Acta Informatica (2018) 55:461–488
https://doi.org/10.1007/s00236-018-0313-1

ORIGINAL ARTICLE

Probabilistic bisimulation for realistic schedulers

Lijun Zhang[1,2] · Pengfei Yang[1,2] · Lei Song[1] · Holger Hermanns[3] · Christian Eisentraut[3] · David N. Jansen[1] · Jens Chr. Godskesen[4]

Received: 23 May 2016 / Accepted: 10 January 2018 / Published online: 24 February 2018

Abstract Weak distribution bisimilarity is an equivalence notion on probabilistic automata, originally proposed for Markov automata. It has gained some popularity as the coarsest behavioral equivalence enjoying valuable properties like preservation of trace distribution equivalence and compositionality. This holds in the classical context of arbitrary schedulers, but it has been argued that this class of schedulers is unrealistically powerful. This paper studies a strictly coarser notion of bisimilarity, which still enjoys these properties in the context of realistic subclasses of schedulers: Trace distribution equivalence is implied for partial information schedulers, and compositionality is preserved by distributed schedulers. The intersection of the two scheduler classes thus spans a coarser and still reasonable compositional theory of behavioral semantics.

1 Introduction

Compositional theories have been an important technique to deal with complex stochastic systems effectively. Their potential ranges from compositional minimization [4,7] approaches to component based verification [27,32]. Due to their expressiveness, Markov automata have attracted many attentions [17,25,39], since they were introduced [15]. Markov automata are a compositional behavioral model for continuous time stochastic and non-deterministic systems [14,15] subsuming interactive Markov chains [29] and probabilistic automata (PAs) [37] (and hence also Markov decision processes and Markov chains).

✉ Lijun Zhang
zhanglj@ios.ac.cn

1 State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, China

2 University of Chinese Academy of Sciences, Beijing, China

3 Saarland University, Saarbrücken, Germany

4 IT University of Copenhagen, Copenhagen, Denmark

On Markov automata, weak probabilistic bisimilarity has been introduced as a powerful way for abstracting from internal computation cascades, and this is obtained by relating subprobability distributions instead of states. In the sequel we call this relation *weak distribution bisimulation,* and focus on probabilistic automata, arguably the most widespread subclass of Markov automata.

On probabilistic automata, weak distribution bisimilarity is strictly coarser than weak bisimilarity, and is the coarsest congruence preserving trace distribution equivalence [9]. More precisely, it is the coarsest reduction-closed barbed congruence [31] with respect to parallel composition. Decision algorithms for weak distribution bisimilarity have also been proposed [18,35].

Weak distribution bisimilarity enables us to equate automata such as the ones on the left in Fig. 1, both of which exhibit the execution of action α followed by states r_1 and r_2 with probability $\frac{1}{2}$ each for an external observer. Specifically, the internal transition of the automaton on the left remains fully transparent. Standard bisimulation notions fail to equate these automata. On the other hand, the automata on the right are not bisimilar even though the situation seems to be identical for an external observer.

The automata on the right of Fig. 1 are to be distinguished, because otherwise compositionality with respect to parallel composition would be broken. However, as observed in [23,37], the general scheduler in the parallel composition is too powerful: the decision of one component may depend on the history of other components; in Fig. 1, whether s_4 or s_5 is visited may influence a scheduler decision regarding some other component. This is especially not desired for partially observable systems, such as multi-agent systems or distributed systems [3,38]. In distributed systems, where components only share the information they gain through explicit communication via observable actions, this behavior is unrealistic. Thus, for practically relevant models, weak distribution bisimilarity is still too fine. The need to distinguish the two automata on the right of Fig. 1 is in fact an unrealistic artifact, and this will motivate our definition of a coarser bisimulation, under which they are equivalent.

In this paper, we present a novel notion of weak bisimilarity on PAs, called *late distribution bisimilarity,* which is coarser than the existing notions of weak bisimilarity. It equates, for instance, all automata in Fig. 1. As weak distribution bisimilarity is the coarsest notion of equivalence that preserves observable behavior and is closed under parallel composition [9], late distribution bisimilarity cannot satisfy these properties in their entirety. However, as we will show, for a natural class of schedulers, late distribution bisimilarity preserves observable behavior, in the sense that trace distribution equivalence (i) is implied by

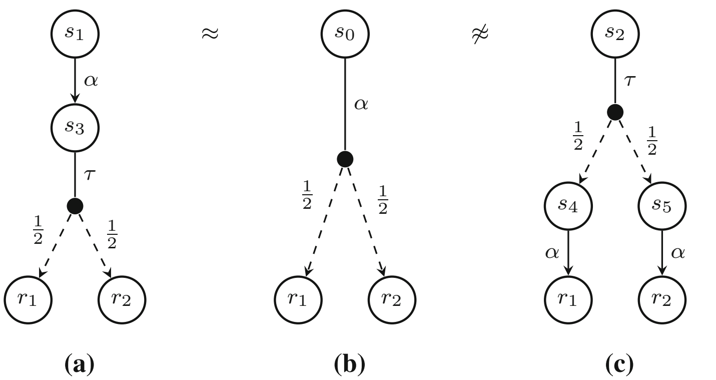

Fig. 1 Distinguishing probabilistic automata

late distribution bisimilarity, and (ii) is preserved in the context of parallel composition. This for instance implies that time-bounded reachability properties are preserved with respect to parallel composition. The class of schedulers under which these properties are satisfied is the intersection of two well-known scheduler classes, namely partial information schedulers [8] and distributed schedulers [23]. Both classes have been coined as principal means to exclude undesired or unrealistically powerful schedulers. We provide a co-inductive definition for late distribution bisimilarity which echoes these considerations on the automaton level, thereby resulting in a very coarse, yet reasonable, notion of equality.

Related work Many variants of bisimulations have been studied for different stochastic models, for instance Markov chains [1], interactive Markov chains [29], probabilistic automata [2,33,37], and alternating automata [12]. These equivalence relations are state-based, as they relate states of the corresponding models. Depending on how internal actions are handled, bisimulation relations can usually be categorized into strong bisimulations and weak bisimulations. The later is our main focus in this paper.

Markov automata arise as a combination of probabilistic automata and interactive Markov chains. In [15], a novel *distribution-based* weak bisimulation has been proposed: it is weaker than the state-based weak bisimulation in [37], and if restricted to continuous-time Markov chains, generates an equivalence established in the Petri net community [17]. Later, another weak bisimulation has been investigated in [9], which is essentially the same as [15]. In this paper, we propose a weaker bisimulation relation – late distribution bisimulation, which is coarser than both of them.

Interestingly, after the *distribution-based* weak bisimulations being introduced in [15], several *distribution-based* strong bisimulations have been proposed. In [28], it is shown that, the strong version of the relation in [15] coincides with the lifting of the classical state-based strong bisimulations. Recently, three different distribution-based strong bisimulations have been defined: paper [21] defines bisimulation relations and metrics which extend the well-known language equivalence [13] of labelled Markov chains; another definition in [30] applies to discrete systems as well as to systems with uncountable state and action spaces. The latter have been investigated in more detail in [41]. In [38], for multi-agent systems, a decentralized strong bisimulation relation is proposed which is shown to be compositional with respect to partial information and distributed schedulers. All these relations enjoy some interesting properties, and they are incomparable to each other: we refer to [38] for a detailed discussion. The current paper extends the decentralized strong bisimulation in [38] to the weak case. The extension is not trivial, as internal transitions need to be handled carefully, particularly when lifting transition relations to distributions. We show that our novel weak bisimulation is weaker than that in [15], and as in [38], we show that it is compositional with respect to partial information and distributed schedulers.

Organization of the paper In the next section we illustrate our approach by a running example. Section 3 recalls some notations used in the paper. Late distribution bisimulation is proposed and discussed in Sect. 4, and its properties are established in Sect. 5 under realistic schedulers. Section 6 concludes the paper. A discussion why all results established in this paper directly carry over to Markov automata can be found in [19].

2 A running example

As discussed in the introduction, the automata on the right of Fig. 1 should be distinguished. We illustrate this with the following example, inspired by [23,37], which considers the

Tossing 1	Tossing 2

print("I am going to toss.");
 $r = rand()$;
 if $r \geq \frac{1}{2}$ **then**
 print("Heads is up.");
 else
 print("Tails is up.");
 end
	$r = rand()$;
	 if $r \geq \frac{1}{2}$ **then**
	 print("I am going to toss.");
	 print("Heads is up.");
	 else
	 print("I am going to toss");
	 print("Tails is up.");
	 end

Fig. 2 Two algorithms simulating a coin toss

automata of Fig. 1 in the intuitive context of a guessing game. The discussion will reveal that the requirement to distinguish them is in fact an unrealistic artifact, and this will motivate our definition of a coarser bisimulation, under which they are equivalent.

Example 1 Figure 2 shows two different algorithms that simulate a coin toss by means of a random number generator. We assume that only the print statement is observable by the environment, while all other statements are internal. In both algorithms, first the initialization message "I am going to toss." will appear on the screen, and then the result of the coin toss, which is either "Heads is up." or "Tails is up."

In algorithm "Tossing 1", the initialization message is printed before the result of the coin throw is determined by a random number r drawn uniformly from (0, 1). Then, with probability $\frac{1}{2}$, "Heads is up" is printed and otherwise "Tails is up." In algorithm "Tossing 2" first r is determined and only afterwards, the initialization message and the result of the coin throw are printed. Intuitively, these two algorithms should not be distinguishable from the outside, as the same messages are printed with the same probability.[1]

Figure 3a, b, respectively, show the algorithms modeled as PAs, where i denotes the printing of the initialization message, while h and t denote the result messages "Heads is up." and "Tails is up.", respectively. Internal computations are modeled by the internal action τ. In Fig. 3c a guesser is modeled. While the tossing is announced (action i), he non-deterministically guesses the outcome, which he announces with the action h or t.

The complete system is obtained by a parallel composition of the coin tosser automaton and the guesser automaton. We use a *CSP*-style parallel composition. Throughout our example, synchronization is enforced for the actions in the set $A = \{i, h, t\}$. These actions synchronize with corresponding actions of the coin tosser. Thus, if the guess was right, the guesser finally performs the action *Suc* to announce that he successfully guessed the outcome. □

In the example, the probability to see head or tail after a (fake) coin toss is one half each, both for tosser (a) and (b). One would expect that hence the chance to guess correct is one half for both tossers. However, $s_0 \parallel_A r_0$ and $s'_0 \parallel_A r_0$ are not weakly bisimilar. We will now show that the executions that distinguish the two systems are actually caused by unrealistic schedulers, which cannot appear in real world applications. Suppose we have a scheduler of $s'_0 \parallel_A r_0$, which chooses the left transition of r_0 when at $s_5 \parallel_A r_0$ and the right one when at $s_6 \parallel_A r_0$, then almost surely $\xrightarrow{Suc}$ will be seen eventually. In contrast, the probability that $\xrightarrow{Suc}$ is executed in $s_0 \parallel_A r_0$ is at most 0.5, for every scheduler.

The intuitive reason why the scheduler for $s'_0 \parallel_A r_0$ is too powerful to be realistic is that it can base its decision which transition to choose in state r_0 on the state the tosser has

[1] We assume that timing differences in the execution are not observable.

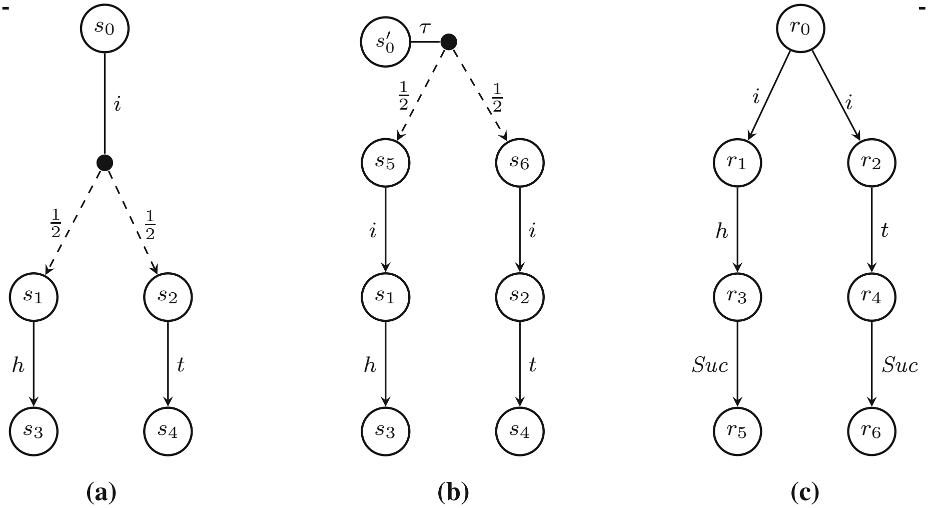

Fig. 3 s_0 and s'_0 represent different ways of tossing a coin and r_0 denotes the guesser

reached by performing his internal probabilistic decision, namely either state s_5 or s_6. If we consider the tosser and the guesser to be independently running processes, this is not a realistic scheduler, as then the guesser would need to see the internal state of the tosser. However, no communication between guesser and tosser has happened at this point in time, by which this information could have been conveyed. Thus, in distributed systems, where components only share the information they gain through explicit communication via observable actions, this behavior is unrealistic. Thus, for practically relevant models, weak distribution bisimilarity is still too fine.

Therefore, we present a novel notion of weak bisimilarity on PAs, called *late distribution bisimilarity,* that is coarser than the existing notions of weak bisimilarity. It equates, for instance, the two automata of Example 1, and the three in Fig. 1. As weak distribution bisimilarity is the coarsest notion of equivalence that preserves observable behavior and is closed under parallel composition [9], late distribution bisimilarity cannot satisfy these properties in their entirety. However, as we will show, for a natural class of schedulers, late distribution bisimilarity preserves observable behavior, in the sense that trace distribution equivalence (i) is implied by late distribution bisimilarity, and (ii) is preserved in the context of parallel composition. This for instance implies that time-bounded reachability properties are preserved with respect to parallel composition. The class of schedulers under which these properties are satisfied is the intersection of two well-known scheduler classes, namely partial information schedulers [8] and distributed schedulers [23]. Both these classes have been coined as principal means to exclude undesired or unrealistically powerful schedulers. We provide a co-inductive definition for late distribution bisimilarity which echoes these considerations on the automaton level, thereby resulting in a very coarse, yet reasonable, notion of equality.

3 Preliminaries

Let S be a finite set of states ranged over by $r, s, \ldots$ A *distribution* is a function $\mu : S \to [0, 1]$ satisfying $\mu(S) = \sum_{s \in S} \mu(s) = 1$. Let $Dist(S)$ denote the set of all distributions. If $\mu(s) = 1$

for some $s \in S$, then μ is called a *Dirac* distribution, written as δ_s. Similarly, a *subdistribution* is a function $\mu : S \to [0, 1]$ satisfying $\mu(S) = \sum_{s \in S} \mu(s) \leq 1$. Let $SubDist(S)$ denote the set of all subdistributions, ranged over by $\mu, \nu, \gamma, \ldots$. Define $Supp(\mu) = \{s \mid \mu(s) > 0\}$ as the support set of μ. Let $|\mu| = \mu(S)$ denote the size of the subdistribution μ. Given a real number x, $x \cdot \mu$ is the subdistribution such that $(x \cdot \mu)(s) = x \cdot \mu(s)$ for each $s \in Supp(\mu)$ if $x \cdot |\mu| \leq 1$, while $\mu - s$ is the subdistribution such that $(\mu - s)(s) = 0$ and $(\mu - s)(r) = \mu(r)$ for all $r \neq s$. Moreover, $\mu = \mu_1 + \mu_2$ whenever $\mu(s) = \mu_1(s) + \mu_2(s)$ for each $s \in S$ and $|\mu| \leq 1$. We often write $\{s : \mu(s) \mid s \in Supp(\mu)\}$ alternatively for a subdistribution μ. For instance, $\{s_1 : 0.4, s_2 : 0.6\}$ denotes a distribution μ with $\mu(s_1) = 0.4$ and $\mu(s_2) = 0.6$.

