



**Project no.:** PIRSES-GA-2011-295261  
**Project full title:** Mobility between Europe and Argentina applying Logics to Systems  
**Project Acronym:** MEALS  
**Deliverable no.:** 1.3 / 2  
**Title of Deliverable:** Probabilistic Bisimulation for Realistic Schedulers

<b>Contractual Date of Delivery to the CEC:</b>	<b>30-Sep-2015</b>
<b>Actual Date of Delivery to the CEC:</b>	<b>30-Sep-2015</b>
<b>Organisation name of lead contractor for this deliverable:</b>	<b>UNC</b>
<b>Author(s):</b>	<b>Christian Eisentraut, Jens Chr. Godskesen, Holger Hermanns, Lei Song, Lijun Zhang</b>
<b>Participants(s):</b>	<b>SAU</b>
<b>Work package contributing to the deliverable:</b>	<b>WP1</b>
<b>Nature:</b>	<b>R</b>
<b>Dissemination Level:</b>	<b>Public</b>
<b>Total number of pages:</b>	<b>19</b>
<b>Start date of project:</b>	<b>1 Oct. 2011    Duration: 48 month</b>

**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.

**Note:**

This deliverable is has been published in *FM 2015: Formal Methods*, LNCS 9109, pp 248-264, Springer, 2015.

*This project has received funding from the European Union Seventh Framework Programme (FP7 2007-2013) under Grant Agreement Nr. 295261.*

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Preliminaries</b>	<b>5</b>
2.1	Probabilistic Automata . . . . .	5
2.2	Trace Distribution Equivalence . . . . .	7
2.3	Partial Information and Distributed Schedulers . . . . .	7
<b>3</b>	<b>Weak Bisimilarities for Probabilistic Automata</b>	<b>8</b>
3.1	Weak Distribution Bisimulation . . . . .	9
3.2	Late Weak Bisimulation . . . . .	10
<b>4</b>	<b>Properties of Late Distribution Bisimilarity</b>	<b>14</b>
<b>5</b>	<b>Conclusion and Future Work</b>	<b>16</b>
	<b>Bibliography</b>	<b>17</b>
	<b>MEALS Partner Abbreviations</b>	<b>19</b>

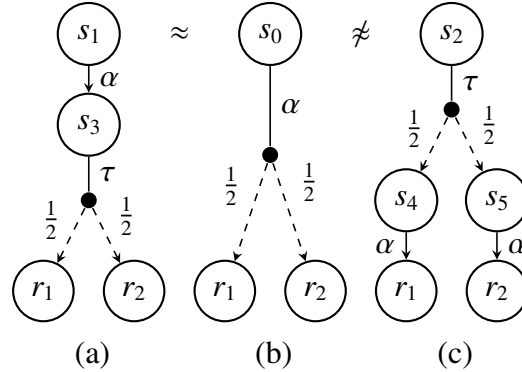


Figure 1: Distinguishing probabilistic automata

## 1 Introduction

Compositional theories have been an important technique to deal with complex stochastic systems effectively. Their potential ranges from compositional minimization [6, 4] approaches to component based verification [26, 21]. Due to their expressiveness, Markov automata have attracted many attentions [33, 13, 19], since they were introduced [16]. Markov automata are a compositional behavioral model for continuous time stochastic and non-deterministic systems [15, 16] subsuming interactive Markov chains (IMCs) [23] and probabilistic automata (PAs) [31] (and hence also Markov decision processes and Markov chains).

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 sub-probability 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. Nevertheless all the results we establish carry over to Markov automata.

On probabilistic automata, weak distribution bisimilarity is strictly coarser than weak bisimilarity, and is the coarsest congruence preserving trace distribution equivalence [8]. More precisely, it is the coarsest reduction-closed barbed congruence [25] with respect to parallel composition. Decision algorithms for weak distribution bisimilarity have also been proposed [14, 29].

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  $\alpha$  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. Surprisingly, 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 [31, 18], the general scheduler in the parallel composition is too powerful: the decision of one component may depend on the history of other components. This is especially not desired for partially observable systems, such as multi-agent systems or distributed systems [3, 32]. 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 the definition of a coarser notion of equality equating them.

In this paper, 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, 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 [8], 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 [7] and distributed schedulers [18]. 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.

**Related Work.** Many variants of bisimulations have been studied for different stochastic models, for instance Markov chains [1], interactive Markov chains [23], probabilistic automata [27, 31, 2], and alternating automata [10]. 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 PAs and IMCs. In [16], a novel *distribution-based* weak bisimulation has been proposed: it is weaker than the state-based weak bisimulation in [31], and if restricted to continuous-time Markov chains, generates an equivalence established in the Petri net community [13]. Later, another weak bisimulation has been investigated in [8], which is essentially the same as [16]. 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 [16], several *distribution-based* strong bisimulations have been proposed. In [22], it is shown that, the strong version of the relation in [16] coincides with the lifting of the classical state-based strong bisimulations. Recently, three different distribution-based strong bisimulations have been defined: paper [17] defines bisimulation relations and metrics which extend the well-known language equivalence [11] of labelled Markov chain; another definition in [24] applies to discrete systems as well as to systems with uncountable state and action spaces; in [32], 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 [32] for a detailed discussion. The current paper extends the decentralized strong bisimulation in [32] 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 [16], and as in [32], we show that it is compositional with respect to partial information and distributed schedulers.

**Organization of the Paper** Section 2 recalls some notations used in the paper. Late distribution bisimulation is proposed and discussed in Section 3, and its properties are established in Section 4 under realistic schedulers. Section 5 concludes the paper. A discussion why all results established in this paper directly carry over to Markov automata can be found in [12].

