutions for models with joint-state dependent transition rates. In Analytical and Stochastic Modeling Techniques and Applications, 17th International Conference, ASMTA 2010, volume 6148 of Lecture Notes in Computer Science, pp. 87–101, (2010)
9. Bernardo, M.: Weak Markovian bisimulation congruences and exact CTMC-level aggregations for concurrent processes. In Proceedings of the 10th Workshop on Quantitative Aspects of Programming Languages and Systems (QALP'12), pp. 122–136. EPTCS, (2012)
10. Bernardo, M.: Weak Markovian bisimulation congruences and exact CTMC-level aggregations for sequential processes. In Proceedings of the 6th international conference on trustworthy global computing (TGC'11), volume 7173 of LNCS, pp. 89–103. Springer, (2012)
11. Bernardo, M.: On the tradeoff between compositionality and exactness in weak bisimilarity for integrated-time markovian process calculi. *Theor. Comput. Sci.* **563**, 99–143 (2015)
12. Bravetti, M.: Revisiting interactive Markov chains. *Electr. Notes Theor. Comput. Sci.* **68**(5), 65–84 (2003)
13. Buchholz, P.: Exact and ordinary lumpability in finite Markov chains. *J. Appl. Probab.* **31**, 59–75 (1994)
14. Cardelli, L., Tribastone, M., Tschairowski, M., Vandin, A.: Symbolic computation of differential equivalences. *Theoret. Comput. Sci.* **777**, 132–154 (2019)
15. Coleman, J.L., Henderson, W., Taylor, P.G.: Product form equilibrium distributions and a convolution algorithm for Stochastic Petri nets. *Perf. Eval.* **26**(3), 159–180 (1996)
16. Daly, D., Buchholz, P., Sanders, W.H.: Bound-preserving composition for markov reward models. In Third International Conference on the Quantitative Evaluation of Systems (QEST 2006), pp. 243–252, (2006)
17. Derisavi, S., Hermanns, H., Sanders, W.H.: Optimal state-space lumping in Markov chains. *Elsevier Inform. Process. Lett.* **87**(6), 309–315 (2003)
18. Franceschinis, G., Muntz, R.: Bounds for quasi-lumpable Markov chains. *Perform. Eval.* **20**(1–3), 223–243 (1994)
19. Franceschinis, G., Muntz, R.: Computing bounds for the performance indices of quasi-lumpable stochastic well-formed nets. *IEEE Trans. Softw. Eng.* **20**(7), 516–525 (1994)
20. Hermanns, H.: *Interactive Markov Chains*. Springer, New York (2002)
21. Hillston, J.: *A Compositional Approach to Performance Modelling*. Cambridge Press, (1996)
22. Kemeny, J.G., Snell, J.L.: *Finite Markov Chains*. Springer, New York (1976)
23. Kuo, J.C.W., Wei, J.: Lumping analysis in monomolecular reaction systems. analysis of approximately lumpable system. *Ind. Eng. Chem. Fund.* **8**(1), 124–133 (1969)
24. Ledoux, J.: A necessary condition for weak lumpability in finite Markov processes. *Oper. Res. Lett.* **13**(3), 165–168 (1993)
25. Li, G., Rabitz, H.: A general analysis of exact lumping in chemical kinetics. *Chem. Eng. Sci.* **44**(6), 1413–1430 (1989)

26. Marin, A., Piazza, C., Rossi, S.: Proportional lumpability. In Formal modeling and analysis of timed systems—17th international conference, formats 2019, proceedings, volume 11750 of lecture notes in computer science, pp. 265–281. Springer, (2019)
27. Marin, A., Rossi, S.: Autoreversibility: exploiting symmetries in Markov chains. In Proceedings of IEEE MASCOTS, pages 151–160, san Francisco, CA, USA (2013)
28. Marin, A., Rossi, S.: On the relations between Markov chain lumpability and reversibility. *Acta Inform.* **54**(5), 447–485 (2017)
29. Milios, D., Gilmore, S.: Component aggregation for PEPA models: An approach based on approximate strong equivalence. *Perform. Eval.* **94**, 43–71 (2015)
30. Molloy, M.K.: Performance analysis using stochastic petri nets. *IEEE Trans. Comput.* **31**(9), 913–917 (1982)
31. Plateau, B.: On the stochastic structure of parallelism and synchronization models for distributed algorithms. *SIGMETRICS Perf. Eval. Rev.* **13**(2), 147–154 (1985)
32. Smith, M.J.A.: Compositional abstractions for long-run properties of stochastic systems. In: Eighth international conference on quantitative evaluation of systems, QEST **2011**, 223–232 (2011)
33. Sumita, U., Rieders, M.: Lumpability and time-reversibility in the aggregation-disaggregation method for large Markov chains. *Commun. Stat. Stoch. Models* **5**, 63–81 (1989)
34. Thomas, Nigel, Harrison, Peter G.: Semi-product-form solution for PEPA models with functional rates. In Analytical and Stochastic Modelling Techniques and Applications - 20th International Conference, ASMTA 2013, volume 7984 of Lecture Notes in Computer Science, pages 416–430, (2013)
35. Tomlin, A.S., Li, G., Rabitz, H., Tóth, J.: The effect of lumping and expanding on kinetic differential equations. *SIAM J. Appl. Math.* **57**(6), 1531–1556 (1997)
36. Towsley, D.F.: Queuing network models with state-dependent routing. *J. ACM* **27**(2), 323–337 (1980)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.