3.1 Probabilistic automata

Initially introduced in [37], *probabilistic automata* (PAs) have been popular models for systems with both non-deterministic choices and probabilistic dynamics. Below we give their formal definition.

Definition 2 A PA $\mathcal{P}$ is a tuple $(S, Act_\tau, \to, \bar{s})$ where

- S is a finite set of states,
- $Act_\tau = Act \,\dot\cup\, \{\tau\}$ is a finite set of actions including the internal or invisible action τ,
- $\to \subset S \times Act_\tau \times Dist(S)$ is a finite set of probabilistic transitions, and
- $\bar{s} \in S$ is the initial state.

In our paper, we assume that in a PA, every state has at least one transition. Let $\alpha, \beta, \gamma, \ldots$ range over the actions in Act_τ. We write $s \xrightarrow{\alpha} \mu$ if $(s, \alpha, \mu) \in \to$. A *path* is a finite or infinite strictly alternating sequence $\pi = s_0, \alpha_0, s_1, \alpha_1, s_2 \ldots$ of states and actions, such that for each $i \geq 0$ there exists a distribution μ with $s_i \xrightarrow{\alpha_i} \mu$ and $\mu(s_{i+1}) > 0$. Some notations are defined as follows: $|\pi|$ denotes the length of π, i.e., the number of states on π, while $\pi\downarrow$ is the last state of π, provided π is finite; $\pi[i] = s_i$ with $i \geq 0$ is the $(i+1)$-th state of π if it exists; $\pi[0..i] = s_0, \alpha_0, s_1, \alpha_1, \ldots, s_i$ is the prefix of π ending at state $\pi[i]$.

Let $Paths^\omega(\mathcal{P}) \subseteq S \times (Act_\tau \times S)^\omega$ and $Paths^*(\mathcal{P}) \subseteq S \times (Act_\tau \times S)^*$ denote the sets containing all infinite and finite paths of $\mathcal{P}$, respectively. Let $Paths(\mathcal{P}) = Paths^\omega(\mathcal{P}) \cup Paths^*(\mathcal{P})$. We will omit $\mathcal{P}$ if it is clear from the context. We also let $Paths(s)$ be the set containing all paths starting from $s \in S$, similarly for $Paths^*(s)$ and $Paths^\omega(s)$.

Due to the non-deterministic choices in PAs, a probability measure over $Paths(\mathcal{P})$ cannot be defined directly. As usual, we shall introduce the definition of *schedulers* to resolve the non-determinism. Intuitively, a scheduler will decide which transition to choose at each step, based on the execution history. Formally,

Definition 3 A scheduler is a function

$$\xi : Paths^* \to Dist(Act_\tau \times Dist(S))$$

such that $\xi(\pi)(\alpha, \mu) > 0$ implies $\pi\downarrow \xrightarrow{\alpha} \mu$. A scheduler ξ is *deterministic* if it returns only Dirac distributions, that is, for each π there is a pair (α, μ) such that $\xi(\pi)(\alpha, \mu) = 1$. ξ is *memoryless* if $\pi\downarrow = \pi'\downarrow$ implies $\xi(\pi) = \xi(\pi')$ for any $\pi, \pi' \in Paths^*$, namely, the decision of ξ only depends on the last state of a path.

Let $\pi \leq \pi'$ iff π is a prefix of π'. Let C_π denote the *cone* of a finite path π, which is the set of infinite paths having π as their prefix, i.e.,

$$C_\pi = \{\pi' \in Paths^\omega \mid \pi \leq \pi'\}.$$

Given a starting state s, a scheduler ξ, and a finite path $\pi = s_0, \alpha_0, s_1, \alpha_1, \ldots, s_k$, the measure Pr^s_ξ of a cone C_π is defined inductively as:

- $Pr^s_\xi(C_\pi) = 0$ if $s \neq s_0$;
- $Pr^s_\xi(C_\pi) = 1$ if $s = s_0$ and $k = 0$;
- otherwise $Pr^s_\xi(C_\pi) =$

$$Pr^s_\xi(C_{\pi[0..k-1]}) \cdot \sum_{s_{k-1} \xrightarrow{\alpha_{k-1}} \mu} \xi(\pi[0..k-1])(\alpha_{k-1}, \mu) \cdot \mu(s_k).$$

Let $\mathcal{B}$ be the smallest σ-algebra that contains all the cones. By standard measure theory [26, 34], Pr^s_ξ can be extended to a unique measure on $\mathcal{B}$.

Large systems are usually built from small components. This is done by using the parallel operator of PAs [37].

Definition 4 (*Parallel Operator*) Let $\mathcal{P}_1 = (S_1, Act_\tau, \rightarrow_1, \bar{s}_1)$ and $\mathcal{P}_2 = (S_2, Act_\tau, \rightarrow_2, \bar{s}_2)$ be two PAs and $A \subseteq Act$. We define $\mathcal{P}_1 \parallel_A \mathcal{P}_2 = (S, Act_\tau, \rightarrow, \bar{s})$ where

- $S = \{s_1 \parallel_A s_2 \mid (s_1, s_2) \in S_1 \times S_2\}$,
- $s_1 \parallel_A s_2 \xrightarrow{\alpha} \mu_1 \parallel_A \mu_2$ iff
  - either $\alpha \in A$ and $\forall i \in \{1, 2\}. s_i \xrightarrow{\alpha}_i \mu_i$,
  - or $\alpha \notin A$ and $\exists i \in \{1, 2\}. (s_i \xrightarrow{\alpha}_i \mu_i$ and $\mu_{3-i} = \delta_{s_{3-i}})$.
- $\bar{s} = \bar{s}_1 \parallel_A \bar{s}_2$,

where $\mu_1 \parallel_A \mu_2$ is a distribution such that $(\mu_1 \parallel_A \mu_2)(s_1 \parallel_A s_2) = \mu_1(s_1) \cdot \mu_2(s_2)$.

Example 5 In Fig. 4, we build the parallel compositions $s_0 \parallel_A r_0$ and $s'_0 \parallel_A r_0$ of the automata of our running example in Fig. 3.

3.2 Trace distribution equivalence

In this subsection we introduce the notion of trace distribution equivalence [36] adapted to our setting with internal actions. Let $\varsigma \in Act^*$ denote a finite trace of a PA $\mathcal{P}$, which is an ordered sequence of visible actions. Each trace ς induces a cylinder C_ς which is defined as follows:

$$C_\varsigma = \bigcup \{C_\pi \mid \pi \in Paths^* \wedge trace(\pi) = \varsigma\}$$

where $trace(\pi) = \epsilon$ denotes an empty trace if $|\pi| \leq 1$, and

$$trace(\pi) = \begin{cases} trace(\pi'), & \text{if } \pi = \pi' \circ (\tau, s') \\ trace(\pi')\alpha, & \text{if } \pi = \pi' \circ (\alpha, s') \text{ and } \alpha \neq \tau \end{cases}.$$

Since C_ς is a countable set of cylinders, it is measurable. Below we define *trace distribution equivalences,* each of which is parametrized by a certain class of schedulers.

Definition 6 (*Trace Distribution Equivalence* $\equiv$) Let s_1 and s_2 be two states of a PA, and $\mathcal{S}$ a set of schedulers. Then, $s_1 \equiv_\mathcal{S} s_2$ iff for each scheduler $\xi_1 \in \mathcal{S}$ there exists a scheduler $\xi_2 \in \mathcal{S}$, such that $Pr^{s_1}_{\xi_1}(C_\varsigma) = Pr^{s_2}_{\xi_2}(C_\varsigma)$ for each finite trace ς and vice versa. If $\mathcal{S}$ is the set of all schedulers, we simply write $\equiv$.

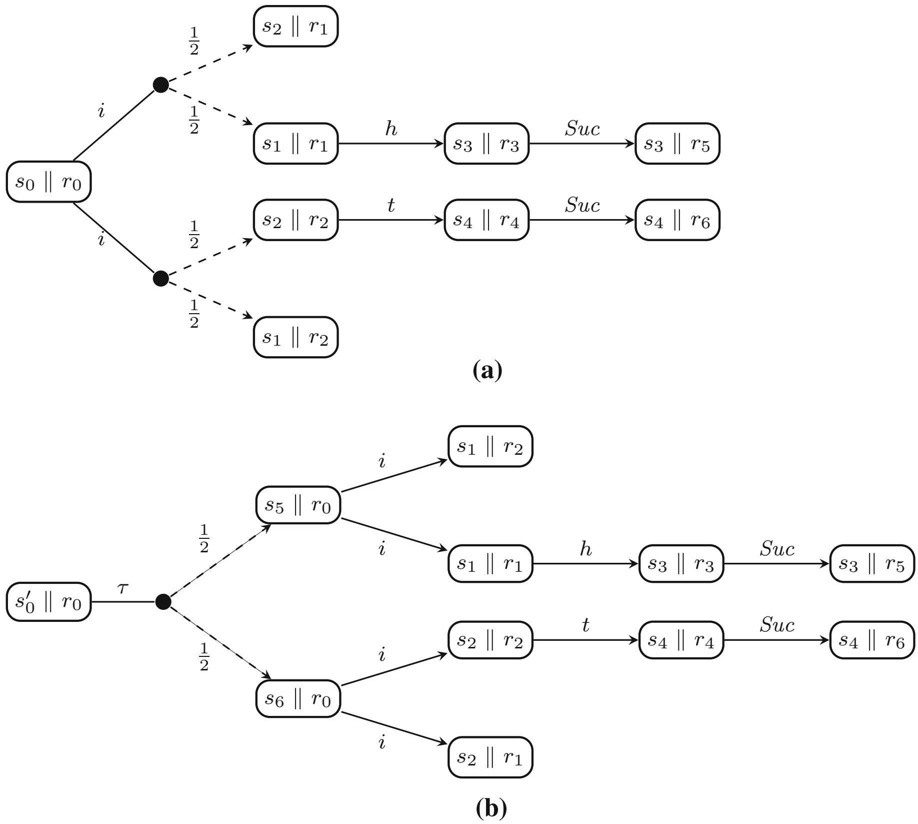

Fig. 4 Parallel compositions $s_0 \parallel_A r_0$ and $s_0' \parallel_A r_0$ (A is omitted from $\parallel_A$ in the pictures)

In contrast to [36,38], we abstract from internal transitions when defining traces of a path. Therefore, the definition above is also a weaker version of the corresponding definition in [36,38].

Below follow examples (and counterexamples) of trace distribution equivalent states:

Example 7 Let s_0 and s_0' be the two states in Fig. 3, then we have $s_0 \equiv s_0'$, since the only trace distribution of s_0 and s_0' is $\{ih : \frac{1}{2}, it : \frac{1}{2}\}$. In contrast, s_0 and s_1 in Fig. 5 are not trace distribution equivalent, since there are two possible trace distributions for s_0: $\{\beta : 1\}$ and $\{\alpha : 1\}$, but for s_1 there are four trace distributions: $\{\alpha : 1\}$, $\{\beta : 1\}$, $\{\alpha : \frac{1}{3}, \beta : \frac{2}{3}\}$, and $\{\beta : \frac{1}{3}, \alpha : \frac{2}{3}\}$. □

3.3 Partial information and distributed schedulers

In this subsection, we are refining the very liberal Definition 3, where the set of all schedulers was introduced. As discussed, this class can be considered too powerful, since it includes unrealistic schedulers. We define two prominent sub-classes of schedulers with limited power. We first introduce some notations. Let $EA : S \mapsto 2^{Act}$ such that

$$EA(s) = \{\alpha \in Act \mid \exists \mu \in Dist(S).s \stackrel{\alpha}{\Rightarrow} \mu\},$$

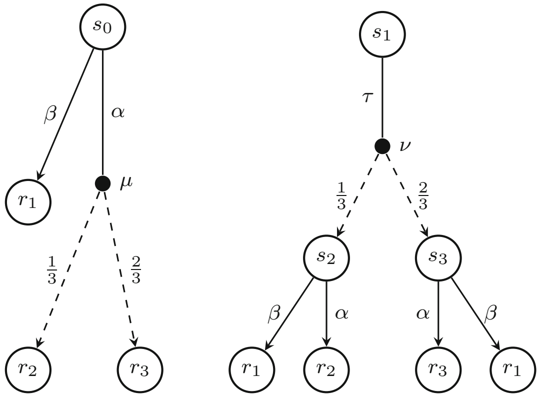

Fig. 5 $s_0 \not\approx^\bullet s_1$

that is, the function *EA* returns the set of visible actions that a state is able to perform, possibly after some internal transitions. In this definition, we use the weak transition $\stackrel{\alpha}{\Rightarrow}$; we will give its formal definition in Sect. 4. We generalize this function to paths as follows: $EA(\pi) =$

$$= \begin{cases} EA(s) & \pi = s \quad (1)\\ EA(\pi') & \pi = \pi' \circ (\tau, s) \wedge EA(\pi'\downarrow) = EA(s) \quad (2)\\ EA(\pi')\alpha EA(s) & \pi = \pi' \circ (\alpha, s) \wedge (\alpha \neq \tau \vee EA(\pi'\downarrow) \neq EA(s)) \quad (3)\end{cases}$$

where Case (2) takes care of a special situation such that internal actions do not change enabled actions. In this case *EA* will not see the difference. Intuitively, $EA(\pi)$ abstracts concrete states on π to their corresponding enabled actions. Whenever an invisible action does not change the enabled actions, it will simply be omitted. In other words, $EA(s)$ can be seen as the interface of s, which is observable by other components. Other components can observe the execution of s, as long as either it performs a visible action $\alpha \neq \tau$, or its interface has been changed ($EA(\pi'\downarrow) \neq EA(s)$). This is similar to the way branching bisimulation disregards τ transitions as far as they do not change the branching structure [24,40]. We are now ready to define the *partial information schedulers* [8] as follows:

Definition 8 (*Partial Information Schedulers*) A scheduler ξ is a partial information scheduler of s if for any $\pi_1, \pi_2 \in Paths^*(s)$, $EA(\pi_1) = EA(\pi_2)$ implies

- either $\xi(\pi_i, \tau) = 1$ for some $i \in \{1, 2\}$,
- or on π_1 and π_2, the scheduler ξ has the same conditional probability to do any visible action α given the condition that a visible action will be done, i.e. for any $\alpha, \beta \in Act$, $\xi(\pi_1, \alpha)\xi(\pi_2, \beta) = \xi(\pi_1, \beta)\xi(\pi_2, \alpha)$, where

$$\xi(\pi, \alpha) := \sum_{\pi\downarrow \stackrel{\alpha}{\to} \mu} \xi(\pi)(\alpha, \mu).$$

ξ is a partial information scheduler of a PA $\mathcal{P}$ iff it is a partial information scheduler for every state of $\mathcal{P}$.

We denote the set of all partial information schedulers by $\mathcal{S}_P$. Intuitively a partial information scheduler can only distinguish histories containing different enabled visible action sequences. A scheduler cannot choose different transitions of states only because they have different state identities. This fits very well to a behavior-oriented rather than state-oriented view, as it is typical for process calculi. Consequently, for two different paths π_1 and π_2 with

$EA(\pi_1) = EA(\pi_2)$, a partial information scheduler either chooses a transition labelled with τ action for π_i $(i = 1, 2)$, or it chooses transitions labelled with the same visible actions for both π_1 and π_2. Partial information schedulers do not impose any restriction on the execution of τ transitions, instead they can be performed independently.