## 2 Preliminaries

Let  $S$  be a finite set of states ranged over by  $r, s, \dots$ . A *distribution* is a function  $\mu : S \rightarrow [0, 1]$  satisfying  $\mu(S) = \sum_{s \in S} \mu(s) = 1$ . Let  $Dist(S)$  to denote the set of all distributions, ranged over by  $\mu, \nu, \gamma, \dots$ . Define  $Supp(\mu) = \{s \mid \mu(s) > 0\}$  as the support set of  $\mu$ . If  $\mu(s) = 1$ , then  $\mu$  is called a *Dirac* distribution, written as  $\delta_s$ . Let  $|\mu| = \mu(S)$  denote the size of the distribution  $\mu$ . Given a real number  $x$ ,  $x \cdot \mu$  is the distribution 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 distribution such that  $(\mu - s)(s) = 0$  and  $(\mu - s)(r) = \mu(r)$  with  $s \neq r$ . 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 distribution  $\mu$ . For instance,  $\{s_1 : 0.4, s_2 : 0.6\}$  denotes a distribution  $\mu$  such that  $\mu(s_1) = 0.4$  and  $\mu(s_2) = 0.6$ .

### 2.1 Probabilistic Automata

Initially introduced in [31], *probabilistic automata* (PAs) have been popular models for systems with both non-deterministic choices and probabilistic dynamics. Below we give their formal definition.

**Definition 1.** A PA  $\mathcal{P}$  is a tuple  $(S, Act_\tau, \rightarrow, \bar{s})$  where

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

Let  $\alpha, \beta, \gamma, \dots$  range over the actions in  $Act_\tau$ . We write  $s \xrightarrow{\alpha} \mu$  if  $(s, \alpha, \mu) \in \rightarrow$ . A *path* is a finite or infinite alternative sequence  $\pi = s_0, \alpha_0, s_1, \alpha_1, s_2, \dots$  of states and actions, such that for each  $i \geq 0$  there exists a distribution  $\mu$  with  $s_i \xrightarrow{\alpha_i} \mu$  and  $\mu(s_{i+1}) > 0$ . Some notations are defined as follows:  $|\pi|$  denotes the length of  $\pi$ , i.e., the number of states on  $\pi$ , while  $\pi \downarrow$  is the last state of  $\pi$ , provided  $\pi$  is finite;  $\pi[i] = s_i$  with  $i \geq 0$  is the  $(i + 1)$ -th state on  $\pi$  if it exists;  $\pi[0..i] = s_0, \alpha_0, s_1, \alpha_1, \dots, s_i$  is the prefix of  $\pi$  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 non-deterministic choices in PAs, a probability measure 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 history execution. Formally,

**Definition 2.** A scheduler is a function

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

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

In this paper, we are restricted to schedulers satisfying the following condition: For any  $\pi \in Paths^*$ ,  $\xi(\pi)(\alpha, \mu) > 0$  and  $\xi(\pi)(\beta, \nu) > 0$  imply  $\alpha = \beta$ . In other words,  $\xi$  always chooses transitions with the same label at each step. This class of schedulers suffices for our purpose.

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

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

Given a starting state  $s$ , a scheduler  $\xi$ , and a finite path  $\pi = s_0, \alpha_0, s_1, \alpha_1, \dots, s_k$ , the measure  $Pr_{\xi, s}$  of a cone  $C_\pi$  is defined inductively as:

- $Pr_{\xi, s}(C_\pi) = 0$  if  $s \neq s_0$ ;
- $Pr_{\xi, s}(C_\pi) = 1$  if  $s = s_0$  and  $k = 0$ ;
- otherwise  $Pr_{\xi, s}(C_\pi) =$

$$Pr_{\xi, s}(C_{\pi[0..k-1]}) \cdot \left( \sum_{(s_{k-1}, \alpha_{k-1}, \mu) \in \rightarrow} \xi(\pi[0..k-1])(\alpha_{k-1}, \mu) \cdot \mu(s_k) \right).$$

Let  $\mathcal{B}$  be the smallest algebra that contains all the cones and is closed under complement and countable unions. By standard measure theory [20, 28], this algebra is a  $\sigma$ -algebra and all its elements are measurable sets of paths. Moreover,  $Pr_{\pi, s}$  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 [31].

**Definition 3.** 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  $A \subseteq Act$ , then  $\mathcal{P}_1 \parallel_A \mathcal{P}_2 = (S, Act_\tau, \rightarrow, \bar{s})$  such that

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

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

## 2.2 Trace Distribution Equivalence

In this subsection we introduce the notion of trace distribution equivalence [30] adapted to our setting with internal actions. Let  $\zeta \in Act^*$  denote a finite trace of a PA  $\mathcal{P}$ , which is an ordered sequence of visible actions. Each trace  $\zeta$  induces a cylinder  $C_\zeta$  which is defined as follows:

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

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

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

Since  $C_\zeta$  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 4.** 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  $Pr_{s_1}^{\xi_1}(C_\zeta) = Pr_{s_2}^{\xi_2}(C_\zeta)$  for each finite trace  $\zeta$  and vice versa. If  $\mathcal{S}$  is the set of all schedulers, we simply write  $\equiv$ .

Different from [30, 32], we abstract internal transitions when defining traces of a path. Therefore, the definition above is also a weaker version of the corresponding definition in [30, 32].

## 2.3 Partial Information and Distributed Schedulers

In this subsection we define two prominent sub-classes of schedulers, where the power of schedulers are limited. We first introduce some notations. Let  $EA : S \mapsto 2^{Act}$  such that

$$EA(s) = \{\alpha \in Act \mid \exists \mu. s \xrightarrow{\alpha} \mu\},$$

that is, the function  $EA$  returns the set of visible actions that a state is able to perform, possibly after some internal transitions. We generalize this function to paths as follows:  $EA(\pi) =$

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

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  $\pi$  to their corresponding enabled actions. Whenever an invisible action does not change the enabled actions, this 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 ( $EA(\pi' \downarrow) \neq EA(s)$ ). We are now ready to define the *partial information schedulers* [7] as follows:

**Definition 5.** A scheduler  $\xi$  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  $\xi(\pi_1) = (\tau, \mu)$  or  $\xi(\pi_2) = (\tau, \mu)$  for some  $\mu$ ,
- or  $\xi(\pi_1) = (\alpha, \mu)$  and  $\xi(\pi_2) = (\alpha, \nu)$  for some  $\mu, \nu$  such that  $\alpha \neq \tau$ .