When composing parallel systems, general schedulers defined in Definition 3 allow one component to make decisions based on full information of other components. Giro and D'Argenio [23] argues that this is unrealistically powerful and introduces another important sub-class of schedulers called *distributed schedulers*. The main idea is to assume that all parallel components run autonomously and make their local scheduling decisions in isolation. In other words, each component can use only that information about other components that has been conveyed to it explicitly. For instance the guesser in Fig. 3 cannot base its local scheduling decision on the tossing outcome at the moment when his guess is to be scheduled.

Below we recall the formal definition of distributed schedulers [23,38].[2] To formalize this locality idea, we first need to define the projection of a path to the path of its components. Let $s = ||_A\{s_i \mid 0 \le i \le n\}$ be a state which is composed from $n + 1$ processes in parallel such that all the processes synchronize on actions in A. Let π be a path starting from s, then the i-projection of π denoted by $[\pi]_i$ is defined as follows: $[\pi]_i = [s]_i$ if $\pi = s$, otherwise if $\pi = \pi' \circ (\alpha, s')$,

$$[\pi]_i = \begin{cases} [\pi']_i \circ (\alpha, [s']_i) & \alpha \in A \vee (\alpha \notin A \wedge [\pi'\downarrow]_i \xrightarrow{\alpha} [s']_i) \\ [\pi']_i & \alpha \notin A \wedge (\exists j \neq i.[\pi'\downarrow]_j \xrightarrow{\alpha} [s']_j) \end{cases}$$

where $[s]_i = s_i$ with $0 \le i \le n$. Intuitively, given a path π of a state s, the i-projection of π is the path that only keeps track of the execution of the i-th component of s during its execution. Also note any scheduler ξ of s can be decomposed into $n + 2$ functions: a global scheduler $\xi_g : Paths^* \times \{0, \ldots, n\} \mapsto \{0, 1\}$ and $n + 1$ local schedulers $\{\xi_i\}_{0 \le i \le n}$ such that for any π with $\pi\downarrow = ||_A\{s_i \mid 0 \le i \le n\}$, and

$$\xi(\pi)(\alpha, \big|\big|_A\{\mu_i\}_{0\le i\le n}) = \prod_{0\le i\le n} \left[\xi_g(\pi, i) \cdot \xi_i(\pi)(\alpha, \mu_i) + (1 - \xi_g(\pi, i)) \cdot Eq(\delta_{s_i}, \mu_i)\right],$$

where $Eq(\delta_{s_i}, \mu_i)$ returns 1 if $\delta_{s_i} = \mu_i$ and 0 otherwise. Intuitively, the global scheduler ξ_g chooses processes which will participate in the next transition, while ξ_i guides the execution of s_i in case the i-th process is chosen by ξ_g. In case the i-th process is not chosen by the global scheduler, it will not change its state. Below we define the distributed schedulers:

Definition 9 (*Distributed Schedulers*) A scheduler ξ is *distributed* for $s = ||_A\{s_i \mid 0 \le i \le n\}$ iff its corresponding global scheduler ξ_g and local schedulers $\{\xi_i\}_{0\le i\le n}$ satisfy: for any $\pi, \pi' \in Paths^*(s)$ and $0 \le i \le n$, $[\pi]_i = [\pi']_i$ implies $\xi_g(\pi, i) = \xi_g(\pi', i)$ and $\xi_i(\pi) = \xi_i(\pi')$. A distributed scheduler for a PA $\mathcal{P}$ is a scheduler distributed for all states in $\mathcal{P}$.

We denote the set of all distributed schedulers by $\mathcal{S}_D$. In case $n = 0$, distributed schedulers degenerate to ordinary schedulers defined in Definition 3. According to Definition 9, a scheduler ξ is distributed, if ξ cannot distinguish different paths starting from s, provided the projections of these paths to each of its parallel component coincide. Note that the lower PA in Fig. 4 allows a scheduler that guesses correctly with probability 1: the scheduler would choose the transitions $s_5 \parallel_A r_0 \xrightarrow{i} s_1 \parallel_A r_1$ and $s_6 \parallel_A r_0 \xrightarrow{i} s_2 \parallel_A r_2$, but this scheduler is not distributed, since the decision of r_0 depends on the execution history of s'_0, i.e., the

[2] In [38], they are called *decentralized schedulers* in the context of probabilistic multi-agent systems.

choice of $\xi_2(\pi)$ depends on whether $[\pi]_1 = s_5$ or s_6. By restricting to the set of distributed schedulers, we can avoid this unrealistic scheduler of $s_0' \parallel_A r_0$.

4 Weak bisimilarities for probabilistic automata

In this section, we first introduce weak distribution bisimulation, which is a variant of weak bisimulation defined in [9], and then define late distribution bisimulation, which is strictly coarser than weak distribution bisimulation.

4.1 Lifting of a transition relation

In the following, let $\mu \xrightarrow{\alpha} \mu'$ iff there exists a transition $s \xrightarrow{\alpha} \mu_s$ for each $s \in \mathit{Supp}(\mu)$ such that $\mu' = \sum_{s \in \mathit{Supp}(\mu)} \mu(s) \cdot \mu_s$. We generalize this as in [9] to the lifting of other relations:

Definition 10 Let S be a nonempty finite set and $\rightsquigarrow \subseteq S \times \mathit{SubDist}(S)$ be a (transition) relation. Then $\rightsquigarrow_{\mathrm{c}} \subseteq \mathit{SubDist}(S) \times \mathit{SubDist}(S)$ is the smallest relation that satisfies:

1. $s \rightsquigarrow \mu$ implies $\delta_s \rightsquigarrow_{\mathrm{c}} \mu$;
2. If $\mu_i \rightsquigarrow_{\mathrm{c}} \nu_i$ for $i = 1, 2, \ldots, n$, then $\sum_{i=1}^n p_i \mu_i \rightsquigarrow_{\mathrm{c}} \sum_{i=1}^n p_i \nu_i$ for any $p_i \in [0, 1]$ with $\sum_{i=1}^n p_i = 1$.

The lifting of a relation has many good properties [9]. We list those we need:

Lemma 11 *Let $\rightsquigarrow \subseteq S \times \mathit{SubDist}(S)$ be a relation. Then its lifting relation $\rightsquigarrow_{\mathrm{c}}$ has the following properties:*

1. *$\rightsquigarrow_{\mathrm{c}}$ is left-decomposable, i.e. $\sum_{i=0}^n p_i \mu_i \rightsquigarrow_{\mathrm{c}} \nu$ implies that ν can be written as $\nu = \sum_{i=0}^n p_i \nu_i$ such that $\mu_i \rightsquigarrow_{\mathrm{c}} \nu_i$ for every i, where $p_i \in [0, 1]$ with $\sum_{i=0}^n p_i = 1$.*
2. *$\mu \rightsquigarrow_{\mathrm{c}} \nu$ iff ν can be written as $\nu = \sum_{s \in \mathit{Supp}(\mu)} \mu(s) \nu_s$, where ν_s can be written as $\nu_s = \sum_{i=0}^n p_i \nu_{s,i}$ such that $s \rightsquigarrow \nu_{s,i}$ for each i, where $p_i \in [0, 1]$ with $\sum_{i=0}^n p_i = 1$.*
3. *$\rightsquigarrow_{\mathrm{c}}$ is σ-linear, i.e. $\mu_i \rightsquigarrow_{\mathrm{c}} \nu_i$ for every $i \geq 0$ implies that $\sum_{i \geq 0} p_i \mu_i \rightsquigarrow_{\mathrm{c}} \sum_{i \geq 0} p_i \nu_i$, where $p_i \in [0, 1]$ with $\sum_{i \geq 0} p_i = 1$.*

4.2 Weak distribution bisimulation

As usual, a standard weak transition relation is needed in the definitions of bisimulation that allows one to abstract from internal actions. Intuitively, $s \overset{\alpha}{\Rightarrow} \mu$ denotes that a distribution μ is reached from s by an α-transition, which may be preceded and followed by an arbitrary sequence of internal transitions. We define them as derivatives [10] for PAs. Formally $s \overset{\tau}{\Rightarrow} \mu$ iff there exists an infinite sequence

$$
\begin{aligned}
\delta_s &= \mu_0^{\rightarrow} + \mu_0^{\times}, \\
\mu_0^{\rightarrow} &\xrightarrow{\tau} \mu_1^{\rightarrow} + \mu_1^{\times}, \\
\mu_1^{\rightarrow} &\xrightarrow{\tau} \mu_2^{\rightarrow} + \mu_2^{\times}, \\
&\ldots
\end{aligned}
$$

where $\mu = \sum_{i \geq 0} \mu_i^{\times}$. We write $s \overset{\alpha}{\Rightarrow} \mu$ iff there exists $s \overset{\tau}{\Rightarrow} \xrightarrow{\alpha} \overset{\tau}{\Rightarrow} \mu$, and $\mu \overset{\alpha}{\Rightarrow} \mu'$ iff for $s \in \mathit{supp}(\mu)$ there exists $s \overset{\alpha}{\Rightarrow} \mu_s$, s.t. $\mu' = \sum_{s \in \mathit{supp}(\mu)} \mu(s) \mu_s$. It is worth noting that the relation $\overset{\tau}{\Rightarrow}$ is transitive (See [9, Thm. A.4]), i.e. $\mu \overset{\tau}{\Rightarrow} \mu'$ and $\mu' \overset{\tau}{\Rightarrow} \mu''$ imply $\mu \overset{\tau}{\Rightarrow} \mu''$, and it is easy to see that $\overset{\tau}{\Rightarrow}_{\mathrm{c}}$ is also transitive.

In [11,20], compactness and continuity are characterized. We say a sequence of subdistributions μ_i converges to μ, denoted by $\lim_{i\to\infty}\mu_i = \mu$, if for any $A \subseteq S$, $\lim_{i\to\infty}\mu_i(A) = \mu(A)$. The following lemma, derived from [20, Lemma 7.5], shows that finite systems are compact and the weak transition relation $\overset{\alpha}{\Rightarrow}_{\mathrm{c}}$ is continuous.

Lemma 12 1. *For any $s \in S$ and $\alpha \in Act_\tau$, the set $\{\mu \mid s \overset{\alpha}{\Rightarrow}_{\mathrm{c}} \mu\}$ is compact.*
2. *The relation $\overset{\alpha}{\Rightarrow}_{\mathrm{c}}$ is continuous, i.e. if $\mu_i \overset{\alpha}{\Rightarrow}_{\mathrm{c}} \nu_i$ with $\lim_{i\to\infty}\mu_i = \mu$ and $\lim_{i\to\infty}\nu_i = \nu$, then $\mu \overset{\alpha}{\Rightarrow}_{\mathrm{c}} \nu$.*

Note the weak transitions in [16,20] are defined via *trees,* whereas our definitions are based on derivatives. The underlying idea is similar, and it is shown in [5] that they agree with each other for systems with finite states. We note also that the proof in [20] can be adapted to a direct proof of the above lemma.

Definition 13 $\mathcal{R} \subseteq Dist(S) \times Dist(S)$ is a *weak distribution bisimulation* iff $\mu\ \mathcal{R}\ \nu$ implies:

1. whenever $\mu \overset{\alpha}{\rightarrow}_{\mathrm{c}} \mu'$, there exists a $\nu \overset{\alpha}{\Rightarrow}_{\mathrm{c}} \nu'$ such that $\mu'\ \mathcal{R}\ \nu'$;
2. whenever $\mu = \sum_{0\le i\le n} p_i \cdot \mu_i$, there exists a $\nu \overset{\tau}{\Rightarrow}_{\mathrm{c}} \sum_{0\le i\le n} p_i \cdot \nu_i$ such that $\mu_i\ \mathcal{R}\ \nu_i$ for each $0 \le i \le n$ where $\sum_{0\le i\le n} p_i = 1$;
3. symmetrically for ν.

We say that μ and ν are weak distribution bisimilar, written as $\mu \mathrel{{}^{\bullet}{\approx}} \nu$, iff there exists a weak distribution bisimulation $\mathcal{R}$ such that $\mu\ \mathcal{R}\ \nu$. Moreover $s \mathrel{{}^{\bullet}{\approx}} r$ iff $\delta_s \mathrel{{}^{\bullet}{\approx}} \delta_r$.

Sometimes we need to consider relations between subdistributions, and for a relation $\mathcal{R} \subseteq Dist(S) \times Dist(S)$, we can extend it to a relation on $SubDist(S)$ (still denoted by $\mathcal{R}$) naturally as follows: $\mu\ \mathcal{R}\ \nu$ iff $\mu = \nu = 0$ or $\mu(S) = \nu(S)$ and $\mu/\mu(S)\ \mathcal{R}\ \nu/\nu(S)$.

The following lemma states that the weak distribution bisimilarity relation $\mathrel{{}^{\bullet}{\approx}}$ is linear and σ-linear. This is natural from the σ-linearity of $\overset{\alpha}{\rightarrow}_{\mathrm{c}}$ and $\overset{\alpha}{\Rightarrow}_{\mathrm{c}}$. We omit its proof.

Lemma 14 *The weak distribution bisimilarity relation $\mathrel{{}^{\bullet}{\approx}}$ is linear and σ-linear.*

In Definition 13, clause 1 is standard. clause 2 says that no matter how we split μ, there always exists a splitting of ν (possibly after internal transitions) to simulate the splitting of μ. Definition 13 is slightly different from Def. 5 in [9], where clause 2 is missing and clause 1 is replaced by: whenever $\mu \overset{\alpha}{\Rightarrow}_{\mathrm{c}} \sum_{0\le i\le n} p_i \cdot \mu_i$, there exists $\nu \overset{\alpha}{\Rightarrow}_{\mathrm{c}} \sum_{0\le i\le n} p_i \cdot \nu_i$ such that $\mu_i\ \mathcal{R}\ \nu_i$ for each $0 \le i \le n$. Essentially, this condition subsumes clause 2, since $\mu = \sum_{0\le i\le n} p_i \cdot \mu_i$ implies $\mu \overset{\tau}{\Rightarrow}_{\mathrm{c}} \sum_{0\le i\le n} p_i \cdot \mu_i$. As we prove in the following lemma, both definitions induce the same equivalence relation on PAs.

Lemma 15 *Let $\mathcal{P} = (S, Act_\tau, \rightarrow, \bar{s})$ be a* PA*. A relation $\mathcal{R} \subseteq Dist(S) \times Dist(S)$ is the weak distribution bisimilarity relation iff it is the largest relation satisfying: $\mu\ \mathcal{R}\ \nu$ implies:*

1. *whenever $\mu \overset{\alpha}{\Rightarrow}_{\mathrm{c}} \mu'$ with $\mu' \in Dist(S)$, there exists $\nu \overset{\alpha}{\Rightarrow}_{\mathrm{c}} \nu'$ such that $\mu'\ \mathcal{R}\ \nu'$,*
2. *whenever $\mu = \sum_{0\le i\le n} p_i \cdot \mu_i$, there exists $\nu \overset{\tau}{\Rightarrow}_{\mathrm{c}} \sum_{0\le i\le n} p_i \cdot \nu_i$ such that $\mu_i\ \mathcal{R}\ \nu_i$ for each $0 \le i \le n$ where $\sum_{0\le i\le n} p_i = 1$,*
3. *symmetrically for ν.*

Proof Since $\mu \overset{\alpha}{\rightarrow}_{\mathrm{c}} \mu'$ implies $\mu \overset{\alpha}{\Rightarrow}_{\mathrm{c}} \mu'$, $\mathcal{R}$ is a weak distribution bisimulation, and $\mathcal{R} \subseteq \mathrel{{}^{\bullet}{\approx}}$. For the other direction, we need to show that $\mathrel{{}^{\bullet}{\approx}}$ satisfies the conditions in Lemma 15, and we only need to check the first clause.

Assume $\alpha = \tau$. According to the definition of derivatives, $\mu \stackrel{\tau}{\Rightarrow}_c \mu'$ iff there exists

$$\begin{aligned} \mu &= \mu_0^{\rightarrow} + \mu_0^{\times}, \\ \mu_0^{\rightarrow} &\stackrel{\tau}{\rightarrow}_c \mu_1^{\rightarrow} + \mu_1^{\times}, \\ \mu_1^{\rightarrow} &\stackrel{\tau}{\rightarrow}_c \mu_2^{\rightarrow} + \mu_2^{\times}, \\ &\vdots \end{aligned}$$

such that $\mu' = \sum_{i \geq 0} \mu_i^{\times}$. By Definition 13, ν can simulate such a derivation at each step, namely, there exists

$$\begin{aligned} \nu &\stackrel{\tau}{\Rightarrow}_c \nu_0^{\rightarrow} + \nu_0^{\times}, \\ \nu_0^{\rightarrow} &\stackrel{\tau}{\Rightarrow}_c \nu_1^{\rightarrow} + \nu_1^{\times}, \\ \nu_1^{\rightarrow} &\stackrel{\tau}{\Rightarrow}_c \nu_2^{\rightarrow} + \nu_2^{\times}, \\ &\vdots \end{aligned}$$

such that $\mu_i^{\rightarrow} \bullet\!\!\approx \nu_i^{\rightarrow}$ and $\mu_i^{\times} \bullet\!\!\approx \nu_i^{\times}$ for each $i \geq 0$. Since $\bullet\!\!\approx$ is σ-linear, $(\sum_{i \geq 0} \mu_i^{\times}) \bullet\!\!\approx (\sum_{i \geq 0} \nu_i^{\times})$. From the transitivity of $\stackrel{\tau}{\Rightarrow}_c$, we have $\nu \stackrel{\tau}{\Rightarrow}_c \sum_{i=0}^{n} \nu_i^{\times} + \nu_n^{\rightarrow}$. Since μ' is a distribution, so is $\sum_{i \geq 0} \nu_i^{\times}$, and we have $\nu_n^{\rightarrow}$ converges to 0. By the continuity of $\stackrel{\tau}{\Rightarrow}_c$, we have $\nu \stackrel{\tau}{\Rightarrow}_c \sum_{i \geq 0} \nu_i^{\times}$.

In case $\mu \stackrel{\alpha}{\Rightarrow}_c \mu'$ with $\alpha \neq \tau$, we have $\mu \stackrel{\tau}{\Rightarrow}_c \mu_1' \stackrel{\alpha}{\rightarrow}_c \mu_2' \stackrel{\tau}{\Rightarrow}_c \mu'$. As shown above, there exists $\nu \stackrel{\tau}{\Rightarrow}_c \nu_1'$ such that $\mu_1' \mathcal{R} \nu_1'$, which indicates that there exists $\nu_1' \stackrel{\alpha}{\Rightarrow}_c \nu_2'$ such that $\mu_2' \mathcal{R} \nu_2'$ by Definition 13, which indicates that there exists $\nu_2' \stackrel{\tau}{\Rightarrow}_c \nu'$ such that $\mu' \mathcal{R} \nu'$. This completes the proof. □

The above lemma implies the transitivity of weak distribution bisimulation. On the other hand, we can restrict $\rightarrow_c$ to $\rightarrow$ in Definition 13 without changing weak distribution similarity:

Lemma 16 *Let $\mathcal{P} = (S, Act_\tau, \rightarrow, \bar{s})$ be a PA. A relation $\mathcal{R} \subseteq Dist(S) \times Dist(S)$ is the weak distribution bisimilarity relation iff it is the largest relation satisfying: $\mu \mathcal{R} \nu$ implies:*

1. *whenever $\mu \stackrel{\alpha}{\rightarrow} \mu'$, there exists $\nu \stackrel{\alpha}{\Rightarrow}_c \nu'$ such that $\mu' \mathcal{R} \nu'$,*
2. *whenever $\mu = \sum_{0 \leq i \leq n} p_i \cdot \mu_i$, there exists $\nu \stackrel{\tau}{\Rightarrow}_c \sum_{0 \leq i \leq n} p_i \cdot \nu_i$ such that $\mu_i \mathcal{R} \nu_i$ for each $0 \leq i \leq n$ where $\sum_{0 \leq i \leq n} p_i = 1$,*
3. *symmetrically for ν.*

Proof Since $\mu \stackrel{\alpha}{\rightarrow} \mu'$ implies $\mu \stackrel{\alpha}{\rightarrow}_c \mu'$, $\bullet\!\!\approx \subseteq \mathcal{R}$. For the other direction, we only need to show that $\mathcal{R}$ is a bisimulation relation.

First it is easy to see that $\mathcal{R}$ is linear. Let $\mu \mathcal{R} \nu$. If $\mu \stackrel{\alpha}{\rightarrow}_c \mu'$, then by Lemma 11, $\mu' = \sum_{s \in Supp(u)} \mu(s)\mu_s$, such that $\mu_s = \sum_{i=1}^{n} p_i \mu_{s,i}$, where $s \stackrel{\alpha}{\rightarrow} \mu_{s,i}$ and $\sum_{i=1}^{n} p_i = 1$. That is to say, there exists $\mu = \sum_{i=1}^{m} w_i \mu_i$, where $\sum_{i=1}^{m} w_i = 1$, such that $\mu_i \stackrel{\alpha}{\rightarrow} \mu_i'$ and $\sum_{i=1}^{m} w_i \mu_i' = \mu'$. By the second clause, there exist $\nu \stackrel{\tau}{\Rightarrow}_c \sum_{i=1}^{m} w_i \nu_i$, such that $\mu_i \mathcal{R} \nu_i$. Then there exists $\nu_i \stackrel{\alpha}{\Rightarrow}_c \nu_i'$, such that $\mu_i' \mathcal{R} \nu_i'$. Then we have $\nu \stackrel{\alpha}{\Rightarrow}_c \sum_{i=1}^{m} w_i \nu_i'$. From the linearity of $\mathcal{R}$, we get $\mu' \mathcal{R} \sum_{i=1}^{m} w_i \nu_i'$. □

4.3 Late weak bisimulation

Clause 2 in Definition 13 allows arbitrary splittings, which is essentially the main reason that weak distribution bisimulation is unrealistically strong. In order to establish a bisimulation relation, all possible splittings of μ must be matched by ν (possibly after some internal transitions). As splittings into Dirac distributions are also considered, the individual behaviors of each single state in $Supp(\mu)$ must be matched too. However, our bisimulation is distribution-based, thus the behaviors of distributions should be matched rather than those of states. We are about to propose a novel definition of late distribution bisimulation. Before that, we still need some notations. The following definition of transition consistency is derived from [22].

Definition 17 A distribution μ is *transition consistent,* written as $\overrightarrow{\mu}$, if for any $s \in Supp(\mu)$ and $\alpha \neq \tau$, $s \stackrel{\alpha}{\Rightarrow} \gamma$ for some γ implies $\mu \stackrel{\alpha}{\Rightarrow}_{\mathrm{c}} \gamma'$ for some γ'.

For a distribution being transition consistent, all states in the support of the distribution should have the same set of *enabled visible actions.* One of the key properties of transition consistent distributions is that $\mu \stackrel{\alpha}{\Rightarrow}$ whenever $s \stackrel{\alpha}{\Rightarrow}$ for some state $s \in Supp(\mu)$. In contrast, when a distribution μ is not transition consistent, there must be a weak α transition of some state in $Supp(\mu)$ being *blocked.* In the sequel, we will adapt the lifting of the transition relation to avoid that a difference in $\stackrel{\tau}{\to}$ transitions leads to blocked $\stackrel{\alpha}{\Rightarrow}$ transitions.

We now introduce $\hookrightarrow$, an alternative lifting of transitions of states to transitions of distributions that differs from the standard definition used in [9,15]. There, a distribution is able to perform a transition labelled with α *if and only if* all the states in its support can perform transitions with the very same label. In contrast, the transition relation $\hookrightarrow$ behaves like a weak transition, where every state in the support of μ may perform an invisible transition independently from other states.

Definition 18 $\mu \stackrel{\alpha}{\hookrightarrow} \mu'$ iff

1. either for each $s \in Supp(\mu)$ there exists $s \stackrel{\alpha}{\to} \mu_s$ such that

$$\mu' = \sum_{s \in Supp(\mu)} \mu(s) \cdot \mu_s,$$

2. or $\alpha = \tau$ and there exists $s \in Supp(\mu)$ and $s \stackrel{\tau}{\to} \mu_s$ such that

$$\mu' = (\mu - s) + \mu(s) \cdot \mu_s.$$

In the definition of late distribution bisimulation, this extension will be used to prevent τ transitions of states from being blocked. Below follows an example:

Example 19 Let $\mu = \{s_1 : 0.4, s_2 : 0.6\}$ such that $s_1 \stackrel{\tau}{\to} \delta_{s_1'} \stackrel{\alpha}{\to} \mu_1$, $s_1 \stackrel{\beta}{\to} \mu_2$, $s_2 \stackrel{\alpha}{\to} \mu_3$, and $s_2 \stackrel{\beta}{\to} \mu_4$, where $\alpha \neq \beta$ are visible actions. According to clause 1 of Definition 18, we will have $\mu \stackrel{\beta}{\hookrightarrow} (0.4 \cdot \mu_2 + 0.6 \cdot \mu_4)$. Without clause 2, this would be the only transition of μ, since the τ transition of s_1 and the α transition of s_2 will be blocked, as s_1 and s_2 cannot perform transitions with labels τ and α at the same time.

Note that the α transition is blocked by the τ transition of s_1, so according to clause 2 of Definition 18, we in addition have

$$\mu \stackrel{\tau}{\hookrightarrow} (0.4 \cdot \delta_{s_1'} + 0.6 \cdot \delta_{s_2}) \stackrel{\alpha}{\hookrightarrow} (0.4 \cdot \mu_1 + 0.6 \cdot \mu_3). \qquad \square$$

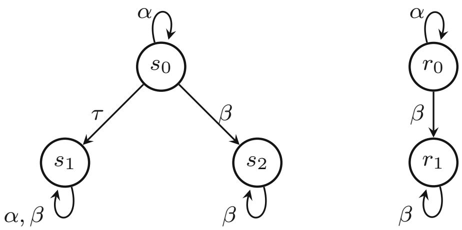

Fig. 6 Using $\rightarrow$ instead of $\hookrightarrow$ will lead to a finer relation

Note that in clause 1 of Definition 13, $\rightarrow$ can be replaced by $\hookrightarrow$ without changing the resulting equivalence relation, as the same effect can be obtained by a suitable splitting in clause 2. In the example, we could have split μ into $0.4 \cdot \delta_{s_1} + 0.6 \cdot \delta_{s_2}$, such that no transition is blocked in the resulting distributions.

With this lifting transition relation $\stackrel{\alpha}{\hookrightarrow}$, we now propose the notion of late distribution bisimulation as follows:

Definition 20 $\mathcal{R} \subseteq \mathit{Dist}(S) \times \mathit{Dist}(S)$ is a *late distribution bisimulation* iff $\mu \; \mathcal{R} \; \nu$ implies:

1. whenever $\mu \stackrel{\alpha}{\hookrightarrow}_{\mathrm{c}} \mu'$, there exists a $\nu \stackrel{\alpha}{\Rightarrow}_{\mathrm{c}} \nu'$ such that $\mu' \; \mathcal{R} \; \nu'$;
2. if not $\overrightarrow{\mu}$, then there exists $\mu = \sum_{0 \le i \le n} p_i \cdot \mu_i$ and $\nu \stackrel{\tau}{\Rightarrow}_{\mathrm{c}} \sum_{0 \le i \le n} p_i \cdot \nu_i$ such that $\overrightarrow{\mu_i}$ and $\mu_i \; \mathcal{R} \; \nu_i$ for each $0 \le i \le n$ where $\sum_{0 \le i \le n} p_i = 1$;
3. symmetrically for ν.

We say that μ and ν are late distribution bisimilar, written as $\mu \approx^{\bullet} \nu$, iff there exists a late distribution bisimulation $\mathcal{R}$ such that $\mu \; \mathcal{R} \; \nu$. Moreover $s \approx^{\bullet} r$ iff $\delta_s \approx^{\bullet} \delta_r$.

In clause 1, this definition differs from Definition 13 by the use of $\hookrightarrow$. It is straightforward to show that $\hookrightarrow$ can also be used in Definition 13 without changing the resulting bisimilarity. However, in Definition 20, using $\rightarrow$ instead of $\hookrightarrow$ will lead to a finer relation.

Example 21 We consider the model in Fig. 6. If we use $\rightarrow$ in defining weak distribution bisimulation, then it is easy to check that $0.5\delta_{s_0} + 0.5\delta_{r_0}$ and δ_{r_0} would be bisimilar. However, according to our definition (where we use $\hookrightarrow$), they are not bisimilar. We show this by contradiction. Notice that $0.5\delta_{s_0} + 0.5\delta_{r_0} \stackrel{\tau}{\hookrightarrow} 0.5\delta_{s_1} + 0.5\delta_{r_0} \stackrel{\beta}{\hookrightarrow} 0.5\delta_{s_1} + 0.5\delta_{r_1}$ can only be simulated by $\delta_{r_0} \stackrel{\tau}{\Rightarrow}_{\mathrm{c}} \delta_{r_0} \stackrel{\beta}{\Rightarrow}_{\mathrm{c}} \delta_{r_1}$, so $0.5\delta_{s_1} + 0.5\delta_{r_1}$ and δ_{r_1} must be bisimilar. However, $0.5\delta_{s_1} + 0.5\delta_{r_1}$ is not transition consistent, so it can be written as $\sum_{i=1}^{n} p_i \mu_i$ such that every μ_i is transition consistent. It is easy to see here every μ_i must be a Dirac distribution, and δ_{s_1} must appear, but $p_i \delta_{r_1}$ and $p_i \delta_{s_1}$ can never be bisimilar.

The key difference between Definitions 13 and 20, however, is in clause 2. As we mentioned, in Definition 13, every split of μ should be matched by a corresponding split of ν, while in Definition 20, we only require that at least one transition consistent split of μ is matched. We do not need to require that ν_i is transition consistent, as we will show later that $\overrightarrow{\mu_i}$ and $\mu_i \; \mathcal{R} \; \nu_i$ implies $\overrightarrow{\nu_i}$. According to Definition 17, splittings to transition consistent distributions ensure that all possible transitions will be considered eventually, as no transition of an individual state is blocked. Therefore, clause 1 suffices to capture every visible behavior.

By introducing transition consistent distributions, we try to group states with the same set of enabled visible actions together and do not distinguish them in a distribution. This idea is mainly motivated by the work in [8], where all states with the same enabled actions are non-distinguishable from the outside. Under this assumption, a model checking algorithm was proposed. By avoiding splitting transition consistent distributions, we essentially delay the probabilistic transitions until transition consistency is broken. This explains the name "late distribution bisimulation". Further, if restricting to models without internal action τ, our notion of late distribution bisimulation agrees with decentralized bisimulation in [38].