$\xi$  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 states via different enabled visible actions. 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  $\pi_1$  and  $\pi_2$  with  $EA(\pi_1) = EA(\pi_2)$ , a partial information scheduler either chooses a transition labelled with  $\tau$  action for  $\pi_i$  ( $i = 1, 2$ ), or it chooses transitions labelled with the same visible actions for both  $\pi_1$  and  $\pi_2$ . Partial information schedulers do not impose any restriction on the execution of  $\tau$  transitions, instead they can be performed spontaneously.

When composing parallel systems, general schedulers defined in Definition 2 allow one component to make decisions based on full information of other components. This may be unrealistically powerful as argued in [18]. To deal with this, another important sub-class of schedulers called *distributed schedulers* has been introduced [18]. The main idea is to assume that all parallel components run in autonomous and can only 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 beforehand. We omit the formal definition of distributed schedulers, which can be found in [18] or [32]. In the sequel we let  $\mathcal{S}_D$  denote the set of all distributed schedulers.

### 3 Weak Bisimilarities for Probabilistic Automata

In this section, we first introduce weak distribution bisimulation, which is a variant of weak bisimulation defined in [8], and then define late distribution bisimulation, which is strictly coarser than weak distribution bisimulation.



### 3.1 Weak Distribution Bisimulation

As usual, a standard weak transition relation is needed in the definitions of bisimulation that allows one to abstract internal actions. Intuitively,  $s \xRightarrow{\alpha} \mu$  denotes that a distribution  $\mu$  is reached from  $s$  by a  $\alpha$ -transition, which may be preceded and followed by an arbitrary sequence of internal transitions. Formally, we define them as derivations [9] for PAs. In the following, let  $\mu \xrightarrow{\alpha} \mu'$  iff there exists a transition  $s \xrightarrow{\alpha} \mu_s$  for each  $s \in \text{Supp}(\mu)$  such that  $\mu' = \sum_{s \in \text{Supp}(\mu)} \mu(s) \cdot \mu_s$ . Then,  $s \xRightarrow{\tau} \mu$  iff there exists

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

where  $\mu = \sum_{i \geq 0} \mu_i^{\times}$ . We write  $s \xRightarrow{\alpha} \mu$  iff there exists  $s \xRightarrow{\tau} \xrightarrow{\alpha} \xRightarrow{\tau} \mu$ .

Given a transition relation  $\rightsquigarrow \subseteq S \times \text{Act}_{\tau} \times \text{Dist}(S)$ , we let  $s \rightsquigarrow_c^{\alpha} \mu$  iff there exists a finite number of real numbers  $w_i > 0$ , and transitions  $s \xrightarrow{\alpha} \mu_i$  such that  $\sum_i w_i = 1$ , and  $\sum_i w_i \cdot \mu_i = \mu$ . We call  $\rightsquigarrow_c$  *combined transitions* (of  $\rightsquigarrow$ ). In general, we lift a transition relation  $\rightsquigarrow \subseteq S \times \text{Act}_{\tau} \times \text{Dist}(S)$  over states to a transition relation  $\text{Dist}(S) \times \text{Act}_{\tau} \times \text{Dist}(S)$  over distributions by letting  $\mu \rightsquigarrow^{\alpha} \mu'$  iff there exists a transition  $s \xrightarrow{\alpha} \mu_s$  for each  $s \in \text{Supp}(\mu)$  such that  $\mu' = \sum_{s \in \text{Supp}(\mu)} \mu(s) \cdot \mu_s$ .

**Definition 6.**  $\mathcal{R} \subseteq \text{Dist}(S) \times \text{Dist}(S)$  is a *weak distribution bisimulation* iff  $\mu \mathcal{R} \nu$  implies:

1. whenever  $\mu \xrightarrow{\alpha}_c \mu'$ , there exists a  $\nu \xrightarrow{\alpha}_c \nu'$  such that  $\mu' \mathcal{R} \nu'$ ;
2. whenever  $\mu = \sum_{0 \leq i \leq n} p_i \cdot \mu_i$ , there exists a  $\nu \xrightarrow{\tau}_c \sum_{0 \leq i \leq n} p_i \cdot \nu_i$  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  $\nu$ .

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

Clause 1 is standard. Clause 2 says that no matter how we split  $\mu$ , there always exists a splitting of  $\nu$  probably after internal transitions to simulate the splitting of  $\mu$ . Definition 6 is slightly different from Definition 5 in [8], where clause 2 is missing and clause 1 is replaced by: whenever  $\mu \xrightarrow{\alpha}_c \sum_{0 \leq i \leq n} p_i \cdot \mu_i$ , there exists  $\nu \xrightarrow{\alpha}_c \sum_{0 \leq i \leq n} p_i \cdot \nu_i$  such that  $\mu_i \mathcal{R} \nu_i$  for each  $0 \leq i \leq n$ . Essentially, this condition subsumes clause 2, since  $\mu = \sum_{0 \leq i \leq n} p_i \cdot \mu_i$  implies  $\mu \xrightarrow{\tau}_c \sum_{0 \leq i \leq n} p_i \cdot \mu_i$ . As we prove in the following lemma, both definitions induce the same equivalence relation on PAs.

**Lemma 1.** *Let  $\mathcal{P} = (S, \text{Act}_{\tau}, \rightarrow, \bar{s})$  be a PA.  $\mathcal{R} \subseteq \text{Dist}(S) \times \text{Dist}(S)$  is a weak distribution bisimulation iff  $\mu \mathcal{R} \nu$  implies that*

1. whenever  $\mu \xrightarrow{\alpha}_c \mu'$ , there exists  $\nu \xrightarrow{\alpha}_c \nu'$  such that  $\mu' \mathcal{R} \nu'$ ,

2. whenever  $\mu = \sum_{0 \leq i \leq n} p_i \cdot \mu_i$ , there exists  $\nu \xRightarrow{\tau}_c \sum_{0 \leq i \leq n} p_i \cdot \nu_i$  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  $\nu$ .