Example 22 We will show that in Fig. 3, $s_0 \approx^\bullet s_0'$. Let

$$\mathcal{R} = \{(\delta_{s_0}, \delta_{s_0'}), (\delta_{s_0}, \{s_5 : 0.5, s_6 : 0.5\})\} \cup \Delta$$

where Δ is the identity relation. It is easy to show that $\mathcal{R}$ is a late distribution bisimulation. The only non-trivial case is when $\delta_{s_0'} \xrightarrow{\tau} \{s_5 : 0.5, s_6 : 0.5\}$. But then δ_{s_0} can simulate it without performing any transition, i.e., $\delta_{s_0} \stackrel{\tau}{\Rightarrow} \delta_{s_0}$. Since $\delta_{s_0}\ \mathcal{R}\ \{s_5 : 0.5, s_6 : 0.5\}$, clause 1 of Definition 20 is satisfied. Moreover both δ_{s_0} and $\{s_5 : 0.5, s_6 : 0.5\}$ are transition consistent, thus we do not need to split them any further. Conversely, we can show that $\mathcal{R}$ is not a weak distribution bisimulation. According to clause 1 of Definition 13, we require that for any split of $\{s_5 : 0.5, s_6 : 0.5\}$, there must exist a matching split of δ_{s_0}, which cannot be established. For instance the split $\{s_5 : 0.5, s_6 : 0.5\} \equiv 0.5 \cdot \delta_{s_5} + 0.5 \cdot \delta_{s_6}$ cannot be matched by any split of δ_{s_0}. □

The following example shows that the transition consistency condition of Definition 20 is necessary to not equate states which should be distinguished.

Example 23 Suppose there are two states s_0 and r_0 such that $s_0 \xrightarrow{\tau} s_1$ and $r_0 \xrightarrow{\tau} \{r_1 : 0.5, r_2 : 0.5\}$ where all of s_1, r_1, and r_2 have a transition to themselves with labels τ, in addition, $r_1 \xrightarrow{\alpha} r_1$ where $\alpha \neq \tau$. Let

$$\mathcal{R} = \{(\delta_{s_0}, \delta_{r_0}), (\delta_{s_1}, \{r_1 : 0.5, r_2 : 0.5\})\}.$$

If we dropped the transition consistency condition from Definition 20, we could show that $\mathcal{R}$ is a late distribution bisimulation, and therefore $s_0 \approx^\bullet r_0$. Because the distribution $\{r_1 : 0.5, r_2 : 0.5\}$ can only perform a τ transition to itself, while the α transition of r_1 would then be blocked. However, s_0 and r_0 should be distinguished, because r_0 can reach r_1 with positive probability, which is a state able to perform a transition with visible label α. Note that as $\{r_1 : 0.5, r_2 : 0.5\}$ is not transition consistent, we should split it further according to Definition 20. Thus we can prove that $\mathcal{R}$ is not a late distribution bisimulation, i.e., $s_0 \not\approx^\bullet r_0$. □

The following lemma states that μ and ν must be transition consistent or not at the same time, if they are late distribution bisimilar.

Lemma 24 *$\mu \approx^\bullet \nu$ and $\overrightarrow{\mu}$ imply $\overrightarrow{\nu}$.*

Proof By contradiction. Assume $\mu \approx^\bullet \nu$ and $\overrightarrow{\mu}$, but not $\overrightarrow{\nu}$. Since $\mu \approx^\bullet \nu$, there exists a late distribution bisimulation $\mathcal{R}$ such that $\mu\ \mathcal{R}\ \nu$. Moreover, $\mu \stackrel{\alpha}{\Rightarrow}$ implies $\nu \stackrel{\alpha}{\Rightarrow}$ and vice versa for any α. Therefore, $EA(\mu) = EA(\nu)$, where $EA(\mu) = \{\alpha \mid \exists \mu'.\mu \stackrel{\alpha}{\Rightarrow} \mu'\}$, similarly for $EA(\nu)$. Since ν is not transition consistent, there exists $s \in Supp(\nu)$ such that $s \stackrel{\alpha}{\Rightarrow}$ with $\alpha \notin EA(\nu)$, i.e., some transitions of states in $Supp(\nu)$ with label α are blocked. This implies that there exists a non-trivial $\nu = \sum_{i \in I} p_i \cdot \nu_i$ with $\overrightarrow{\nu_i}$ for each $i \in I$ such that $\nu_j \stackrel{\alpha}{\Rightarrow}$ for some $j \in I$. Since $\overrightarrow{\mu}$ and $\alpha \notin EA(\mu)$, there does not exist $\mu \stackrel{\tau}{\Rightarrow} \sum_{i \in I} p_i \cdot \mu_i$ such that $\mu_i \stackrel{\alpha}{\Rightarrow}$ for some $i \in I$. This contradicts the assumption that $\mu \approx^\bullet \nu$. □

The following lemma shows that the late distribution bisimilarity relation $\approx^\bullet$ is linear.

Lemma 25 *The late distribution bisimilarity relation $\approx^{\bullet}$ is linear.*

Proof By contradiction we assume $\approx^{\bullet}$ is not linear. We consider the smallest linear relation containing $\approx^{\bullet}$ (i.e. the intersection of all linear relations containing $\approx^{\bullet}$, which is still linear and strictly larger than $\approx^{\bullet}$) and denote it by $\approx'$. We show that $\approx'$ is a late distribution bisimulation. Let $\mu \approx' \nu$. If $\mu \approx^{\bullet} \nu$, naturally the conditions in Definition 20 hold. If $\mu \not\approx^{\bullet} \nu$, then $\mu \approx' \nu$ because of linearity, so there exist $\mu_i \approx^{\bullet} \nu_i$ for $0 \leq i \leq n$ s.t. $\mu = \sum_{i=0}^{n} p_i \mu_i$ and $\nu = \sum_{i=0}^{n} p_i \nu_i$. From the linearity of $\stackrel{\alpha}{\hookrightarrow}_{\mathrm{c}}$ and $\stackrel{\alpha}{\Rightarrow}_{\mathrm{c}}$, the first clause in Definition 20 holds. For the second clause, if μ is not transition consistent, then for those μ_i which are not transition consistent, there exists $\nu_i \stackrel{\tau}{\Rightarrow} \nu_i'$, s.t. $\mu_i \approx^{\bullet} \nu_i'$, and naturally we get a split of μ by combining all splits of μ_i and corresponding weak transition of ν. Then from the linearity of $\stackrel{\tau}{\Rightarrow}_{\mathrm{c}}$, we can see the second clause holds. Therefore, $\approx'$ is a late distribution bisimulation. However, $\approx^{\bullet}$ is the largest late distribution bisimulation, which is a contradiction! □

In Definition 20, we can split μ in the coarsest way when μ is not transition consistent. Basically we split μ as

$$\mu = \sum_{A \subseteq Act} \mu_A \qquad \text{where } \mu_A = \sum_{EA(s)=A} \mu(s)\delta_s. \tag{4}$$

It is obvious that the split given in (4) is the coarsest one that makes every component μ_A in the split transition consistent. By using this coarsest split, we can get the same relation as that in Definition 20, as is stated in the following lemma:

Lemma 26 *A relation $\mathcal{R} \subseteq Dist(S) \times Dist(S)$ is the late distribution bisimilarity relation iff it is the largest relation satisfying: $\mu\ \mathcal{R}\ \nu$ implies:*

1. *whenever $\mu \stackrel{\alpha}{\hookrightarrow}_{\mathrm{c}} \mu'$, there exists a $\nu \stackrel{\alpha}{\Rightarrow}_{\mathrm{c}} \nu'$ such that $\mu'\ \mathcal{R}\ \nu'$;*
2. *if not $\overrightarrow{\mu}$, then for the coarsest split $\mu = \sum_A \mu_A$, there exists $\nu \stackrel{\tau}{\Rightarrow}_{\mathrm{c}} \nu'$ such that $\nu' = \sum_A \nu_A'$ and $\mu_A\ \mathcal{R}\ \nu_A'$ for each $A \subseteq Act$;*
3. *symmetrically for ν.*

Proof Let $\mathcal{R}$ be a relation satisfying the conditions in the lemma, then it is a late distribution bisimulation and naturally $\mathcal{R} \subseteq \approx^{\bullet}$. For the other direction, we show that $\approx^{\bullet}$ satisfies the conditions in the lemma so that $\approx^{\bullet} \subseteq \mathcal{R}$. We only need to check the second clause. If μ is not transition consistent, then there exists a split $\mu = \sum_{i=0}^{n} p_i \mu_i$ and $\nu \stackrel{\tau}{\Rightarrow}_{\mathrm{c}} \sum_{i=0}^{n} p_i \nu_i$, s.t. $\mu_i \approx^{\bullet} \nu_i$. This split is finer than $\mu = \sum_A \mu_A$, and we can combine some components to this coarsest split. From the linearity of $\approx^{\bullet}$, we can see the second clause holds. □

From Lemma 26, we can know that $\approx^{\bullet}$ is σ-linear.

Lemma 27 *The late distribution bisimilarity relation $\approx^{\bullet}$ is σ-linear.*

Basically we adopt a similar proof as Lemma 25, replacing n with ∞. The only difference is that, when we check the second clause, we need Lemma 26 to do the split of every μ_i so that the resulting combined split is still finite and we can simulate it from the linearity of $\approx^{\bullet}$. We omit this proof.

The lemma below resembles Lemma 15, which can be proved similarly as Lemma 15.

Lemma 28 *Let $\mathcal{P} = (S, Act_\tau, \rightarrow, \bar{s})$ be a PA. A relation $\mathcal{R} \subseteq Dist(S) \times Dist(S)$ is the late distribution bisimilarity relation iff it is the largest relation satisfying: $\mu\ \mathcal{R}\ \nu$ implies:*

1. *whenever $\mu \stackrel{\alpha}{\Rightarrow}_{\mathrm{c}} \mu'$ with $\mu' \in Dist(S)$, there exists $\nu \stackrel{\alpha}{\Rightarrow}_{\mathrm{c}} \nu'$ such that $\mu'\ \mathcal{R}\ \nu'$,*

2. *if not* $\overrightarrow{\mu}$, *then there exists* $\mu = \sum_{0 \le i \le n} p_i \cdot \mu_i$ *and* $\nu \stackrel{\tau}{\Rightarrow}_{\mathrm{c}} \sum_{0 \le i \le n} p_i \cdot \nu_i$ *such that* $\overrightarrow{\mu_i}$ *and* $\mu_i \mathcal{R} \nu_i$ *for each* $0 \le i \le n$ *where* $\sum_{0 \le i \le n} p_i = 1$;
3. *symmetrically for* ν.

Proof It suffices to check that $\approx^{\bullet}$ satisfies the first clause in Lemma 28. We assume $\mu \approx^{\bullet} \nu$.

If $\mu \stackrel{\tau}{\Rightarrow}_{\mathrm{c}} \mu'$, then there exists

$$\mu = \mu_0^{\rightarrow} + \mu_0^{\times},$$
$$\mu_0^{\rightarrow} \stackrel{\tau}{\rightarrow}_{\mathrm{c}} \mu_1^{\rightarrow} + \mu_1^{\times},$$
$$\mu_1^{\rightarrow} \stackrel{\tau}{\rightarrow}_{\mathrm{c}} \mu_2^{\rightarrow} + \mu_2^{\times},$$
$$\vdots$$

such that $\mu' = \sum_{i \ge 0} \mu_i^{\times}$. Then we have $\mu \stackrel{\tau}{\hookrightarrow}_{\mathrm{c}} \mu_1^{\rightarrow} + \mu_0^{\times} + \mu_1^{\times} \stackrel{\tau}{\hookrightarrow}_{\mathrm{c}} \mu_2^{\rightarrow} + \mu_0^{\times} + \mu_1^{\times} + \mu_2^{\times} \stackrel{\tau}{\hookrightarrow}_{\mathrm{c}} \cdots$. Then there exists $\nu \stackrel{\tau}{\Rightarrow}_{\mathrm{c}} \nu_n$, such that $\mu_n^{\rightarrow} + \sum_{i=1}^{n} \mu_i^{\times} \approx^{\bullet} \nu_n$, for each $n \in \mathbb{N}$. Since $\{\mu_n\}$ is a bounded sequence in $\mathbb{R}^S$, there exists a converging subsequence, which we still denote by $\{\nu_n\}$. Let $\nu' = \lim_{n \to \infty} \nu_n$. From the continuity of $\stackrel{\tau}{\Rightarrow}_{\mathrm{c}}$, we get $\nu \stackrel{\tau}{\Rightarrow}_{\mathrm{c}} \nu'$.

If $\mu \stackrel{\alpha}{\Rightarrow}_{\mathrm{c}} \mu'$ with $\alpha \ne \tau$, we have $\mu \stackrel{\tau}{\Rightarrow}_{\mathrm{c}} \mu_1' \stackrel{\alpha}{\rightarrow}_{\mathrm{c}} \mu_2' \stackrel{\tau}{\Rightarrow}_{\mathrm{c}} \mu'$. As shown above, there exists $\nu \stackrel{\tau}{\Rightarrow}_{\mathrm{c}} \nu_1'$ such that $\mu_1' \mathcal{R} \nu_1'$, which indicates that there exists $\nu_1' \stackrel{\alpha}{\Rightarrow}_{\mathrm{c}} \nu_2'$ such that $\mu_2' \mathcal{R} \nu_2'$, which indicates that there exists $\nu_2' \stackrel{\tau}{\Rightarrow}_{\mathrm{c}} \nu'$ such that $\mu' \mathcal{R} \nu'$. This completes the proof. □

Similarly, we have the following result, which can be proved analogously to Lemma 16.

Lemma 29 *Let* $\mathcal{P} = (S, Act_\tau, \rightarrow, \bar{s})$ *be a* PA. *A relation* $\mathcal{R} \subseteq Dist(S) \times Dist(S)$ *is the late distribution bisimilarity relation iff it is the largest relation satisfying:* $\mu \mathcal{R} \nu$ *implies:*

1. *whenever* $\mu \stackrel{\alpha}{\hookrightarrow} \mu'$, *there exists* $\nu \stackrel{\alpha}{\Rightarrow}_{\mathrm{c}} \nu'$ *such that* $\mu' \mathcal{R} \nu'$,
2. *if not* $\overrightarrow{\mu}$, *then there exists* $\mu = \sum_{0 \le i \le n} p_i \cdot \mu_i$ *and* $\nu \stackrel{\tau}{\Rightarrow}_{\mathrm{c}} \sum_{0 \le i \le n} p_i \cdot \nu_i$ *such that* $\overrightarrow{\mu_i}$ *and* $\mu_i \mathcal{R} \nu_i$ *for each* $0 \le i \le n$ *where* $\sum_{0 \le i \le n} p_i = 1$;
3. *symmetrically for* ν.

The following theorem shows that $\approx^{\bullet}$ is an equivalence relation and $\approx^{\bullet}$ is strictly coarser than ${}^{\bullet}\!\approx$.

Theorem 30

1. ${}^{\bullet}\!\approx \subset \approx^{\bullet}$;
2. $\approx^{\bullet}$ *is an equivalence relation.*

Proof The first clause ${}^{\bullet}\!\approx \subset \approx^{\bullet}$ is easy to establish, since the second condition of Definition 13 implies the second condition of Definition 20. The PA in Fig. 1 shows that the inclusion is strict.

Now we prove that $\approx^{\bullet}$ is an equivalence relation. We prove transitivity (other parts are easy). For any μ, ν, and γ, assume $\mu \approx^{\bullet} \nu$ and $\nu \approx^{\bullet} \gamma$, we prove that $\mu \approx^{\bullet} \gamma$. We shall prove:

1. Whenever $\mu \stackrel{\alpha}{\Rightarrow}_{\mathrm{c}} \mu'$, there exists $\gamma \stackrel{\alpha}{\Rightarrow}_{\mathrm{c}} \gamma'$ such that $\mu' \mathcal{R} \gamma'$. This is achieved by applying Lemma 28.

2. If not $\overrightarrow{\mu}$, then for the coarsest split $\mu = \sum_A \mu_A$, there exists $\nu \stackrel{\tau}{\Rightarrow}_c \nu'$, s.t. $\nu' = \sum_A \nu'_A$ and $\mu_A \approx^{\bullet} \nu'_A$. Because $\nu \approx^{\bullet} \gamma$, there exists $\gamma \stackrel{\tau}{\Rightarrow}_c \gamma'$, s.t. $\nu' \approx^{\bullet} \gamma'$. From Lemma 24, ν' is not transition consistent, and by adopting Lemma 26 again we know there exists $\gamma' \stackrel{\tau}{\Rightarrow}_c \gamma''$, s.t. $\gamma'' = \sum_A \gamma''_A$ and $\nu'_A \approx^{\bullet} \gamma''_A$. Since $\mu_A, \nu'_A, \gamma''_A$ are all transition consistent, we have $\mu_A \approx^{\bullet} \gamma''_A$. From the transitivity of $\stackrel{\tau}{\Rightarrow}_c$, we know that $\gamma \stackrel{\tau}{\Rightarrow}_c \sum_A \gamma''_A$.

This completes our proof. □

5 Properties of late distribution bisimilarity

In this section we show that results established in [38] can be extended to the setting where internal transitions are abstracted away.

We concentrate on two properties of late distribution bisimulation: compositionality and preservation of trace distributions. When general schedulers are considered, the two properties do not hold, hence we will restrict ourselves to partial information distributed schedulers. We mention that both partial information and distributed schedulers were proposed to rule out unrealistic behaviors of general schedulers; see [8,23] for more details.

We first define some notations. In the following, we restrict ourselves to schedulers satisfying the following condition: For any $\pi \in \mathit{Paths}^*$, $\xi(\pi)(\alpha, \mu) > 0$ and $\xi(\pi)(\beta, \nu) > 0$ imply $\alpha = \beta$. In other words, ξ always chooses transitions with the same label at each step. This class of schedulers suffices for our purpose. To distinguish between scheduler classes, we parameterize transition relations with schedulers explicitly. A transition from s to μ with label α is induced by a scheduler ξ, written as $s \stackrel{\alpha}{\rightarrow}_\xi \mu$, iff $\mu \equiv \sum_{\mu' \in \mathit{Dist}(S)} \xi(s)(\alpha, \mu') \cdot \mu'$. As before, such a transition relation can be lifted to distributions: $\mu \stackrel{\alpha}{\rightarrow}_\xi \nu$ to denote that μ can evolve into ν by performing a transition with label α under the guidance of ξ, where $s \stackrel{\alpha}{\rightarrow}_\xi \nu_s$ for each $s \in \mathit{Supp}(\mu)$ and $\nu \equiv \sum_{s \in \mathit{Supp}(\mu)} \mu(s) \cdot \nu_s$. Since no a priori information is available, given a distribution μ, for each $s \in \mathit{Supp}(\mu)$, we simply use s as the history information for ξ to guide the execution, which corresponds to a *memoryless* scheduler and suffices for the purpose of defining bisimulation.

We can also define $\mu \stackrel{\alpha}{\hookrightarrow}_\xi \mu'$ analogously to Definition 18. Moreover, weak transitions $s \stackrel{\alpha}{\Rightarrow}_\xi \mu$ and their lifting to distributions can be defined similarly.

Below we give an alternative to Definition 20, where schedulers are considered explicitly.

Definition 31 Let $\mathcal{S}$ be a fixed class of schedulers. $\mathcal{R} \subseteq \mathit{Dist}(S) \times \mathit{Dist}(S)$ is a late distribution bisimulation with respect to $\mathcal{S}$ iff $\mu \; \mathcal{R} \; \nu$ implies:

1. whenever $\mu \stackrel{\alpha}{\hookrightarrow}_{\xi_1} \mu'$ for some $\xi_1 \in \mathcal{S}$, there exist $\xi_2 \in \mathcal{S}$ and $\nu \stackrel{\alpha}{\Rightarrow}_{\xi_2} \nu'$ such that $\mu' \; \mathcal{R} \; \nu'$;
2. if not $\overrightarrow{\mu}$, then there exist $\xi \in \mathcal{S}$ and $\mu = \sum_{0 \le i \le n} p_i \cdot \mu_i$ and $\nu \stackrel{\tau}{\Rightarrow}_\xi \sum_{0 \le i \le n} p_i \cdot \nu_i$ such that $\overrightarrow{\mu_i}$ and $\mu_i \; \mathcal{R} \; \nu_i$ for each $0 \le i \le n$ where $\sum_{0 \le i \le n} p_i = 1$;
3. symmetrically for ν.

We write $\mu \approx^{\bullet}_{\mathcal{S}} \nu$ iff there exists a late distribution bisimulation $\mathcal{R}$ with respect to $\mathcal{S}$ such that $\mu \; \mathcal{R} \; \nu$, and we write $s \approx^{\bullet}_{\mathcal{S}} r$ iff $\delta_s \approx^{\bullet}_{\mathcal{S}} \delta_r$.

In contrast to Definition 20, in Definition 31, every transition is induced by a scheduler in $\mathcal{S}$. Obviously, when $\mathcal{S}$ is the set of all schedulers, these two definitions coincide. Thus, $s_1 \approx^{\bullet} s_2 \iff s_1 \approx^{\bullet}_{\mathcal{S}_D} s_2$, provided s_1 and s_2 contain no parallel operators, as in this case $\mathcal{S}_D$ is the set of all schedulers.

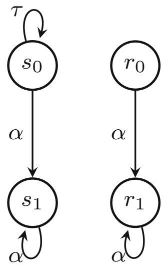

Fig. 7 Late distribution bisimulation cannot imply trace distribution equivalence

The lemma below shows that distribution bisimulation and partial information schedulers are closely related. It shows that partial information schedulers are enough to discriminate late distribution bisimilarity with respect to arbitrary schedulers.

Lemma 32 *For any states s_1 and s_2, $s_1 \approx^{\bullet} s_2$ iff $s_1 \approx^{\bullet}_{\mathcal{S}_P} s_2$.*

Proof This equivalence is straightforward from Definition 20, since we always group states with the same enable visible actions together and let them either perform transitions with the same visible action at the same time, or an internal transition independently, which never breaks the conditions of partial information schedulers. In other words, all transitions we consider in Definition 20 are induced by some scheduler in $\mathcal{S}_P$. □

In the following we investigate the implication between late distribution bisimulation and trace distribution equivalence. In the classical setting, weak bisimulation and trace equivalence are incomparable, and divergence sensitivity is added to guarentee the implication. It is the same for our probabilistic setting. The following example shows that only late distribution bisimulation cannot imply trace distribution equivalence.

Example 33 Our model is shown in Fig. 7. It is obvious that $s_0 \approx^{\bullet} r_0$. Let the scheduler $\xi_1(s_0) = (\tau, \delta_{s_0})$. Then $Pr^{s_0}_{\xi_1}(C_\alpha) = 0$. However, for any scheduler ξ_2, $\xi_2(r_0) = (\alpha, \delta_{r_1})$, so $Pr^{r_0}_{\xi_2}(C_\alpha) = 1$. Therefore, s_0 and r_0 are not trace distribution equivalent with respect to any set of schedulers containing ξ_1.

Definition 34 Given a relation $\mathcal{R} \subseteq Dist(S) \times Dist(S)$, we say $\mu \in Dist(S)$ is $\mathcal{R}$-divergent, if there exists an infinite sequence $\mu \xrightarrow{\tau} \mu_1 \xrightarrow{\tau} \mu_2 \xrightarrow{\tau} \cdots$, such that $\mu \;\mathcal{R}\; \mu_i$ for each $i \geq 1$. We say $s \in S$ is $\mathcal{R}$-divergent, if δ_s is $\mathcal{R}$-divergent.

Definition 35 We say a relation $\mathcal{R} \subseteq Dist(S) \times Dist(S)$ is *divergence-sensitive,* if $\mu \;\mathcal{R}\; \nu$ implies that μ and ν are both or neither $\mathcal{R}$-divergent.

Now we add divergence sensitivity to late distribution bisimulation as follows:

Definition 36 A relation $\mathcal{R} \subseteq Dist(S) \times Dist(S)$ is a *divergence-sensitive late distribution bisimulation,* iff $\mathcal{R}$ is divergence-sensitive and a late distribution bisimulation. We write $\mu \approx^{\bullet \mathrm{div}} \nu$ iff there exists a divergence-sensitive late distribution bisimulation $\mathcal{R}$ such that $\mu \;\mathcal{R}\; \nu$. Moreover $s \approx^{\bullet \mathrm{div}} r$ iff $\delta_s \approx^{\bullet \mathrm{div}} \delta_r$.

The following result shows that divergence-sensitive late distribution bisimulation implies trace distribution equivalence.

Theorem 37 *For any states s_1 and s_2, $s_1 \approx^{\bullet \mathrm{div}} s_2$ implies $s_1 \equiv_{\mathcal{S}_P} s_2$.*

Proof (Sketch, see the appendix for details) We assume given states $s_1 \approx^{\bullet \mathrm{div}} s_2$, a partial information scheduler ξ_1, and a trace ς starting with $EA(s_1) = EA(s_2)$. Basically we need to construct a partial information scheduler ξ_2 that ensures $Pr^{s_1}_{\xi_1}(C_\varsigma) = Pr^{s_2}_{\xi_2}(C_\varsigma)$.

Since $s_1 \approx^{\bullet \mathrm{div}} s_2$, for $\pi \in Paths^*(s_1)$ and $\alpha \in Act_\tau$ with $\xi_1(\pi, \alpha) > 0$, there exists a sequence of transitions $s_1 \overset{\alpha_1}{\hookrightarrow}_c \mu_1 \overset{\alpha_2}{\hookrightarrow}_c \cdots \overset{\alpha_n}{\hookrightarrow}_c \mu_n$, which can be simulated by some sequence $s_2 \overset{\alpha_1}{\Rightarrow}_c \nu_1 \overset{\alpha_1}{\Rightarrow}_c \cdots \overset{\alpha_n}{\Rightarrow}_c \nu_n$. From Definition 20, μ_i and ν_i can be splitted into subdistributions which are transition consistent and they keep the bisimilarity relation. We put strong transitions in $s_2 \overset{\alpha_1}{\Rightarrow}_c \nu_1 \overset{\alpha_2}{\Rightarrow}_c \ldots \overset{\alpha_n}{\Rightarrow}_c \nu_n$ to ξ_2 to form this scheduler. Because all paths in ξ_2 are simulated through ξ_1, given any path in ξ_2, we can always find a path in ξ_1 that has the same trace of enabled actions. Therefore its probability of choosing transitions with a specific visible action can be chosen to be equal to ξ_1; this also ensures that the scheduler ξ_2 is a partial information scheduler.

Now we need to show that, for any finite trace ς, $Pr^{s_1}_{\xi_1}(C_\varsigma) = Pr^{s_2}_{\xi_2}(C_\varsigma)$. It is not difficult to see this. We first prove that the two schedulers have the same distribution on traces with length at most n which share the same prefix ς, which is natural through the construction of ξ_2. With divergence sensitivity, distribution equivalence on finite traces suffices for trace distribution equivalence. This completes the proof. □

Theorem 37 does not hold if we consider general schedulers:

Example 38 Let s_0 and s_1 be two states in Fig. 5. In Example 7 we have shown that $s_0 \not\equiv s_1$. It is also not hard to check that $s_0 \approx^\bullet s_1$. But we also notice that the schedulers giving rise to the trace distributions $\{\alpha : \frac{1}{3}, \beta : \frac{2}{3}\}$ and $\{\beta : \frac{1}{3}, \alpha : \frac{2}{3}\}$ are not partial information schedulers. Since at states s_2 and s_3 with the same enabled visible actions, the schedulers can choose transitions with different labels. By restricting to partial information schedulers we exclude these two distributions and can indeed show that $s_0 \equiv_{\mathcal{S}_P} s_1$. □

If looking at the effect of parallel composition, we can establish compositionality if distributed schedulers are considered:

Theorem 39 *For any states s_1, s_2, and s_3,*

$$s_1 \approx^\bullet_{\mathcal{S}_D} s_2 \text{ implies } s_1 \parallel_A s_3 \approx^\bullet_{\mathcal{S}_D} s_2 \parallel_A s_3.$$

Proof 1. In case that s_1 and s_2 contain no parallel operators, all schedulers of s_1 and s_2 are distributed schedulers according to Definition 9. Therefore $s_1 \approx^\bullet s_2$ implies $s_1 \approx^\bullet_{\mathcal{S}_D} s_2$ and vice versa.

2. Let $\mathcal{R} = \{(\mu_1 \parallel_A \mu_3, \mu_2 \parallel_A \mu_3) \mid \mu_1 \approx^\bullet_{\mathcal{S}_D} \mu_2\}$. It suffices to prove that $\mathcal{R}$ is a late distribution bisimulation with respect to $\mathcal{S}_D$. Let $(\mu_1 \parallel_A \mu_3)\ \mathcal{R}\ (\mu_2 \parallel_A \mu_3)$ and $\mu_1 \parallel_A \mu_3 \overset{\alpha}{\rightarrow}_{\xi_1} \nu$ for some $\xi_1 \in \mathcal{S}_D$. We have to show that there exists $\mu_2 \parallel_A \mu_3 \overset{\alpha}{\Rightarrow}_{\xi_2} \nu'$ for some $\xi_2 \in \mathcal{S}_D$ such that $\nu\ \mathcal{R}\ \nu'$. We distinguish several cases:

   (a) $\alpha \in Act \backslash A$:
   Since ξ_1 is a distributed scheduler, we have either *(i)* $\mu_1 \overset{\alpha}{\rightarrow}_{\xi_1} \nu_1$ such that $\nu \equiv \nu_1 \parallel_A \mu_3$, or *(ii)* $\mu_3 \overset{\alpha}{\rightarrow}_{\xi_1} \nu_3$ such that $\nu \equiv \mu_1 \parallel_A \nu_3$. We first consider case *(i)*. Since $\mu_1 \approx^\bullet_{\mathcal{S}_D} \mu_2$, there exists $\mu_2 \overset{\alpha}{\Rightarrow}_{\xi_2} \nu_2$ for some $\xi_2 \in \mathcal{S}_D$ such that $\nu_1 \approx^\bullet_{\mathcal{S}_D} \nu_2$, therefore there exists $\mu_2 \parallel_A \mu_3 \overset{\alpha}{\Rightarrow}_{\xi_2} \nu_2 \parallel_A \mu_3$. According to the definition of $\mathcal{R}$, we have $\nu \equiv (\nu_1 \parallel_A \mu_3)\ \mathcal{R}\ (\nu_2 \parallel_A \mu_3) \equiv \nu'$ as desired. The proof of case *(ii)* is similar and omitted here.

(b) $\alpha \in A$:
As before ξ_1 is a distributed scheduler, according to the definition of parallel operator, it must be the case that $\mu_1 \xrightarrow{\alpha}_{\xi_1} \nu_1$ and $\mu_3 \xrightarrow{\alpha}_{\xi_1} \nu_3$ such that $\nu \equiv \nu_1 \parallel_A \nu_3$. Since $\mu_1 \approx^{\bullet}_{\mathcal{S}_D} \mu_2$, there exists $\mu_2 \overset{\alpha}{\Rightarrow}_{\xi_2} \nu_2$ with $\nu_1 \approx^{\bullet}_{\mathcal{S}_D} \nu_2$. Hence there exists $\mu_2 \parallel_A \mu_3 \overset{\alpha}{\Rightarrow}_{\xi_2} \nu_2 \parallel_A \nu_3$ such that

$$\nu \equiv (\nu_1 \parallel_A \nu_3)\ \mathcal{R}\ (\nu_2 \parallel_A \nu_3) \equiv \nu'.$$

□

This is demonstrated by the following two examples:

Example 40 Let $s_0' \parallel_A r_0$ be a state as in Example 1, whose execution is depicted in Fig. 4b. Additionally, let r be a sequential state whose execution is same as $s_0' \parallel_A r_0$, such sequential state always exists (simply introducing a state for each node in Fig. 4b). By construction, we have $s_0' \parallel_A r_0 \approx^{\bullet} r$. However, if restricted to schedulers in $\mathcal{S}_D$, $s_0' \parallel_A r_0 \approx^{\bullet}_{\mathcal{S}_D} r$ does not hold. Since the scheduler inducing the execution of $s_0' \parallel_A r_0$ with probability 1 guessing correctly in Fig. 4b is not distributed, while the scheduler inducing the corresponding execution of r is distributed. Essentially, every possible scheduler of r is distributed because r is sequential.□

Example 41 Let s_0, s_0', and r_0 be the states in Fig. 3. We have shown in Example 22 that $s_0 \approx^{\bullet} s_0'$, but we have $s_0 \parallel_A r_0 \not\approx^{\bullet} s_0' \parallel_A r_0$ if general schedulers are considered: The composed PA of $s_0' \parallel_A r_0$, depicted in Fig. 4 allows a (non-distributed) scheduler that guesses correctly with probability 1, but this behaviour cannot be simulated by $s_0 \parallel_A r_0$, no matter how we schedule the transitions of $s_0 \parallel_A r_0$. For instance the maximal probability to reach states in the set $\{s_3 \parallel_A r_3, s_4 \parallel_A r_4\}$ from $s_0' \parallel_A r_0$ is 1, while the probability to reach this set from $s_0 \parallel_A r_0$ is always 0.5.