*Proof.* Let  $\mathcal{R} \subseteq \text{Dist}(S) \times \text{Dist}(S)$ . If  $\mathcal{R}$  is a weak distribution bisimulation by Lemma 1, then trivially we can show that  $\mathcal{R}$  is also a weak distribution bisimulation by Definition 6, since  $\rightarrow_c \subset \xRightarrow{\alpha}_c$ . In the sequel, we let  $\mathcal{R}$  be a weak distribution bisimulation by Definition 6 and we show that  $\mathcal{R}$  also satisfies conditions of Lemma 1. Let  $\mu \mathcal{R} \nu$ . It suffices to show that whenever  $\mu \xRightarrow{\alpha}_c \mu'$ , there exists a  $\nu \xRightarrow{\alpha}_c \nu'$  such that  $\mu' \mathcal{R} \nu'$ ,

Assume  $\alpha = \tau$ . According to the definition of derivations (P. 9),  $\mu \xRightarrow{\tau}_c \mu'$  iff there exists

$$\begin{aligned} \mu &= \mu_0^{\rightarrow} + \mu_0^{\times}, \\ \mu_0^{\rightarrow} &\xrightarrow{\tau}_c \mu_1^{\rightarrow} + \mu_1^{\times}, \\ \mu_1^{\rightarrow} &\xrightarrow{\tau}_c \mu_2^{\rightarrow} + \mu_2^{\times}, \\ &\vdots \end{aligned} \tag{4}$$

such that  $\mu' \equiv \sum_{i \geq 0} \mu_i^{\times}$ . By Definition 6,  $\nu$  can simulate such a derivation at each step, namely, there exists

$$\begin{aligned} \nu &\xRightarrow{\tau}_c \nu_0^{\rightarrow} + \nu_0^{\times}, \\ \nu_0^{\rightarrow} &\xRightarrow{\tau}_c \nu_1^{\rightarrow} + \nu_1^{\times}, \\ \nu_1^{\rightarrow} &\xRightarrow{\tau}_c \nu_2^{\rightarrow} + \nu_2^{\times}, \\ &\vdots \end{aligned} \tag{5}$$

such that  $\mu_i^{\rightarrow} \mathcal{R} \nu_i^{\rightarrow}$  and  $\mu_i^{\times} \mathcal{R} \nu_i^{\times}$  for each  $i \geq 0$ . Note  $\mathcal{R}$  satisfies *infinite linearity*, which can be proved in a similar way as [8, Thm. A.6]. Therefore,  $(\sum_{i \geq 0} \mu_i^{\times}) \mathcal{R} (\sum_{i \geq 0} \nu_i^{\times})$ . Since  $\xRightarrow{\alpha}_c$  is transitive [8, Thm. A.4], there exists  $\nu \xRightarrow{\tau}_c \nu'$  such that  $\mu' \mathcal{R} \nu'$  as desired.

In case  $\mu \xRightarrow{\alpha}_c \mu'$  with  $\alpha \neq \tau$ , we have  $\mu \xRightarrow{\tau}_c \mu_1 \xrightarrow{\alpha}_c \mu_2 \xRightarrow{\tau}_c \mu'$ . As shown above, there exists  $\nu \xRightarrow{\tau}_c \nu_1$  such that  $\mu_1 \mathcal{R} \nu_1$ , which indicates that there exists  $\nu_1 \xrightarrow{\alpha}_c \nu_2$  such that  $\mu_2 \mathcal{R} \nu_2$  by Definition 6, which indicates that there exists  $\nu_2 \xRightarrow{\tau}_c \nu'$  such that  $\mu' \mathcal{R} \nu'$ . This completes the proof.  $\square$

The above lemma implies the transitivity of the weak distribution bisimulation, and will be useful for establishing different bisimulation relations.

## 3.2 Late Weak Bisimulation

Clause 2 in Definition 6 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  $\mu$  must be matched by  $\nu$  (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 will fix this in the definition of late distribution bisimulation. Before that, we still need some notations.

**Definition 7.** A distribution  $\mu$  is *transition consistent*, written as  $\vec{\mu}$ , if for any  $s \in Supp(\mu)$  and  $\alpha \neq \tau$ ,  $s \xrightarrow{\alpha} \gamma$  for some  $\gamma$  implies  $\mu \xrightarrow{\alpha} \gamma'$  for some  $\gamma'$ .

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  $\mu \xrightarrow{\alpha}$  whenever  $s \xrightarrow{\alpha}$  for some state  $s \in Supp(\mu)$ . In contrast, when a distribution  $\mu$  is not transition consistent, there must be a weak  $\alpha$  transition of some state in  $Supp(\mu)$  being *blocked*. In the sequel, when we adopt the notion of blocked states accordingly for non-weak transition relations, also  $\tau$  transitions can be blocked.

We now introduce  $\hookrightarrow$ , an alternative lifting of transitions of states to transitions of distributions that differs from the standard definition used in [16, 8]. There, a distribution is able to perform a transition labelled with  $\alpha$  *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  $\mu$  may at most perform one transition.

**Definition 8.**  $\mu \xrightarrow{\alpha} \mu'$  iff

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

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

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

*Example 1.* Let  $\mu = \{s_1 : 0.4, s_2 : 0.6\}$  such that  $s_1 \xrightarrow{\tau} \delta_{s'_1} \xrightarrow{\alpha} \mu_1$ ,  $s_1 \xrightarrow{\beta} \mu_2$ ,  $s_2 \xrightarrow{\alpha} \mu_3$ , and  $s_2 \xrightarrow{\beta} \mu_4$ , where  $\alpha \neq \beta$  are visible actions. According to clause 1 of Definition 8, we will have  $\mu \xrightarrow{\beta} (0.4 \cdot \mu_2 + 0.6 \cdot \mu_4)$ . Without clause 2, this would be the only transition of  $\mu$ , since the  $\tau$  transition of  $s_1$  and the  $\alpha$  transition of  $s_2$  will be blocked by each other, as  $s_1$  and  $s_2$  cannot perform transitions with labels  $\tau$  and  $\alpha$  at the same time.