However, when restricting to distributed schedulers, we can show that both $s_0 \parallel_A r_0$ and $s_0' \parallel_A r_0$ reach this state set with probability 0.5, since schedulers of $s_0' \parallel_A r_0$ that guess correctly with a different probability are not distributed. The reason is that at states $s_5 \parallel_A r_0$ and $s_6 \parallel_A r_0$, r_0 makes different decisions by looking at s_5 and s_6, which should not happen in a distributed scheduler. □

As in the strong setting [38], by restricting to the set of schedulers in $\mathcal{S}_P \cap \mathcal{S}_D$, late distribution bisimulation is compositional and preserves trace distribution equivalence. Furthermore, late distribution bisimulation is the coarsest congruence satisfying the two properties with respect to schedulers in $\mathcal{S}_P \cap \mathcal{S}_D$.

Theorem 42 *Let $\mathcal{S} = \mathcal{S}_P \cap \mathcal{S}_D$. $s_1 \approx^{\bullet}_{\mathcal{S}} s_2$ iff $s_1 \equiv^{\mathrm{cgr}}_{\mathcal{S}} s_2$ for any s_1 and s_2, where $s_1 \equiv^{\mathrm{cgr}}_{\mathcal{S}} s_2$ iff $s_1 \equiv_{\mathcal{S}} s_2$ and $s_1 \parallel_A s_3 \equiv_{\mathcal{S}} s_2 \parallel_A s_3$ for any s_1, s_2, s_3, and A.*

We mention that schedulers in $\mathcal{S}_P \cap \mathcal{S}_D$ arise very natural in practice, for instance in decentralized multi-agent systems [3], where all agents are autonomous (corresponding to distributed schedulers) and states are partially observable (corresponding to partial information schedulers).

In [30] an algorithm was proposed to compute distribution-based bisimulation relations. We show briefly how the algorithm can be adapted to compute late distribution bisimulation. First observe that the relation $\approx^{\bullet}$ is linear, namely, $\mu_1 \approx^{\bullet} \nu_1$ and $\mu_2 \approx^{\bullet} \nu_2$ imply $(p \cdot \mu_1 + (1-p) \cdot \mu_2) \approx^{\bullet} (p \cdot \nu_1 + (1-p) \cdot \nu_2)$ for any $p \in [0, 1]$. By fixing an arbitrary order on the state space of a given PA, each distribution can be viewed as a vector in $[0, 1]^n$ with n being the number of states. Then for any s and α, it is easy to see that $\{\mu \mid s \xrightarrow{\alpha}_{\mathrm{c}} \mu\}$ constitutes a convex set. According to [6, Props. 3 and 4], every such convex set has a finite number of extreme points, which can be enumerated by restricting to Dirac memoryless schedulers. For deciding $\approx^{\bullet}$, it suffices to restrict to these finitely many extreme distributions. By doing so, all weak transitions can be handled in the same way as non-deterministic strong transitions in [30]. Not surprisingly, this will cause an exponential blow-up. We refer readers to [30] for more details of the remaining procedure.

6 Conclusion and future work

In this paper, we have proposed the notion of late distribution bisimilarity for PAs, which enjoys some interesting properties if restricted to two well-known subclasses of schedulers: partial information schedulers and distributed schedulers. Under partial information schedulers, late distribution bisimulation implies trace distribution equivalence, while under distributed schedulers, compositionality can be derived. Furthermore, if restricted to partial information distributed schedulers, late distribution bisimulation has been shown to be the coarsest relation which is compositional and preserves trace distribution equivalence.

As future work we intend to study reduction barbed congruences [9] under subclasses of schedulers, in order to pinpoint the characteristics of late distribution bisimilarity. The axiom system and logical characterization of $\approx^{\bullet}$ would be also interesting. The algorithm in [30] is exponential in the worst case. We will work out whether or not more efficient algorithms exist.

Acknowledgements Many thanks to the anonymous referees for their valuable suggestions on an early version of this paper. This work has been supported by the National Natural Science Foundation of China (Grant Nos. 61532019, 61472473), by the CAP project GZ1023, by the National 973 Program (Grant No. 2014CB340701) and by the CAS/SAFEA International Partnership Program for Creative Research Team. Part of this work was done while Lei Song was at Saarland University in Saarbrücken, Germany.

A Proofs

Theorem 37 *For any states s_1 and s_2, $s_1 \approx^{\bullet\mathrm{div}} s_2$ implies $s_1 \equiv_{\mathcal{S}_P} s_2$.*

Proof $s_1 \approx^{\bullet} s_2 \Longrightarrow s_1 \equiv_{\mathcal{S}_P} s_2$: let μ and ν be two distributions such that $Supp(\mu) = \{s_i\}_{i \in I}$ and $Supp(\nu) = \{r_j\}_{j \in J}$ where I and J are two finite sets of indexes. Let $\{\pi_i\}_{i \in I}$ and $\{\pi_j\}_{j \in J}$ be two sets of finite paths such that $\pi_i{\downarrow} = s_i$ and $\pi_j{\downarrow} = r_j$ for each $i \in I$ and $j \in J$. We prove a more general result: $\mu \approx^{\bullet} \nu$ implies for each partial information scheduler ξ_1, there exists a partial information scheduler ξ_2 such that

$$Pr^{\mu}_{\xi_1}(C_{\varsigma}, \{\pi_i\}_{i \in I}) = Pr^{\nu}_{\xi_2}(C_{\varsigma}, \{\pi_j\}_{j \in J})$$

for each finite trace ς, provided the following conditions hold:

1. $EA(s_{i_1}) = EA(s_{i_2})$ implies $EA(\pi_{i_1}) = EA(\pi_{i_2})$ for each $i_1, i_2 \in I$,
2. $EA(r_{j_1}) = EA(r_{j_2})$ implies $EA(\pi_{j_1}) = EA(\pi_{j_2})$ for each $j_1, j_2 \in J$,
3. $EA(s_i) = EA(r_j)$ implies $EA(\pi_i) = EA(\pi_j)$ for each $i \in I$ and $j \in J$.

Intuitively, $Pr^{\mu}_{\xi_1}(C_{\varsigma}, \{\pi_i\}_{i \in I})$ denotes the probability of C_{ς} starting from μ given execution history π_i for each $s_i \in Supp(\mu)$ under the guidance of scheduler ξ_1.

Since $EA(\pi_{i_1}) = EA(\pi_{i_2})$ if $EA(s_{i_1}) = EA(s_{i_2})$ for any $i_1, i_2 \in I$. If $\overrightarrow{\mu}$ and ξ_1 is a partial information scheduler,

$$Pr^{\mu}_{\xi_1}(C_\varsigma, \{\pi_i\}_{i\in I}) = Pr^{\mu}_{\xi_1}(C_\varsigma, EA(\pi_i))$$

for any $i \in I$. Fix a π_i in the sequel. Let $p = Pr^{\mu}_{\xi_1}(C_\varsigma, EA(\pi_i), n)$ with $n \geq 0$ and $\overrightarrow{\mu}$ be defined as follows, where all transitions are induced by ξ_1:

1. If $|\varsigma| > 0$ and $n = 0$, $p = 0$,
2. else if $|\varsigma| = 0$, $p = 1$,
3. else if $\mu \xhookrightarrow{\tau}_{\mathrm{c}} \sum_{k\in K} p_k \cdot \mu_k$ such that $\overrightarrow{\mu_k}$ for each $k \in K$, then
$$p = \sum_{k\in K} p_k \cdot Pr^{\mu_k}_{\xi_1}(C_\varsigma, EA(\pi_i \circ (\tau, s_k)), n-1)$$
for any $s_k \in Supp(\mu_k)$,
4. else if $\varsigma = \alpha\varsigma'$ and $\mu \xhookrightarrow{\alpha}_{\mathrm{c}} \sum_{k\in K} p_k \cdot \mu_k$ such that $\overrightarrow{\mu_k}$ for each $k \in K$, then
$$p = \sum_{k\in K} p_k \cdot Pr^{\mu_k}_{\xi_1}(C_{\varsigma'}, EA(\pi_i \circ (\alpha, s_k)), n-1)$$
for any $s_k \in Supp(\mu_k)$,
5. otherwise $p = 0$.

If $\neg\overrightarrow{\mu}$ and $\mu \equiv \sum_{k\in K} \mu_k$ such that $\overrightarrow{\mu_k}$ for each $k \in K$, then

$$Pr^{\mu}_{\xi_1}(C_\varsigma, \{\pi_i\}_{i\in I}, n) = \sum_{k\in K} p_k \cdot Pr^{\mu_k}_{\xi_1}(C_\varsigma, EA(\pi_k), n)$$

where $\pi_k = \pi_i$ for some $s_i \in Supp(\mu_k)$.

Now we prove by induction on n that for each partial information scheduler ξ_1, there exists a partial information scheduler ξ_2, such that

$$Pr^{\mu}_{\xi_1}(C_\varsigma, \{\pi_i\}_{i\in I}, n) \leq Pr^{\nu}_{\xi_2}(C_\varsigma, \{\pi_j\}_{j\in J})$$

for any $n \geq 0$ and ς.

First we assume that μ is transition consistent, which indicates $\overrightarrow{\nu}$ by Lemma 24. It is equivalent to show that for some $i \in I$ and $j \in J$,

$$Pr^{\mu}_{\xi_1}(C_\varsigma, EA(\pi_i), n) \leq Pr^{\nu}_{\xi_2}(C_\varsigma, EA(\pi_j)).$$

We distinguish the following cases:

1. $n = 0$ or $|\varsigma| = 0$. Trivial.
2. $n > 0$, $|\varsigma| > 0$, and there exists $\mu \xhookrightarrow{\tau}_{\mathrm{c}} \mu' = \sum_{k\in K} p_k \cdot \mu_k$ such that $\overrightarrow{\mu_k}$ for each $k \in K$, and
$$Pr^{\mu}_{\xi_1}(C_\varsigma, EA(\pi_i), n) = \sum_{k\in K} p_k \cdot Pr^{\mu_k}_{\xi_1}(C_\varsigma, EA(\pi_i \circ (\tau, s_k)), n-1)$$
for some $s_k \in Supp(\mu_k)$. Since $\mu \approx^\bullet \nu$, there exists
$$\nu \xRightarrow{\tau}_{\mathrm{c}} \nu' = \sum_{k\in K} p_k \cdot \nu_k$$
such that $\mu_k \approx^\bullet \nu_k$, thus $\overrightarrow{\nu_k}$ by Lemma 24 for each $k \in K$. Moreover, $EA(\mu_k) = EA(\nu_k)$. Let ξ_2 be a scheduler mimicking the transition $\nu \xRightarrow{\tau}_{\mathrm{c}} \nu'$. According to Definition 8 such

partial information scheduler ξ_2 always exists, since only τ transitions are involved. Since $\mu_k \approx^\bullet \nu_k$,

$$Pr_{\xi_2}^{\nu_k}(C_\varsigma, EA(\pi_i \circ (\tau, r_k))) \geq Pr_{\xi_1}^{\mu_k}(C_\varsigma, EA(\pi_i \circ (\tau, s_k)), n-1)$$

by induction hypothesis, where $r_k \in Supp(\nu_k)$ for each $k \in K$. Therefore

$$Pr_{\xi_2}^{\nu}(C_\varsigma, EA(\pi_i)) \geq Pr_{\xi_1}^{\mu}(C_\varsigma, EA(\pi_i), n).$$

3. $n > 0$, $\varsigma = \alpha\varsigma'$, and there exists $\mu \xhookrightarrow{\alpha}_{\text{c}} \mu' = \sum_{k\in K} p_k \cdot \mu_k$ such that $\overrightarrow{\mu_k}$ for each $k \in K$, and

$$Pr_{\xi_1}^{\mu}(C_\varsigma, EA(\pi_i), n) = \sum_{k\in K} p_k \cdot Pr_{\xi_1}^{\mu_k}(C_{\varsigma'}, EA(\pi_i \circ (\alpha, s_k)), n-1)$$

for any $i \in I$ and $s_k \in Supp(\mu_k)$. Since $\mu \approx^\bullet \nu$, there exists $\nu \overset{\alpha}{\Rightarrow}_{\text{c}} \sum_{k\in K} p_k \cdot \nu_k$ such that $\mu_k \approx^\bullet \nu_k$ for each $k \in K$. Let ξ_2 be the scheduler which mimic the weak transition of ν. The ξ_2 is guaranteed to be a partial information scheduler, since all states will perform a transition with label α. By induction we have:

$$Pr_{\xi_2}^{\nu_k}(C_{\varsigma'}, EA(\pi_i \circ (\alpha, r_k))) \geq Pr_{\xi_1}^{\mu_k}(C_{\varsigma'}, EA(\pi_i \circ (\alpha, s_k)), n-1)$$

where $r_k \in Supp(\nu_k)$ for each $k \in K$. Therefore

$$Pr_{\xi_2}^{\nu}(C_\varsigma, EA(\pi_i)) \geq Pr_{\xi_1}^{\mu}(C_\varsigma, EA(\pi_i), n).$$

In case $\varsigma = \beta\varsigma'$ such that $\beta \neq \alpha$, it is trivial, since $Pr_{\xi_1}^{\mu}(C_\varsigma, EA(\pi_i), n) = 0$.

Secondly, if $\neg\overrightarrow{\mu}$ and $\mu = \sum_{k\in K} \mu_k$ such that $\overrightarrow{\mu_k}$ for each $k \in K$, then

$$Pr_{\xi_1}^{\mu}(C_\varsigma, \{\pi_i\}_{i\in I}, n) = \sum_{k\in K} p_k \cdot Pr_{\xi_1}^{\mu_k}(C_\varsigma, EA(\pi_k), n)$$

where $\pi_k = \pi_i$ for any $s_i \in Supp(\mu_k)$. Since $\mu \approx^\bullet \nu$, there exists $\nu \overset{\tau}{\Rightarrow}_{\text{c}} \sum_{k\in K} p_k \cdot \nu_k$ such that $\mu_k \approx^\bullet \nu_k$ for each $k \in K$. Since $\overrightarrow{\mu_k}$ and we have proved that there exists ξ_2 such that

$$Pr_{\xi_2}^{\nu_k}(C_\varsigma, EA(\pi_k')) \geq Pr_{\xi_1}^{\mu_k}(C_\varsigma, EA(\pi_k), n)$$

for each $k \in K$. Again let ξ_2 mimic the transition of ν in a stepwise manner, we get

$$Pr_{\xi_2}^{\nu}(C_\varsigma, \{\pi_i'\}_{i\in J}) \geq Pr_{\xi_1}^{\mu}(C_\varsigma, \{\pi_i\}_{i\in I}, n)$$

as desired. Such ξ_2 is a partial information scheduler by construction, since only τ actions are involved.

Note that

$$Pr_{\xi_1}^{\mu}(C_\varsigma, \{\pi_i\}_{i\in I}) = \lim_{n\to\infty} Pr_{\xi_1}^{\mu}(C_\varsigma, \{\pi_i\}_{i\in I}, n).$$

If every trace has infinite many visible actions, then it is straightforward to see trace distribution equivalence. For those traces which only have finite visible actions, then there must exist an infinite sequence $\mu' \xrightarrow{\tau} \mu_1 \xrightarrow{\tau} \mu_2 \cdots$. Because of divergence sensitivity, ξ_2 can also simulate such infinite τ-transitions. So the distribution on finite traces are the same, and then it is easy to see trace distribution equivalence. □

In order to prove Theorem 42, we shall first introduce the following lemma:

Lemma 43 *Let $\mathcal{S} = \mathcal{S}_P \cap \mathcal{S}_D$, then $\mu_1 \approx^\bullet_{\mathcal{S}} \mu_2$ implies*

1. $\mu_1 \equiv_{\mathcal{S}} \mu_2$;
2. $\mu_1 \parallel_A \mu_3 \approx^{\bullet}_{\mathcal{S}} \mu_2 \parallel_A \mu_3$ *for any* μ_3.

Proof 1. Refer to the proof of Theorem 37.
2. The proof is similar as the proof of clause 2 of Theorem 39.

□

Theorem 42 *Let* $\mathcal{S} = \mathcal{S}_P \cap \mathcal{S}_D$. $s_1 \approx^{\bullet}_{\mathcal{S}} s_2$ *iff* $s_1 \equiv^{\mathrm{cgr}}_{\mathcal{S}} s_2$ *for any* s_1 *and* s_2, *where* $s_1 \equiv^{\mathrm{cgr}}_{\mathcal{S}} s_2$ *iff* $s_1 \equiv_{\mathcal{S}} s_2$ *and* $s_1 \parallel_A s_3 \equiv_{\mathcal{S}} s_2 \parallel_A s_3$ *for any* s_1, s_2, s_3, *and* A.

Proof – $\approx^{\bullet}_{\mathcal{S}} \Rightarrow \equiv^{\mathrm{cgr}}_{\mathcal{S}}$:

$$\mu_1 \approx^{\bullet}_{\mathcal{S}} \mu_2 \xrightarrow{\text{Lemma 43}} \mu_1 \equiv_{\mathcal{S}} \mu_2 \text{ and } \mu_1 \parallel_A \mu_3 \approx^{\bullet}_{\mathcal{S}} \mu_2 \parallel_A \mu_3$$
$$\xrightarrow{\text{Definition of } \equiv^{\mathrm{cgr}}_{\mathcal{S}}} \mu_1 \equiv^{\mathrm{cgr}}_{\mathcal{S}} \mu_2$$

– $\equiv^{\mathrm{cgr}}_{\mathcal{S}} \Rightarrow \approx^{\bullet}_{\mathcal{S}}$:
Let $\mathcal{R} = \{(\mu_1, \mu_2) \mid \mu_1 \equiv^{\mathrm{cgr}}_{\mathcal{S}} \mu_2\}$, we show that $\mathcal{R}$ is a late distribution bisimulation with respect to $\mathcal{S}$. Let $\mu_1 \mathcal{R} \mu_2$. We first assume that $\overrightarrow{\mu_1}$ and $\mu_1 \xrightarrow{\alpha}_{\xi_1} \mu_1'$ for some α and $\xi_1 \in \mathcal{S}$. We need to prove that there exists $\mu_2 \xRightarrow{\alpha}_{\xi_2} \mu_2'$ for some $\xi_2 \in \mathcal{S}$ such that $\mu_1' \mathcal{R} \mu_2'$. By contraposition. Assume $\mu_1' \not{\mathcal{R}} \mu_2'$, i.e., $\mu_1' \not\equiv^{\mathrm{cgr}}_{\mathcal{S}} \mu_2'$ for all μ_2'. We distinguish two cases as follows, where the main idea is to construct a distribution μ_3 with a proper set A such that $\mu_1 \parallel_A \mu_3 \not\equiv_{\mathcal{S}} \mu_2 \parallel_A \mu_3$, i.e., $\mu_1 \not{\mathcal{R}} \mu_2$.

1. $\mu_1' \not\equiv_{\mathcal{S}} \mu_2'$:
Given a set of actions A, we let $s' = A.s'$ denote a state which can only perform self loop transitions with labels in A. We can see that for any distribution μ such that $\overrightarrow{\mu}$, $\mu \parallel_A \delta_{s'}$ induces the same trace distribution as μ, where A contains all possible actions which can be performed by states in $Supp(\mu)$ and their successors.
Now let A contains all actions which can be performed by states in $Supp(\mu_1)$ and $Supp(\mu_2)$ and their successors. Let $s = \alpha.s'$ where s' is defined as above. Then for each $\xi_2 \in \mathcal{S}$, there exists $\xi_1 \in \mathcal{S}$ such that
$$Pr^{\xi_1}_{\mu_1 \parallel_{A'} \delta_s}(C_{\alpha\varsigma}) = Pr^{\xi_2}_{\mu_1' \parallel_{A'} \delta_{s'}}(C_{\varsigma}) = Pr^{\xi_2}_{\mu_1'}(C_{\varsigma}),$$
for each ς and vice versa, where $A' = A \cup \{\alpha\}$, similarly for $\mu_2 \parallel_{A'} \delta_s$. Since $\mu_1' \not\equiv_{\mathcal{S}} \mu_2'$, we conclude that $\mu_1 \parallel_{A'} \delta_s \not\equiv_{\mathcal{S}} \mu_2 \parallel_{A'} \delta_s$, which contradicts the assumption that $\mu_1 \equiv^{\mathrm{cgr}}_{\mathcal{S}} \mu_2$ (by letting $\mu_3 = \delta_s$).
2. $\mu_1' \equiv_{\mathcal{S}} \mu_2'$, but there exists μ_3' and A such that $\mu_1' \parallel_A \mu_3' \not\equiv_{\mathcal{S}} \mu_2' \parallel_A \mu_3'$:
Let s_i be a state such that the only transition of s_i is $s_i \xrightarrow{\alpha} \mu_i'$ with $\alpha \notin A$ a novel action and $i \in \{1, 2, 3\}$. It is easy to see that $\mu_1 \equiv^{\mathrm{cgr}}_{\mathcal{S}} \mu_2$ implies $\delta_{s_1} \equiv^{\mathrm{cgr}}_{\mathcal{S}} \delta_{s_2}$. By construction, for each $\xi_1 \in \mathcal{S}$, there exists $\xi_2 \in \mathcal{S}$ such that $Pr^{\xi_1}_{\mu_1' \parallel_A \mu_3'}(C_{\varsigma}) = Pr^{\xi_2}_{\delta_{s_1} \parallel_{A'} \delta_{s_3}}(C_{\alpha\varsigma})$ for each ς and vice versa, where $A' = A \cup \{\alpha\}$, similarly for $\mu_2' \parallel_A \mu_3'$ and $\delta_{s_1} \parallel_A \delta_{s_3}$. Therefore, $\delta_{s_1} \parallel_{A'} \delta_{s_3} \not\equiv_{\mathcal{S}} \delta_{s_2} \parallel_{A'} \delta_{s_3}$. Contradiction.
3. The cases when $\alpha = \tau$ can be proved in a similar way.

For now we have only considered case when μ_1 and μ_2 are transition consistent. Since all schedulers in $\mathcal{S}$ are partial information schedulers, in case that μ_1 is not transition consistent, we can always find a split $\mu_1 \equiv \sum_{i \in I} p_i \cdot \nu_i$ such that $\sum_{i \in I} p_i = 1$ and $\overrightarrow{\nu_i}$

for each $i \in I$. Moreover there exists $\mu_2 \xRightarrow{\tau} \sum_{i \in I} p_i \cdot \nu'_i$ such that $\nu_i \equiv_{\mathcal{S}}^{\text{cgr}} \nu'_i$ for each $i \in I$. Then we can apply the same arguments as when μ_1 is transition consistent.

□

References

1. Baier, C., Katoen, J.P., Hermanns, H., Wolf, V.: Comparative branching-time semantics for Markov chains. Inf. Comput. **200**(2), 149–214 (2005)
2. Bernardo, M., De Nicola, R., Loreti, M.: Relating strong behavioral equivalences for processes with nondeterminism and probabilities. Theor. Comput. Sci. **546**, 63–92 (2014)
3. Bernstein, D.S., Givan, R., Immerman, N., Zilberstein, S.: The complexity of decentralized control of Markov decision processes. Math. Oper. Res. **27**(4), 819–840 (2002)
4. Boudali, H., Crouzen, P., Stoelinga, M.: A rigorous, compositional, and extensible framework for dynamic fault tree analysis. IEEE Trans. Dependable Secur. Comput. **7**(2), 128–143 (2010)
5. Brengel, M.: Probabilistic weak transitions. Bachelor's thesis, Universität des Saarlandes, Saarbrücken, Germany (2013)
6. Cattani, S., Segala, R.: Decision algorithms for probabilistic bisimulation. In: CONCUR, pp. 371–385 (2002)
7. Chehaibar, G., Garavel, H., Mounier, L., Tawbi, N., Zulian, F.: Specification and verification of the PowerScale™ bus arbitration protocol: an industrial experiment with lotos. In: FORTE, pp. 435–450 (1996)
8. De Alfaro, L.: The verification of probabilistic systems under memoryless partial-information policies is hard. Technical report, DTIC document (1999)
9. Deng, Y., Hennessy, M.: On the semantics of Markov automata. Inf. Comput. **222**, 139–168 (2013)
10. Deng, Y., van Glabbeek, R., Hennessy, M., Morgan, C.: Testing finitary probabilistic processes. In: CONCUR, pp. 274–288 (2009)
11. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Weak bisimulation is sound and complete for pCTL*. Inf. Comput. **208**(2), 203–219 (2010). https://doi.org/10.1016/j.ic.2009.11.002
12. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Weak bisimulation is sound and complete for pCTL*. Inf. Comput. **208**(2), 203–219 (2010)
13. Doyen, L., Henzinger, T.A., Raskin, J.: Equivalence of labeled Markov chains. Int. J. Found. Comput. Sci. **19**(3), 549–563 (2008)
14. Eisentraut, C., Hermanns, H., Zhang, L.: Concurrency and composition in a stochastic world. In: CONCUR, pp. 21–39 (2010)
15. Eisentraut, C., Hermanns, H., Zhang, L.: On probabilistic automata in continuous time. In: LICS, pp. 342–351 (2010)
16. Eisentraut, C., Hermanns, H., Zhang, L.: On probabilistic automata in continuous time. In: Proceedings of the 25th Annual IEEE Symposium on Logic in Computer Science, LICS 2010, 11–14 July 2010, pp 342–351. IEEE Computer Society, Edinburgh (2010). https://doi.org/10.1109/LICS.2010.41
17. Eisentraut, C., Hermanns, H., Katoen, J., Zhang, L.: A semantics for every GSPN. In: PETRI NETS. Lecture Notes in Computer Science, vol. 7927, pp. 90–109. Springer (2013)
18. Eisentraut, C., Hermanns, H., Kraemer, J., Turrini, A., Zhang, L.: Deciding bisimilarities on distributions. In: QEST. Lecture Notes in Computer Science, vol. 8054, pp. 72–88. Springer, Berlin (2013)
19. Eisentraut, C., Godskesen, J.C., Hermanns, H., Song, L., Zhang, L.: Late weak bisimulation for Markov automata. CoRR arXiv:1202.4116 (2014)
20. Eisentraut, C.G.: Principles of Markov automata. Ph.D. thesis, Universität des Saarlandes, Saarbrücken, Germany (2017)
21. Feng, Y., Zhang, L.: When equivalence and bisimulation join forces in probabilistic automata. In: FM, Lecture Notes in Computer Science, vol. 8442, pp. 247–262. Springer (2014). https://doi.org/10.1007/978-3-319-06410-9_18
22. Feng, Y., Song, L., Zhang, L.: Distribution-based bisimulation and bisimulation metric in probabilistic automata. CoRR arXiv:1512.05027 (2015)
23. Giro, S., D'Argenio, P.R.: Quantitative model checking revisited: neither decidable nor approximable. In: FORMATS. Lecture Notes in Computer Science, vol. 4763, pp. 179–194. Springer (2007)
24. Groote, J.F., Jansen, D.N., Keiren, J.J.A., Wijs, A.: An $o(m \log n)$ algorithm for computing stuttering equivalence and branching bisimulation. ACM Trans. Comput. Log. (2017). https://doi.org/10.1145/3060140, article 13

25. Guck, D., Timmer, M., Hatefi, H., Ruijters, E., Stoelinga, M.: Modelling and analysis of Markov reward automata. In: ATVA. Lecture Notes in Computer Science, vol. 8837, pp. 168–184. Springer (2014)
26. Halmos, P.R.: Measure Theory, vol. 1950. Springer, New York (1974)
27. He, F., Gao, X., Wang, B., Zhang, L.: Leveraging weighted automata in compositional reasoning about concurrent probabilistic systems. In: POPL, pp. 503–514. ACM (2015)
28. Hennessy, M.: Exploring probabilistic bisimulations. Form. Asp. Comput. **24**(4–6), 749–768 (2012)
29. Hermanns, H.: Interactive Markov Chains: And the Quest for Quantified Quality. Springer, Berlin (2002)
30. Hermanns, H., Krcál, J., Kretínský, J.: Probabilistic bisimulation: naturally on distributions. In: CONCUR. Lecture Notes in Computer Science, vol. 8704. Springer (2014)
31. Honda, K., Tokoro, M.: On asynchronous communication semantics. In: Object-Based Concurrent Computing, pp. 21–51 (1991)
32. Kwiatkowska, M., Norman, G., Parker, D., Qu, H.: Assume-guarantee verification for probabilistic systems. In: TACAS, Lecture Notes in Computer Science, vol. 6015, pp. 23–37. Springer (2010)
33. Philippou, A., Lee, I., Sokolsky, O.: Weak bisimulation for probabilistic systems. In: CONCUR, pp. 334–349 (2000)
34. Rudin, W.: Real and Complex Analysis. Tata McGraw-Hill Education, Delhi (2006)
35. Schuster, J., Siegle, M.: Markov automata: deciding weak bisimulation by means of non-navely vanishing states. Inf. Comput. **237**, 151–173 (2014)
36. Segala, R.: A compositional trace-based semantics for probabilistic automata. In: CONCUR. Lecture Notes in Computer Science, vol. 962, pp. 234–248. Springer (1995)
37. Segala, R.: Modeling and verification of randomized distributed realtime systems. Ph.D. thesis, MIT (1995)
38. Song, L., Feng, Y., Zhang, L.: Decentralized bisimulation for multiagent systems. In: AAMAS'15: Autonomous Agents and Multiagent Systems, pp. 209–217. ACM, New York (2015)
39. Timmer, M., van de Pol, J., Stoelinga, M.: Confluence reduction for Markov automata. In: FORMATS. Lecture Notes in Computer Science, vol. 8053, pp. 243–257. Springer (2013)
40. van Glabbeek, R.J., Weijland, P.W.: Branching time and abstraction in bisimulation semantics. J. ACM **43**(3), 555–600 (1996). https://doi.org/10.1145/233551.233556
41. Yang, P., Jansen, D.N., Zhang, L.: Distribution-based bisimulation for labelled Markov processes. In: Abate, A., Geeraerts, G. (eds.) Formal Modeling and Analysis of Timed Systems: FORMATS. Lecture Notes in Computer Science, vol. 10419, pp. 170–186. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-65765-3_10