Note that the  $\alpha$  transition is blocked by the  $\tau$  transition of  $s_1$ , so according to clause 2 of Definition 8, we in addition have

$$\mu \xrightarrow{\tau} (0.4 \cdot \delta_{s'_1} + 0.6 \cdot \delta_{s_2}) \xrightarrow{\alpha} (0.4 \cdot \mu_1 + 0.6 \cdot \mu_3).$$

Note that in clause 1 of Definition 6,  $\rightarrow$  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 this example, we could let  $\mu$  be split into  $0.4 \cdot \delta_{s_1} + 0.6 \cdot \delta_{s_2}$ , such that no transition is blocked in the resulting distributions.  $\square$

**Definition 9.**  $\mathcal{R} \subseteq \text{Dist}(S) \times \text{Dist}(S)$  is a *late distribution bisimulation* iff  $\mu \mathcal{R} \nu$  implies:

1. whenever  $\mu \xrightarrow{\alpha}_c \mu'$ , there exists a  $\nu \xrightarrow{\alpha}_c \nu'$  such that  $\mu' \mathcal{R} \nu'$ ;
2. if not  $\vec{\mu}$ , then there exists  $\mu = \sum_{0 \leq i \leq n} p_i \cdot \mu_i$  and  $\nu \xrightarrow{\tau}_c \sum_{0 \leq i \leq n} p_i \cdot \nu_i$  such that  $\vec{\mu}_i$  and  $\mu_i \mathcal{R} \nu_i$  for each  $0 \leq i \leq n$  where  $\sum_{0 \leq i \leq n} p_i = 1$ ;
3. symmetrically for  $\nu$ .

We say that  $\mu$  and  $\nu$  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 6 by the use of  $\xrightarrow{\alpha}_c$ . It is straightforward to show that  $\xrightarrow{\alpha}_c$  can also be used in Definition 6 without changing the resulting bisimilarity. However, in Definition 9, using  $\rightarrow$  instead of  $\xrightarrow{\alpha}_c$  will lead to a finer relation. The key difference between Definition 6 and 9, however, is clause 2. As we mentioned, in Definition 6, any split of  $\mu$  should be matched by  $\nu$ , while in Definition 9, we require to split  $\mu$  only if it is not transition consistent. Additionally, the resulting distributions  $\mu_i$  must be transition consistent as well. We do not need to require that  $\nu_i$  is transition consistent, as we will show later that  $\vec{\mu}_i$  and  $\mu_i \mathcal{R} \nu_i$  implies  $\vec{\nu}_i$ . According to Definition 7, splittings to transition consistent distributions ensure that all possible transitions will be considered eventually, as no transition of individual states 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 [7], 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 the transition consistent condition is broken. This explains the name “late distribution bisimulation”. Further, if restricting to models without internal action  $\tau$ , our notion of late distribution bisimulate agrees with the decentralized bisimulations in [32].

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

**Theorem 1.** 1.  $\approx^\bullet$  is an equivalence relation;

2.  $\approx \subset \approx^\bullet$ .

Before proving Theorem 1, we shall introduce two lemmas. The lemma below resembles Lemma 1, which can be proved similarly as Lemma 1.

**Lemma 2.** Let  $\mathcal{P} = (S, \text{Act}_\tau, \rightarrow, \bar{s})$  be a PA.  $\mathcal{R} \subseteq \text{Dist}(S) \times \text{Dist}(S)$  is a weak distribution bisimulation iff  $\mu \mathcal{R} \nu$  implies that

1. whenever  $\mu \xrightarrow{\alpha}_c \mu'$ , there exists  $\nu \xrightarrow{\alpha}_c \nu'$  such that  $\mu' \mathcal{R} \nu'$ ,
2. if not  $\vec{\mu}$ , then there exists  $\mu = \sum_{0 \leq i \leq n} p_i \cdot \mu_i$  and  $\nu \xrightarrow{\tau}_c \sum_{0 \leq i \leq n} p_i \cdot \nu_i$  such that  $\vec{\mu}_i$  and  $\mu_i \mathcal{R} \nu_i$  for each  $0 \leq i \leq n$  where  $\sum_{0 \leq i \leq n} p_i = 1$ ;

### 3. symmetrically for $\nu$ .

*Proof.* The proof is almost the same as Lemma 1 with two exceptions related to the transition consistent requirement:

- In Eq. 4 and 5, derivations should respect the transition consistent requirement, namely, states with the same set of enable actions should be in the support of either  $\mu_i^{\rightarrow}$  or  $\mu_i^{\times}$ , similarly for  $\nu_i^{\rightarrow}$  and  $\nu_i^{\times}$ .
- The infinite linearity of late distribution bisimulation can be proved as follows: Let

$$\mathcal{R} = \{(\sum_{i \geq 0} p_i \cdot \mu_i, \sum_{i \geq 0} p_i \cdot \nu_i) \mid \sum_{i \geq 0} p_i = 1 \wedge \forall i \geq 0, \mu_i \approx^{\bullet} \nu_i\}.$$

We prove that  $\mathcal{R}$  is a late distribution bisimulation. Let  $\mu \mathcal{R} \nu$ . Suppose  $\mu \xrightarrow{\alpha}_c \mu'$ , then for all  $i \geq 0$ , there exists  $\mu_i \xrightarrow{\alpha}_c \mu'_i$  such that  $\mu' \equiv \sum_{i \geq 0} p_i \cdot \mu'_i$ . Since  $\mu_i \approx^{\bullet} \nu_i$ , there exists  $\nu_i \xrightarrow{\alpha}_c \nu'_i$  such that  $\mu'_i \approx^{\bullet} \nu'_i$ , which implies that  $\nu \xrightarrow{\alpha}_c \nu' \equiv \sum_{i \geq 0} p_i \cdot \nu'_i$ . Therefore,  $\mu' \mathcal{R} \nu'$  by the definition of  $\mathcal{R}$ .

Now assume  $\mu$  is not transition consistent and  $\mu \equiv \sum_{1 \leq j \leq n} q_j \cdot \gamma_j$  such that  $\vec{\gamma}_j$ . Let  $\mu_i \equiv \sum_{1 \leq j \leq n} q_j^i \cdot \gamma_j^i$  where  $\vec{\gamma}_j^i, \gamma_j = \sum_{i \geq 0} q_j^i \cdot \gamma_j^i$ , and  $\sum_{i \geq 0} q_j^i = q_j$  for each  $1 \leq j \leq n$ . Then for each  $i \geq 0$ , there exists  $\nu_i \xrightarrow{\tau}_c \sum_{1 \leq j \leq n} q_j^i \cdot \gamma_j^i$  such that  $\vec{\gamma}_j^i$  and  $\gamma_j^i \approx^{\bullet} \gamma_j^i$ . Therefore, there exists  $\nu \xrightarrow{\tau}_c \sum_{1 \leq j \leq n} q_j \cdot \gamma_j$ , where  $\gamma_j \equiv \sum_{i \geq 0} q_j^i \cdot \gamma_j^i$ . By construction of  $\mathcal{R}$ ,  $\gamma_j \mathcal{R} \gamma_j$  for each  $1 \leq j \leq n$  as desired. □

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

**Lemma 3.**  $\mu \approx^{\bullet} \nu$  and  $\vec{\mu}$  imply  $\vec{\nu}$ .

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

Finally, we are ready to show the proof of Theorem 1.

*Proof of Theorem 1.* First, the second clause  $\overset{\bullet}{\approx} \subset \overset{\circ}{\approx}$  is easy to establish: Since the second condition of Definition 6 implies the second condition of Definition 9, but not vice versa. PA in Fig. 1 shows that the inclusion is strict.

Now we prove that  $\overset{\circ}{\approx}$  is an equivalence relation. We prove transitivity (other parts are easy). For any  $\mu, \nu$ , and  $\gamma$ , assume  $\mu \overset{\circ}{\approx} \nu$  and  $\nu \overset{\circ}{\approx} \gamma$ , we prove that  $\mu \overset{\circ}{\approx} \gamma$ . According to Definition 9, there exists late distribution bisimulations  $\mathcal{R}_1$  and  $\mathcal{R}_2$  such that  $\mu \mathcal{R}_1 \nu$  and  $\nu \mathcal{R}_2 \gamma$ . Let

$$\mathcal{R} = \mathcal{R}_1 \circ \mathcal{R}_2 = \{(\mu, \gamma) \mid \exists \nu. (\mu \mathcal{R}_1 \nu \wedge \nu \mathcal{R}_2 \gamma)\},$$

it then suffices to prove that  $\mathcal{R}$  is also a late distribution bisimulation.

Let  $\mu \mathcal{R} \gamma$  such that  $\mu \mathcal{R}_1 \nu$  and  $\nu \mathcal{R}_2 \gamma$  for some  $\nu$ . We shall prove:

1. Whenever  $\mu \xrightarrow{\alpha}_c \mu'$ , there exists  $\gamma \xrightarrow{\alpha}_c \gamma'$  such that  $\mu' \mathcal{R} \gamma'$ . This is achieved by applying Lemma 3.
2. If not  $\vec{\mu}$ , there exists  $\mu = \sum_{i \in I} p_i \cdot \mu_i$  and  $\gamma \xrightarrow{\tau}_c \sum_{i \in I} p_i \cdot \gamma_i$  such that  $\mu_i \mathcal{R} \gamma_i$  for each  $i \in I$ , where  $\sum_{i \in I} p_i = 1$ . Assume  $\mu$  is not transition consistent; otherwise it is easy. Since  $\mu \overset{\circ}{\approx} \nu$ , there exists  $\nu \xrightarrow{\tau}_c \sum_{i \in I} p_i \cdot \nu_i$  such that  $\vec{\mu}_i$  and  $\mu_i \mathcal{R}_1 \nu_i$  for each  $i \in I$ . By Lemma 3,  $\vec{\nu}_i$  for each  $i \in I$ . We distinguish the following two cases:
  - (a)  $\nu = \sum_{i \in I} p_i \cdot \nu_i$ .  
According to Lemma 3,  $\nu$  is not transition consistent, and moreover, we have  $\vec{\nu}_i$  for each  $i \in I$ . Since  $\nu \mathcal{R}_2 \gamma$ , there exists  $\gamma \xrightarrow{\tau}_c \sum_{i \in I} p_i \cdot \gamma_i$  such that  $\nu_i \mathcal{R}_2 \gamma_i$ , thus we have  $\mu_i \mathcal{R} \gamma_i$  by the definition of  $\mathcal{R}$  for each  $i \in I$ .
  - (b)  $\nu \xrightarrow{\tau}_c \nu' = \sum_{i \in I} p_i \cdot \nu_i$ .  
Since  $\nu \mathcal{R}_2 \gamma$ , there exists  $\gamma \xrightarrow{\tau}_c \gamma'$  such that  $\nu' \mathcal{R}_2 \gamma'$  according to the first clause of Definition 9. Since  $\mu$  is not transition consistent, so there exists  $i, j \in I$  such that  $i \neq j$  and  $EA(\mu_i) \neq EA(\mu_j)$ , which indicates that  $EA(\nu_i) \neq EA(\nu_j)$ . Therefore,  $\nu'$  is not transition consistent. As a result there exists  $\gamma' \xrightarrow{\tau}_c \sum_{i \in I} p_i \cdot \gamma_i$  such that  $\nu_i \mathcal{R}_2 \gamma_i$ , thus  $\mu_i \mathcal{R} \gamma_i$  for each  $i \in I$ .

This completes our proof. □

## 4 Properties of Late Distribution Bisimilarity

In this section we show that results established in [32] can be extended to the setting, where internal transitions are abstracted. 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 [7] and [18] for more details.

We first define some notations. To play with schedulers, we parameterize transition relations with schedulers explicitly. A transition from  $s$  to  $\mu$  with label  $\alpha$  is induced by a scheduler  $\xi$ ,

written as  $s \xrightarrow{\alpha}_{\xi} \mu$ , iff  $\mu \equiv \sum_{\mu' \in \text{Dist}(S)} \xi(s)(\alpha, \mu') \cdot \mu'$ . As before, such a transition relation can be lifted to distributions:  $\mu \xrightarrow{\alpha}_{\xi} \nu$  to denote that  $\mu$  can evolve into  $\nu$  by performing a transition with label  $\alpha$  under the guidance of  $\xi$ , where  $s \xrightarrow{\alpha}_{\xi} \nu_s$  for each  $s \in \text{Supp}(\mu)$  and  $\nu \equiv \sum_{s \in \text{Supp}(\mu)} \mu(s) \cdot \nu_s$ . Since no a priori information is available, given a distribution  $\mu$ , for each  $s \in \text{Supp}(\mu)$ , we simply use  $s$  as the history information for  $\xi$  to guide the execution, which correspond to *memoryless* schedulers and suffice for the purpose of defining bisimulations. Moreover, weak transitions  $s \xRightarrow{\alpha}_{\xi} \mu$  and their lifting to distributions can be defined similarly; see Section 3.1.

Below we define an alternative definition of Definition 9, where schedulers are considered explicitly.

**Definition 10.** Let  $\xi_1, \xi_2, \xi \in \mathcal{S}$  for a given set of schedulers  $\mathcal{S}$ .  $\mathcal{R} \subseteq \text{Dist}(S) \times \text{Dist}(S)$  is a late distribution bisimulation with respect to  $\mathcal{S}$  iff  $\mu \mathcal{R} \nu$  implies:

1. whenever  $\mu \xrightarrow{\alpha}_{\xi_1} \mu'$ , there exists  $\nu \xRightarrow{\alpha}_{\xi_2} \nu'$  such that  $\mu' \mathcal{R} \nu'$ ;
2. if not  $\vec{\mu}$ , then there exists  $\mu = \sum_{0 \leq i \leq n} p_i \cdot \mu_i$  and  $\nu \xRightarrow{\tau}_{\xi} \sum_{0 \leq i \leq n} p_i \cdot \nu_i$  such that  $\vec{\mu}_i$  and  $\mu_i \mathcal{R} \nu_i$  for each  $0 \leq i \leq n$  where  $\sum_{0 \leq i \leq n} p_i = 1$ ;
3. symmetrically for  $\nu$ .

We write  $\mu \approx_{\mathcal{S}} \nu$  iff there exists a late distribution bisimulation  $\mathcal{R}$  with respect to  $\mathcal{S}$  such that  $\mu \mathcal{R} \nu$ . And we write  $s \approx_{\mathcal{S}} r$  iff  $\delta_s \approx_{\mathcal{S}} \delta_r$ .

Different from Definition 9, in Definition 10, 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,  $s_1 \approx^{\bullet} s_2 \iff s_1 \approx_{\mathcal{S}_D}^{\bullet} s_2$ , provided  $s_1$  and  $s_2$  contain no parallel operators, as in this case  $\mathcal{S}_D$  represents the set of all schedulers.

Below is a theorem showing 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. Furthermore, late distribution bisimulation implies trace distribution equivalence under partial information schedulers.

**Theorem 2.** For any states  $s_1$  and  $s_2$ ,

1.  $s_1 \approx^{\bullet} s_2$  iff  $s_1 \approx_{\mathcal{S}_p}^{\bullet} s_2$ ;
2.  $s_1 \approx^{\bullet} s_2$  implies  $s_1 \equiv_{\mathcal{S}_p} s_2$ .

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

**Theorem 3.** For any states  $s_1, s_2$ , and  $s_3$ ,

$$s_1 \approx_{\mathcal{S}_D}^{\bullet} s_2 \text{ implies } s_1 \parallel_A s_3 \approx_{\mathcal{S}_D}^{\bullet} s_2 \parallel_A s_3.$$

As in the strong setting [32], 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 4.** *Let  $\mathcal{S} = \mathcal{S}_P \cap \mathcal{S}_D$ .  $s_1 \approx_{\mathcal{S}}^c s_2$  iff  $s_1 \equiv_{\mathcal{S}}^c s_2$  for any  $s_1$  and  $s_2$ , where  $s_1 \equiv_{\mathcal{S}}^c s_2$  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 multiagent systems [3], where all agents are autonomous (corresponding to distributed schedulers) and states are partially observable (corresponding to partial information schedulers).

In [24] an algorithm was proposed to compute distribution-based bisimulation relations. We discuss briefly that the algorithm can also 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  $\alpha$ , it is easy to see that  $\{\mu \mid s \xrightarrow{\alpha} \mu\}$  constitutes a convex hull. According to [5, Prop. 3 and 4], every such convex hull 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 [24]. Not surprisingly, this will cause an exponential blow-up. We refer readers to [24] for more details of the remaining procedure.

## 5 Conclusion and Future Work

In this paper, we proposed the notion of late distribution bisimilarity for PAs, which enjoys some interesting properties if restricted to the 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 shown to be the coarsest relation which is compositional and preserves trace distribution equivalence.

As future work we intend to study reduction barbed congruences [8] 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 [24] is exponential in the worst case. We will work out whether or not more efficient algorithms exist.

**Acknowledgments.** Many thanks to the anonymous referees for their valuable suggestions on an early version of this paper. This work has been supported by the DFG as part of the SFB/TR 14 “Automatic Verification and Analysis of Complex Systems” (AVACS), by the European Union Seventh Framework Programme under grant agreements 295261 (MEALS) and 318490 (SENSATION), by the National Natural Science Foundation of China (Grant Nos. 61428208,



61472473 and 61361136002), and by the CAS/SAFEA International Partnership Program for Creative Research Team. Part of this work was done while Lei Song was at Max-Planck Institute for Informatics and Saarland University in Saarbrücken, Germany.

## Bibliography

- [1] C. Baier, J.-P. Katoen, H. Hermanns, and V. Wolf. Comparative branching-time semantics for Markov chains. *Inf. Comput.*, 200(2):149–214, 2005.
- [2] M. Bernardo, R. De Nicola, and M. Loreti. Relating strong behavioral equivalences for processes with nondeterminism and probabilities. *Theor. Comput. Sci.*, 546:63–92, 2014.
- [3] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein. The complexity of decentralized control of Markov decision processes. *Math. Oper. Res.*, 27(4):819–840, 2002.
- [4] H. Boudali, P. Crouzen, and M. Stoelinga. A rigorous, compositional, and extensible framework for dynamic fault tree analysis. *IEEE Trans. Dependable Sec. Comput.*, 7(2):128–143, 2010.
- [5] S. Cattani and R. Segala. Decision algorithms for probabilistic bisimulation. In *CONCUR*, pages 371–385, 2002.
- [6] G. Chehaibar, H. Garavel, L. Mounier, N. Tawbi, and F. Zulian. Specification and Verification of the PowerScale™ bus arbitration protocol: An industrial experiment with lotos. In *FORTE*, pages 435–450, 1996.
- [7] L. De Alfaro. The verification of probabilistic systems under memoryless partial-information policies is hard. Technical report, DTIC Document, 1999.
- [8] Y. Deng and M. Hennessy. On the semantics of Markov automata. *Information and Computation*, 222:139–168, 2013.
- [9] Y. Deng, R. van Glabbeek, M. Hennessy, and C. Morgan. Testing finitary probabilistic processes. In *CONCUR*, pages 274–288, 2009.
- [10] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Weak bisimulation is sound and complete for pCTL<sup>\*</sup>. *Inf. Comput.*, 208(2):203–219, 2010.
- [11] L. Doyen, T. A. Henzinger, and J. Raskin. Equivalence of labeled Markov chains. *Int. J. Found. Comput. Sci.*, 19(3):549–563, 2008.
- [12] C. Eisentraut, J. C. Godskesen, H. Hermanns, L. Song, and L. Zhang. Late Weak Bisimulation for Markov Automata. *CoRR*, abs/1202.4116, 2014. <http://arxiv.org/abs/1202.4116>.

- [13] C. Eisentraut, H. Hermanns, J. Katoen, and L. Zhang. A semantics for every GSPN. In *PETRI NETS*, volume 7927 of *Lecture Notes in Computer Science*, pages 90–109. Springer, 2013.
- [14] C. Eisentraut, H. Hermanns, J. Kraemer, A. Turrini, and L. Zhang. Deciding bisimilarities on distributions. In *QEST*, volume 8054 of *Lecture Notes in Computer Science*, pages 72–88, Berlin, 2013. Springer.
- [15] C. Eisentraut, H. Hermanns, and L. Zhang. Concurrency and composition in a stochastic world. In *CONCUR*, pages 21–39, 2010.
- [16] C. Eisentraut, H. Hermanns, and L. Zhang. On probabilistic automata in continuous time. In *LICS*, pages 342–351, 2010.
- [17] Y. Feng and L. Zhang. When equivalence and bisimulation join forces in probabilistic automata. In *FM*, volume 8442 of *Lecture Notes in Computer Science*, pages 247–262. Springer, 2014.
- [18] S. Giro and P. R. D’Argenio. Quantitative model checking revisited: neither decidable nor approximable. In *FORMATS*, volume 4763 of *Lecture Notes in Computer Science*, pages 179–194. Springer, 2007.
- [19] D. Guck, M. Timmer, H. Hatefi, E. Ruyjters, and M. Stoelinga. Modelling and analysis of Markov reward automata. In *ATVA*, volume 8837 of *Lecture Notes in Computer Science*, pages 168–184. Springer, 2014.
- [20] P. R. Halmos. *Measure theory*, volume 1950. Springer, 1974.
- [21] F. He, X. Gao, B. Wang, and L. Zhang. Leveraging weighted automata in compositional reasoning about concurrent probabilistic systems. In *POPL*, pages 503–514. ACM, 2015.
- [22] M. Hennessy. Exploring probabilistic bisimulations, part I. *Formal Asp. Comput.*, 24(4-6):749–768, 2012.
- [23] H. Hermanns. *Interactive Markov chains: and the quest for quantified quality*. Springer, Berlin, Heidelberg, 2002.
- [24] H. Hermanns, J. Krcál, and J. Kretínský. Probabilistic bisimulation: Naturally on distributions. In *CONCUR*, volume 8704 of *Lecture Notes in Computer Science*. Springer, 2014.
- [25] K. Honda and M. Tokoro. On asynchronous communication semantics. In *Object-Based Concurrent Computing*, pages 21–51, 1991.
- [26] M. Kwiatkowska, G. Norman, D. Parker, and H. Qu. Assume-guarantee verification for probabilistic systems. In *TACAS*, volume 6015 of *Lecture Notes in Computer Science*, pages 23–37. Springer, 2010.

- [27] A. Philippou, I. Lee, and O. Sokolsky. Weak bisimulation for probabilistic systems. In *CONCUR*, pages 334–349, 2000.
- [28] W. Rudin. *Real and complex analysis*. Tata McGraw-Hill Education, 2006.
- [29] J. Schuster and M. Siegle. Markov automata: Deciding weak bisimulation by means of non-nately vanishing states. *Information and Computation*, 237(0):151 – 173, 2014.
- [30] R. Segala. A compositional trace-based semantics for probabilistic automata. In *CONCUR*, volume 962 of *Lecture Notes in Computer Science*, pages 234–248. Springer, 1995.
- [31] R. Segala. *Modeling and Verification of Randomized Distributed Realtime Systems*. PhD thesis, MIT, 1995.
- [32] L. Song, Y. Feng, and L. Zhang. Decentralized bisimulation for multiagent systems. In *AAMAS*. IFAAMAS, to appear, 2015.
- [33] M. Timmer, J. van de Pol, and M. Stoelinga. Confluence reduction for Markov automata. In *FORMATS*, volume 8053 of *Lecture Notes in Computer Science*, pages 243–257. Springer, 2013.

## MEALS Partner Abbreviations

**SAU:** Saarland University, D

**RWT:** RWTH Aachen University, D

**TUD:** Technische Universität Dresden, D

**INR:** Institut National de Recherche en Informatique et en Automatique, FR

**IMP:** Imperial College of Science, Technology and Medicine, UK

**ULEIC:** University of Leicester, UK

**TUE:** Technische Universiteit Eindhoven, NL

**UNC:** Universidad Nacional de Córdoba, AR

**UBA:** Universidad de Buenos Aires, AR

**UNR:** Universidad Nacional de Río Cuarto, AR

**ITBA:** Instituto Tecnológico Buenos Aires, AR